Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2015, 7(6):622-626

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Artificial neural network model of algae density in Xiangxi Bay of Three Gorges Reservoir

Jun-Yu Liu[1], Hua-Jun Luo[2*] and Ying-Ping Huang[3]

[1]*Math Department, Califonia State University of Northridge, Los Angeles, USA*
[2]*College of Biological and Pharmaceutical Science, China Three Gorges University, Yichang, China*
[3]*Engineering Research Center of Eco-environment in Three Gorges Reservoir Region, Ministry of Education, China Three Gorges University, Yichang, China*

ABSTRACT

Algae density model of diatom and blue algae in Xiangxi Bay of Three Gorges Reservoir was established by artificial neural network (ANN) technology. Using stepwise multiple linear regression method, the important environmental factors (nitrate, ammonium nitrogen, total phosphates, phosphate, silicate and water temperature) were selected as input variables in ANN model. The optimal structure of the ANN model was determined, which has two hidden layers (number of neurons in the first hidden layer: 5; number of neurons in the second hidden layer: 4). The ANN model has been shown to perform well for simulating the algal density of diatom and blue algae simultaneously, in which the training set R^2 values are 0.738 (diatom) and 0.949 (blue algae), the test set R^2 values are 0.773 (diatom) and 0.870 (blue algae), respectively.

Keywords: Algae density; Artificial neural network; Xiangxi Bay; Three Gorges Reservoir

---

INTRODUCTION

Since 2009 Three Gorges reservoir (TGR) was created by Three-Gorge Dam in China measuring an area of 1080 km^2 [1]. With impoundment of Three Gorges Reservoir, the Xiangxi River, which is the largest tributary in the Hubei portion of TGR, was inundated and the water flow velocity in Xiangxi Bay has dropped from the original 0.43-0.92 m/s [2] to 0.0020-0.0041 m/s [3]. So with prolonged retention time and high nutrient concentrations, there were algal blooms in Xiangxi Bay.

To monitoring and controlling algal blooms, it is necessary to establish an algal response model that can effectively simulate the timing and magnitude of algae density. Although statistical and deterministic models such as hydrodynamic and water quality models have been the traditional approaches for modeling water environment, these require a significant amount of field data to support the analysis due to the complicated nonlinear processes in algal blooms [4]. Recently, artificial neural network (ANN) technology has been successfully applied in water quality prediction [5-7] and provides an effective tool to analysis and modeling nonlinear relationships in ecology [8]. Because ANN can imitate the basic characteristics of the human brain such as self-adaptability, self- organization and error tolerant [9, 10], it is able to map the nonlinear relationships. So in this study, the algal density of diatom and blue algae in Xiangxi Bay of Three Gorges Reservoir was simultaneously simulated using artificial neural network model. The proposed model may contribute to more efficient management in TGR.

EXPERIMENTAL SECTION

Sampling and Analysis

Samplings were performed monthly at 10 stations (X0-X9) in Xiangxi River (Fig. 1) from January 23 to December

20 in 2013 (no monitoring data in March, April and October). Water samples were collected at 0.5 m depth from surface in the middle of the river using a 5-L Niskin sampler (Hydrobios-Kiel). Water temperature (WT) was recorded in situ using multi-parameter water quality analyzer (Hydrolab DS5). Light intensity (LI) was monitored by light intensity meter. Flow velocity (v) was measured by current meter (Vector). Total nitrogen (TN), nitrate (NO_3), ammonium nitrogen (NH_4), total phosphates (TP), phosphate (PO_4), and silicate (SiO_4) were determined in the laboratory using State Environmental Protection Administration (SEPA) standard methods [11]. The algal density (AD) of diatom and blue algae were recorded with the algae counter (Algacount S300) [12].

