Research Journal of Applied Sciences, Engineering and Technology 6(2): 345-348, 2013
DOI:10.19026/rjaset.6.4084
ISSN: 2040-7459; e-ISSN: 2040-7467
© 2013 Maxwell Scientific Organization Corp.
Submitted: November 20, 2012 Accepted: December 15, 2012 Published: June 10, 2013

Research Article

A Wavelet Neural Network Hybrid Model for Monthly Ammonia Forecasting in River Water

[1]Yi Wang, [1]Yuanyuan Wang, [1, 2]Liang Guo, [1, 2]Ying Zhao,
[1]Zhichao Zhang and [1, 2]Peng Wang
[1]School of Municipal and Environment Engineering, Harbin Institute of Technology, Harbin, 150090, Heilongjiang Province, China
[2]State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology, Harbin 150090, China

Abstract: Forecasting water quality is always an effective approach for water environmental management. This study presents a combined Wavelet transform (WA) and Artificial Neural Network (ANN) model for monthly ammonia nitrogen series prediction in river water. The WA decomposed original time series into different subseries, in which the most significant one was chosen as the training data instead of the original series. Compared to the traditional ANN, the WA-ANN models were found more accurate and reliable. The results of the study indicate that WA could remove the noise of the original datasets and the WA-ANN could help environment decision-maker manage water quality more effective.

Keywords: Artificial neural network, environment management, Harbin region, time series, water quality forecasting, wavelet analysis

INTRODUCTION

Forecasting water quality is always an effective approach for water quality management. In many rivers, ammonia nitrogen (NH_3-N) is still an important water quality variable because it could lead to potential adverse effects on human health and environment (Powlson *et al*., 2008). In order to predict the concentration of NH_3-N in river water, mechanism models are often used to describe the physical or biochemical processes occurring in river water system. However, they are constrained by numerous historical data, which are usually insufficient or missing in some watershed. Sometimes obtaining accurate predictions is more important than understanding underlying mechanism. Thus, many studies have focused on data-driven models such as Artificial Neural Network (ANN) as a suitable alternative to predict the environmental time series (Taormina *et al*., 2012).

The ANN has capability to treat non-linear and complicated environmental problems and it had a wide range of applicability and was ideally suited for modeling the non-linear time series (Zou *et al*., 2010). Nevertheless, ANNs have limitations that they often handle non-stationary data ineffectively without input data preprocessing (Cannas *et al*., 2006). To address the problem, Wavelet Analysis (WA) appears to be an effective method employing wavelet basis function as pre-transforming tool to decompose observed time series into various components, which were set as new time series inputs of different sub-ANNs (Abaurrea *et al*., 2011; Rajaee, 2011; Tiwari and Chatterjee, 2010). In this study, WA-ANN models were developed and compared with just ANN models for monthly time lag forecasting of ammonia nitrogen in Songhua River-Harbin Region, China.

METHODOLOGY

Artificial neural network: Artificial Neural Network (ANN) imitates human neural system to form the model structure of a neural network that is highly flexible for simulating complex nonlinear relationships, particularly the relation between unknown variables involved (Kuo *et al*., 2007). The basic structure of the ANN model is usually comprised of three distinctive layers and the nodes from one layer are connected to all nodes in the next adjacent layers. Trained by Back Propagation (BP) algorithm, the ANN can be proposed when arriving at a best fit with minimum error. Then the output y_n and the training error E are computed as:

$$y_n = \sum_{j=1}^{J} w_{jk} f\left(\sum_{i=1}^{h} w_{ij} x_i\right) \tag{1}$$

Corresponding Author: Peng Wang, School of Municipal and Environment Engineering, Harbin Institute of Technology, Harbin, 150090, Heilongjiang Province, China
This work is licensed under a Creative Commons Attribution 4.0 International License (URL: http://creativecommons.org/licenses/by/4.0/).

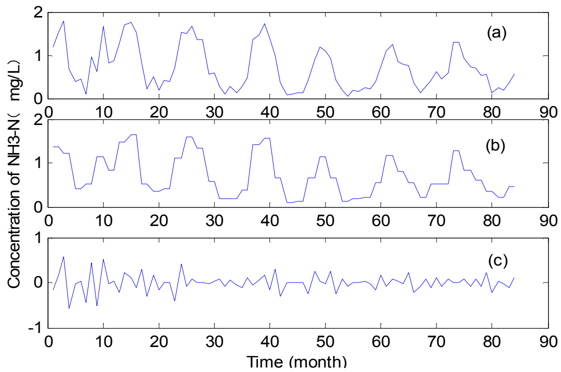

Fig. 1: Original time series (a), discrete components approximation A1 (b) and detail D1 (c) wavelet of NH_3-N concentrations of Zhushuntun during the training years

$$E = \frac{1}{2}\sum_{n=1}^{N}(y_n - t_n)^2 \tag{2}$$

where,

w_{ij} & w_{jk} : The connections weights whose values are trained
f : Normally the sigmoid function
h & J : The respective number of input and hidden layers
x_i : The model input variable

Different time lags x_{t-1}, x_{t-2} and x_{t-3} are selected respectively as the input variable to predict the object x_t. The target training goal E and the maximum epochs were set to 10^{-3}, 1000, respectively. Then the ANNs were trained for 1-10 hidden nodes and the structure with the lowest E was chosen as the optimum.

