Amirkabir Journal of Civil Engineering

Amirkabir Journal of Civil Engineering, 49(2) (2017), 83-85
DOI:10.22060/ceej.2016.710

Performance Improvement of Biological BOD in Rivers based on De-noising Comparison Wavelet-ANN Conjunction, GP, ANN and MLR Methods (Case Study: Karaj Dam Outlet Station)

Taher Rajaee* , Hamideh Jafari, Roghaye Rahimi

Civil Engineering Department, University of Qom, Qom, Iran

ABSTRACT: This study considered artificial neural network (ANN), multi-linear regression (MLR), Genetic Programming (GP) and wavelet analysis and ANN combination (WANN) models for monthly water biological oxygen demand (BOD) in station Karaj Dam outlet and investigates the effects of data preprocessing on model performance using discrete wavelet. For this purpose, in the first proposed model, observed time series of BOD were decomposed into several sub-time series at different scales by discrete wavelet transform. Then these sub-time series were imposed as inputs to the ANN method. In the second proposed model, observed time series of BOD were decomposed at ten scales by wavelet analysis. Then, total effective time series BOD were imposed as inputs to the neural network model for prediction of BOD in one month ahead. Results showed that the wavelet neural network models performance was better in prediction rather than the neural network and multi-linear regression models. The wavelet analysis model produced reasonable predictions for the extreme values. This model dropped the mean absolute percentage error for the MLR, GP, ANN and the first hybrid models from 1.87, 0.91, 0.65 and 0.46 respectively, to 0.44 and increased the Nash-Sutcliffe model efficiency coefficient from 0.23, 0.53, 0.73 and 0.81 to 0.83..

Review History:
Received: 27 May 2014
Revised: 23 August 2015
Accepted: 12 August 2016
Available Online: 11 March 2017

Keywords:
Artificial neural network
BOD
De-noising
Karaj River
Wavelet transformy

1- . Introduction

Surface water as one of the most important resources of supplying drinking water is in danger of contaminants entry. Water-quality monitoring is one of the significant concerns all over the world [1]. There are stable links between physical, chemical and biological activities in surface water. Hence, with entry of the exterior materials, environmental interactions begin simultaneously. Environmental contamination indexes in a river can be assessed by considering BOD [2]. The BOD of any aquatic system is the foremost parameter needed for assessment of the water quality as well as development of management strategies for the protection of water resources [3]. Most of the methods common for determining BOD are hard to use and also include lots of measurement errors. So having a pattern to exactly show the present condition, predict the later condition and also to determine the essential limitations for accessing to quality standards is necessary. Water quality prediction is the foundation of water pollution control project. It predicts how water quality changes to the entry of contaminants to the river ecosystem. Due to the complexity of water quality in a river, the black box (lumped) modeling may have some avails over the modeling by the theoretical ruling equations. Several linear and nonlinear methods have been applied in the prediction of water quality in rivers and successful results have been reported. In recent years, artificial intelligence (AI) techniques such as ANN and GP have been pronounced as a branch of computer science to model wide range of water and environmental processes [4]. ANNs were used in various fields of water quality prediction. In the last decade, GP model has been successfully used for non-linear system modeling to solve wide range of modeling problems in water resources engineering such as hydrological and environmental modeling. Due to the non-linearity and non-stationarity of water quality indicator series, the accuracy of the commonly used conventional methods, including MLR, ANN and GP models has been limited. Hence, pre-analysis on the model inputs can improve the results. One of the pre-analysis methods which are used in the time series is wavelet transform. Using wavelet transform for analyzing and predicting time series is an efficient method which is expanding recently. Wavelet transform can decompose a non-stationary signal into a certain number of stationary sub-signals. Then, ANN is combined with the wavelet transform to improve precision of the prediction. Hybrid WANN model has been successfully applied in recent years to predict hydrological and hydro geological processes [5]. The current study is a new research that uses the conjunction of ANN and wavelet-based data pre-processing (WANN) in order to predict time series of qualitative parameter of the BOD. For this purpose, different combinations of inputs will be created and it will be modeled by the WANN model, MLR method and ANN. The rest of the paper is organized as following: the study area and data analysis are presented in section one. In section two, MLRs, GPs, ANNs, wavelet transform, WANNs model applications are described.