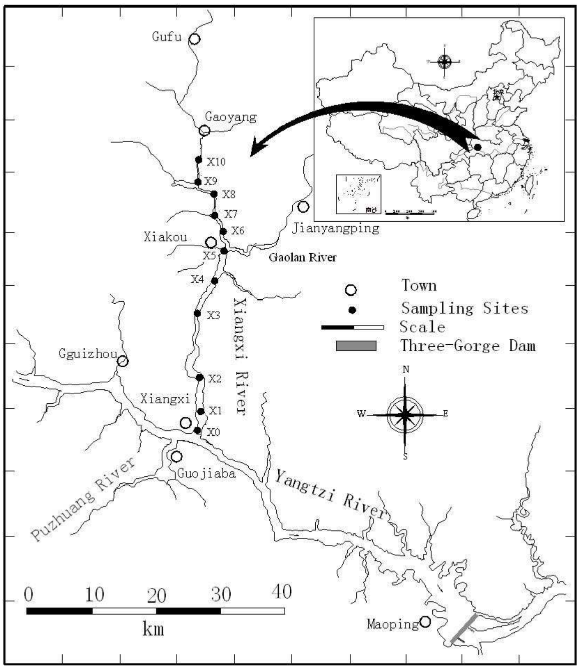


Fig.1 Sampling stations in Xiangxi Bay

Artificial Neural Network

The basic structure of an ANN model is usually comprised of three distinctive layers: the input layer, the hidden layer or layers, and the output layer. Each layer consists of one or more neuron(s). A neuron is a nonlinear algebraic function, parameterized with boundary values. The signal passing through the neuron is modified by weights and transfer functions. This process is repeated until the output layer is reached [13]. In this study, back propagation (BP) algorithm was used in ANN model. This training algorithm is a technique that helps distribute the error in order to arrive at minimum error. After the information has gone through the network in a forward direction and the network has predicted an output, the back propagation algorithm redistributes the error associated with this output back through the model, and weights are adjusted accordingly [14]. Minimization of the error is achieved through several iterations. One iteration of this algorithm can be written as

$$x_{k+1} = x_k - \alpha_k g_k \tag{1}$$

Where x_k is a vector of current weights and biases, g_k is the current gradient, α_k is the learning rate.

RESULTS AND DISCUSSION

Selection of Input Variables

In ANN model, input variables (independent variables) were selected from environmental factors using stepwise multiple linear regression (MLR) method, while algal density as the output variable (dependent variable). The algal density of diatom during the observation period in Xiangxi bay ranged from 0 to 8.3983×10^7 with mean value of 7.1798×10^6, while the algal density of blue algae ranged from 0 to 1.1640×10^8 with mean value of 6.9044×10^6 （Fig. 2）. The MLR model was obtained as follows:

$$AD\_diatom = 5.559\times10^6 + 6.906\times10^7 TP - 7.841\times10^7 PO_4$$
$$- 3.067\times10^6 SiO_4 + 6.069\times10^5 WT \tag{2}$$

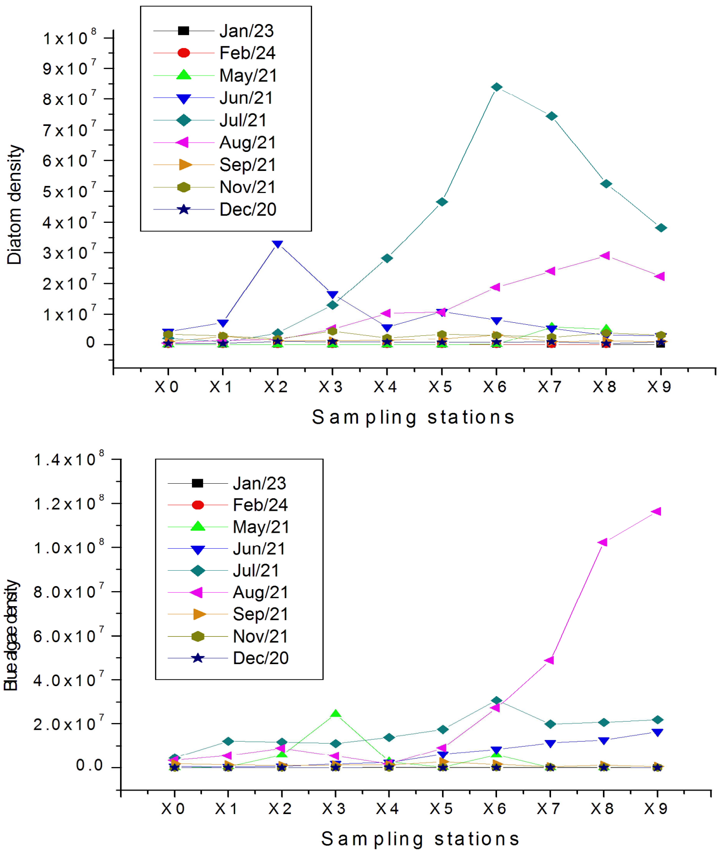
$(R = 0.667, R_{adj}^2 = 0.418, F = 16.648, P = 0.000, n = 88)$