Wavelet analysis: A series of sub-wavelet $\Psi_{a,b}$ (x) can be derived by wavelet transform:

$$\psi_{a,b}(x) = |a|^{-1/2}\,\psi\left(\frac{x-b}{a}\right) \tag{3}$$

where,

a = The scale factor for dilation
b = The translation factor for temporal translation or shift of the wavelet function
$\Psi(x)$ = The mother wavelet

Discrete wavelet transform is convenient to apply and suited for the situation in which the available data are sparse. This can be achieved by the modified way (Grossman and Morlet, 1984):

$$\psi_{m,n}(t) = s_0^{-m/2}\,\psi\left(\frac{t - np_0 s_0^m}{s_0^m}\right) \tag{4}$$

where,

s : The wavelet scale
t : The time
p : The translation parameter
m : Numbers that control scale
n : Numbers that control time
s_0 : A specified fixed dilation step greater than 1
p_0 : The location parameter that greater than 0

Then the original time series is decomposed to an approximation (low-pass filter) and other one or more details (high-pass filter) representing the original information. In this study, discrete db1 wavelet basis function was applied as the node activation function.

Hybrid WA and ANN: The original datasets was decomposed into L levels, where L = int [log (N)] (Nourani *et al*., 2009). N = 60 is the number of time series data, so L is one in this study. The components are shown in Fig. 1. The effective wavelet subseries are determined using the correlation coefficients between each subseries and observed NH_3-N at Zhushuntun. The coefficients of the Approximation (A1) and Detail (D1) are respect 0.9223 and 0.3865 that the WA-ANN hybrid models use only the A1, instead of original data, as the new training data.

Model performance: Nash and Shutcliff (1970) Efficiency (NSE) and Root Mean Squared Error (RMSE) are used to evaluation the models. The NSE Eq. (5) is a practical index for assessing model predictive power, which has statistical character to measure the difference between the predictions and the mean of observations. The best value of NSE is 1, representing best fit of the model. In this study, the forecasting is considered satisfactory if the NSE is bigger than 0.5. The RMSE Eq. (6) provides statistical information about the effects of root mean square error on observations and lower value is preferred:

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}(t_i - y_i)^2}{\sum_{i=1}^{N}(t_i - \bar{t}_i)^2} \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N}\frac{(t_i - y_i)^2}{N}} \tag{6}$$

where,

t_i = The i^{th} observations
y_i = The i^{th} predictions
$\bar{t}_i$ = The mean of the observations
N = The number of observations in testing data sets

Study area and data: Harbin is the capital of Heilongjiang province in Northeast of China, locating on the Songhua River which is the third biggest river in

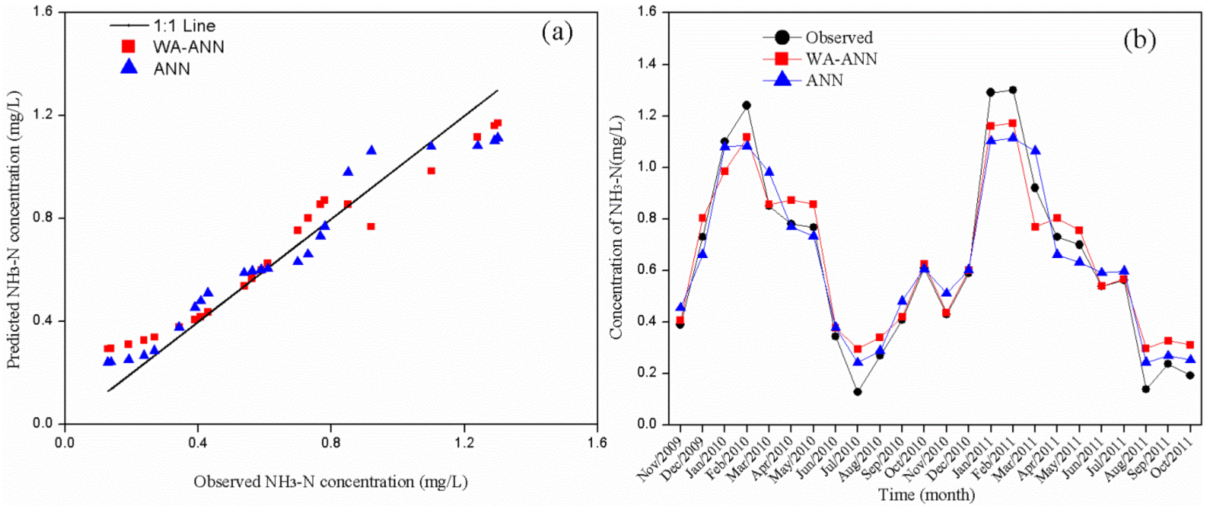

Fig. 2: Scatter (a) and curve (b) comparison of predicted and observed concentration of NH_3-N in Zhushuntun station using WA-ANN and ANN model with 1 month lag during the testing period

China. The concentrations of NH_3-N were collected monthly in Zhushuntun sampling site, used to be water intake of Harbin city, during years 2005-2011. For ANN models, the data series were divided into a training set (January/2005-Octover/2009) and a testing set (November/2009-Octover/2011).