*Corresponding Author, Email: Trajaee@qom.ac.ir

2- Study Area

Karaj River is a river of northern *Iran* in a section of the south-central Alborz protected area between the geographical latitudes of 35_550 to 36_90N and geographical longitudes of 51 to 51_450W, located between the cities of Tehran, Karaj, Chalus, Noshahr and Nour. Karaj River is the most important source of water supply for Tehran province. Most of the Karaj River and its branches flow are used for agricultural, municipal and industrial uses of Tehran province (including Tehran, Karaj, Damavand, Varamin and Shahriar), and the rest of it flows into Qom salt lake. The data derived from the Karaj River in Karaj Dam output was employed to train and test of all the models developed in this study. Data used in this research includes a 145-month time series in these stations. About 80 percent of data are used for train set and the rest of them are used for validation set. In different prediction water quality methods, training and validation sets are chosen therefore, in this study, data were selected by non-random sampling to better compare the methods done. In this research, BOD data of Tehran's regional water company were used to the assessment of water quality of Karaj River.

Table 1: Statistical analysis for testing, training and all data sets.

Statistic parameters	All Data	Testing set	Training set
Mean	1.57	1.5	1.58
S_d	0.73	0.74	0.73
C_{sx}	0.55	0.76	0.34
Minimum	0.3	0.3	0.32
Maximum	4.8	4.8	3.8
R_1	0.7	0.68	0.71
R_2	0.54	0.51	0.54
R_3	0.51	0.56	0.48
R_4	0.68	0.66	0.65

The statistical analysis is given in Table 1, which include the mean, standard deviation (S_d), skewness coefficient (C_{sx}), minimum, maximum, one month lag autocorrelation coefficient (R_1), two month lag autocorrelation coefficient (R_2), three month lag autocorrelation coefficient (R_3), and four month lag autocorrelation coefficient (R_4) of data [6, 7].

3- Wavelet-ANN Model

The WANN tries to combine the attributes of ANNs with the characteristics of wavelet transform. According to the time series length (145 values) of the observed data, the maximum calculated decomposition level was 5. Afterward, the decomposed BOD time series were used at different levels as inputs to the ANN method for predicting the one-month-ahead BOD.

4- Model Application

4- 1- Model Development

With respect to the statistical analysis in Table 1, the following combinations containing different numbers of input values of the BOD were considered in the input layer to predict the unique BOD value at time t (BODt) in the output layer of the research models [6]:

1. BOD_{t-1} ;
2. BOD_{t-1}, BOD_{t-3};
3. BOD_{t-1} , BOD_{t-6};
4. BOD_{t-1} , BOD_{t-3}, BOD_{t-6};
5. BOD_{t-1} , BOD_{t-3}, BOD_{t-4}, BOD_{t-6};

4- 2- Performance Evaluation

The MAE, RMSE and E performance evaluation criteria employed in this paper can be computed utilizing the following equations:

Mean absolute error (*MAE*):

$$MAE = \frac{\sum_{i=1}^{N} \left| BOD_i^{Observed} - BOD_i^{Predicted} \right|}{N} \quad (1)$$

Root mean square error (*RMSE*):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(BOD_i^{Observed} - BOD_i^{Predicted} \right)^2}{N-1}} \quad (2)$$

Nash-Sutcliffe criteria (*E*):

$$E = 1 - \frac{\sum_{t=1}^{N} \left(BOD_i^{Observed} - BOD_i^{Predicted} \right)^2}{\sum_{t=1}^{N} \left(BOD_i^{Observed} - \overline{BOD^{Observed}} \right)^2} \quad (3)$$

E is a non-dimensional criterion which it can be employed to compare the relative performance of the model. A perfect agreement between the observed and predicted values yields an E equal to 1. The RMSE, defined as a percentage of the observed mean value of BOD, gives a measure of the difference of the predictions with the observed values [6].

5- Results and Discussion

Time series analysis methods are helpful tools to investigate long-term ecological observations. They serve as means to understand the time and frequency structure of measurements.

The goal of multiple regression analysis is to evaluate the relationship between several independent or predictor variables and a dependent or criterion variable.

This reveals that the traditional ANN and GP model predictions are not consistent and depict uncertainty, whereas the WANN models predictions are more consistent and can be considered more reliable.