RESULTS AND DISCUSSION

For all the best WA-ANN models, with different input time lag, have the same five hidden neurons and they are all satisfactory, while the one month ahead as the input has the highest NSE the lowest value of RMSE (Table 1). Similarly, for all the best ANN models have six hidden neurons and the one-month ahead as the input has the highest NSE the lowest value of RMSE. However, the performances of the best WA-ANN are better than the best ANN results.

Figure 2a compares the observations and the predictions during the testing period for the best WA-ANN and ANN models. It can be seen that they all have satisfactory fit, while WA-ANN performance better. Besides, the predictions of ANN are overestimate in high-flow months (i.e., from June to October) with lower concentration and the underestimation occurs in low-flow period (i.e., January and February) (Fig. 2b, which may due to a systematic shift. In Harbin region, this can be attributed to the severe changes of concentration of NH_3-N. The mean concentration of NH_3-N in low-flow period is usually near eight times greater than that in high-flow period, but the transitional period when ice freezing or thawing, between the high and low-flow periods lasts only one month. Therefore, the monthly value of NH_3-N concentration in transitional period fells or arises

Table 1: Model performance of WA-ANN and ANN models with different time lags input variables

	Best WA-ANN		Best ANN	
Input	NSE	RMSE (mg/L)	NSE	RMSE (mg/L)
x_{t-1}	0.7588	0.0837	0.7380	0.0909
x_{t-2}	0.7397	0.0906	0.6962	0.1055
x_{t-3}	0.7295	0.0939	0.6488	0.1219

sharply. The comparisons are due to de-noise of WA that can deal with the severely fluctuating data sets, which may distort the original information. Thus, adding wavelet transform as the preprocessing for ANN are more accurate then without it when forecasting the monthly NH_3-N series in Songhua River Harbin Region.

CONCLUSION

In this study, artificial neural network coupling discrete Wavelet transforms (WA) for forecasting NH_3-N in river water was presented to help environmental management. The NH_3-N time series was decomposed into an approximation series with correlation coefficient 0.9223 and a detail series with 0.3865. Then the significant series A1 was used based on the correlation coefficient to train the ANN models. Different month lags were compared and one-month lag was determinate to the best input variable to predict the concentration of NH_3-N. However, the WA-ANN was more accurate, with NSE0.7588 and RMSE 0.0837, than the only ANN models in Zhushuntun Station of Harbin Region, because the WA could remove the noise of the original datasets.

ACKNOWLEDGMENT

This study was supported by the Funds NSFC for Creative Research Groups of China (No. 51121062) and National Natural Science Foundation of China (No. 71203041). The authors thank the Environmental Monitoring Center of Harbin City for data supply.

REFERENCES

Abaurrea, J., J. Asín, A.C. Cebrián and M.A. García-Vera, 2011. Trend analysis of water quality series based on regression models with correlated errors. J. Hydrol., 400(3-4): 341-352.

Cannas, B., A. Fanni, L. See and G. Sias, 2006. Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. Phy. Chem. Earth., 31(18): 1164-1171.

Grossman, A. and J. Morlet, 1984. Decompositions of hardy functions into square integrable wavelets of constant shape. SIAM J. Math. Analys., 15: 723-736.

Kuo, J.T., M.H. Hsieh, W.S. Lung and N. She, 2007. Using artificial neural network for reservoir eutrophication prediction. Ecol. Model., 200(1-2): 171-177.

Nash, J.E. and J.V. Shutcliff, 1970. River flow forcasting through conceptual models. I. J. Hydrol., 10: 282-290.

Nourani, V., M. Alami and M. Aminfar, 2009. A combined neural-wavelet model for prediction of ligvanchai watershed precipitation. Eng. Appl. Artif. Intel., 22: 466-472.

Powlson, D.S., T.M. Addiscott, N. Benjamin, K.G. Cassman and T.M. Dekok, 2008. When does nitrate become a risk for humans. J. Environ. Qual., 37(2): 291-295.

Rajaee, T., 2011. Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers. Sci. Total. Environ., 409(15): 2917-2928.

Taormina, R., K.W. Chaua and R. Sethib, 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. Eng. Appl. Artif. Intell., 25(8): 1670-1676.

Tiwari, M.K. and C. Chatterjee, 2010. Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. J. Hydrol., 394: 458-470.

Zou, P., J. Yang, J. Fub, G. Liua and D. Lia, 2010. Artificial neural network and time series models for predicting soil salt and water content. Agr. Water Manage., 97(12): 2009-2019.