For successful implementation of wavelets based on forecasting methods, selection of the appropriate mother wavelet form and number of decomposition levels play an important role. In this study, the Dmey mother wavelet

provides considerably better outcomes than the others based on efficiency criteria. Water quality is the result of different factors in a river such as geological, climate and human pollution sources. Also time series of each parameter encompasses complicated data that are including rich information about water resources behaviors.

The results showed that WANN method performs satisfactorily and may be considered to be alternatives to the other methods. By using different decomposition levels for each variable in the WANN model, different characteristics are considered, including long, intermediate, and short levels. When sub-time series are entered the hybrid models as input conditions, their assigned weights by the ANN model will be distinct at different decomposition levels; therefore high weights will be applied to the high levels of the signals. Thus, the network magnifies its weight compared to the other sub-series. Comparing the efficiency coefficients of the WANN model to those of the ANN model shows that decomposing of the input time series before applying the ANN model decreases the RMSE and MAE as well as increases E. The same result is achieved for the ANN models with three inputs. Among all applied models, the WANN model has the best efficiency.

6- Conclusion

This section presents the predictions performed by the MLR, GP and the ANN models for all input combinations. The SPSS 22.0 software package is used for regression calculations. When multi-level sub-time series are entered the hybrid models as input conditions, their assigned weights by the ANN technique will be different at different decomposition levels; therefore, high weights will be applied to the high levels of the signals. Thus, the network magnifies its weight compared to the other sub-signals. Another reason for good performance of the hybrid models is that the wavelet transformation reduces the noises in the water level time series and it causes forecasts to be more reliable and accurate. This paper presents a comparative evaluation for different wavelet forms when employed for forecasting future states of various kinds of time series. The results showed that wavelet is a useful tool for increasing the prediction of ANN and GP methods. According to the trend of the time series, it seems that the water is polluted in the summer and fall and this is showing the relations between BOD and discharge value severity. BOD changes in Karaj River were related to climate changes, the slope of the river, the river erosion, water flow rate and discharge. These changes are within the specified limits. Increase or decrease in the temperature can directly change the level of chemical reactions. It is recommended that future studies explore the use of the WANN method in BOD forecasting for other watersheds, in different geographical areas, for other lead times (such as daily, weekly, or yearly forecasting), comparing the WANN model to other new methods such as support vector regression with localized multiple kernel learning; and ensemble forecasting by the use of the bootstrap method to develop wavelet-boot strap neural network models.

References

[1] T., Rajaee; Wavelet-ANN combination model for prediction of daily suspended sediment load in rivers, Science of The Total Environment,Vol. 409, No. 15, pp. 2917- 2928, 2011.

[2] J., Adamowski; F. H., Chan; A wavelet neural network conjunction model for groundwater level forecasting, Journal of Hydrology, Vol. 407, No. 1-4, pp.28- 40, 2011.

[3] J.G., Han; W.X., Ren; Z.S., Sun; Wavelet packet based damage identification of beam structures, International Journal of Solids and Structures, Vol. 42, No. 26, pp. 6610-6627, 2005.

[4] V., Nourani; M.T., Alami; M.H., Aminfar; A Combined Neural-Wavelet Model for Prediction of Ligvanchai Watershed Precipitation, Engineering Application of Artificial Intelligence, Vol. 22, No. 3, pp.146-477, 2009.

[5] V., Nourani; A., Hosseini Baghanam; A., Adamowski; M., Gebremichael; Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and run off data in neural network based rainfall–runoff modeling, Journal of Hydrology,Vol. 476, pp.228-243, 2013.

[6] A., Najah; A., Elshafie; O., Karimi; O., Jaffer; Prediction of Johor river water quality parameters using artificial neural networks, European Journal of Scientific Research, Vol. 28, pp. 422-435, 2009.

[7] T., Rajaee; V., Nourani; M., Zounemat-Kermani; O., Kisi; River Suspended Sediment Load Prediction: Application of ANN and Wavelet Conjunction Model, Journal of Hydrologic Engineering, Vol. 16, pp.613-627, 2011.

Please cite this article using:
T. Rajaee, H. Jafari, R. Rahimi, "Performance Improvement of Biological BOD in Rivers based on De-noising Comparison Wavelet-ANN Conjunction, GP, ANN and MLR Methods (Case Study: Karaj Dam Outlet Station)".*Amirkabir J. Civil Eng.,* 49(2) (2017) 83-85.
DOI:10.22060/ceej.2016.710