$$AD\_bluea\lg ae = -2.289\times10^7 - 1.054\times10^7 NO_3 - 2.028\times10^7 NH_4 + 1.034\times10^8 TP + 1.695\times10^6 WT \quad (3)$$

$(R = 0.678, R_{adj}^2 = 0.433, F = 17.643, P = 0.000, n = 88)$

The statistical quality of the regression equation was examined using parameters such as the correlation coefficient (R), the squared adjusted correlation coefficient (R_{adj}^2), the Fisher ratio at the 95% confidence level (F).

Six important environmental factors (NO_3, NH_4, TP, PO_4, SiO_4 and WT) were selected as input variables (Table 1).



(A) (B)

Fig. 2 Algae density in Xiangxi bay: (A) Diatom density; (B) Blue algae density

Table 1 Mean Values of input factors during the observation period in Xiangxi Bay

	NO_3 (mg/L)	NH_4 (mg/L)	TP (mg/L)	PO_4 (mg/L)	SiO_4 (mg/L)	WT (℃)
Jan	1.316	0.201	0.133	0.048	6.729	13.57
Feb	0.851	0.424	0.180	0.139	5.019	12.29
May	0.748	0.272	0.174	0.065	4.667	20.77
Jun	1.588	0.497	0.097	0.031	5.032	25.85
Jul	1.328	0.532	0.087	0.012	0.794	30.93
Aug	0.764	0.235	0.086	0.026	5.467	29.47
Sep	1.029	0.574	0.051	0.009	6.789	28.18
Nov	1.772	0.027	0.050	0.016	4.626	20.75
Dec	1.377	0.038	0.112	0.036	7.833	17.83

Optimization of the ANN Structure

The resulting environmental factors decided by stepwise MLR were used for the input layer of ANN model. The number of neurons in the hidden layers depends on the problem. If the number of hidden neurons is small, the network may not have sufficient degrees of freedom to learn the process correctly. On the other hand, if the number is too high, the training will take a longer time and the network may over-fit the data [15]. To determine the performance of each ANN model, two different criteria were used: the squared correlation coefficient (R^2) of training set and test set. The training set (80%) and test set (20%) were divided randomly. In this study, the parameter values of the learning rate (0.1) and momentum (0.2), as well as maximum training epochs (1000) were used. Based on a four-layer network, using different numbers of hidden layer neurons, a number of ANN models were developed (Table 2). As a result, the ANN model ANN8, which has two hidden layers (number of neurons in the first hidden layer: 5; number of neurons in the second hidden layer: 4), was more predictive with the training set R^2 value of 0.738 (diatom) and 0.949 (blue algae), the test set R^2 value of 0.773 (diatom) and 0.870 (blue algae). While in MLR model the test set R^2 values are 0.392 (diatom) and 0.426 (blue algae). It can be concluded that the ANN modeled results are in good agreement with the measured values (Fig. 3).

Table 2 Features of different ANN models and MLR model

No.	ANN structure	Diatom		Blue algae	
		R^2(training)	R^2(test)	R^2(training)	R^2(test)
ANN1	6-4-1-2	0.734	0.760	0.945	0.830
ANN2	6-4-2-2	0.729	0.770	0.949	0.821
ANN3	6-4-3-2	0.746	0.758	0.946	0.802
ANN4	6-4-4-2	0.730	0.768	0.945	0.841
ANN5	6-5-1-2	0.732	0.763	0.936	0.832
ANN6	6-5-2-2	0.748	0.773	0.938	0.855
ANN7	6-5-3-2	0.744	0.767	0.938	0.817
ANN8	6-5-4-2	0.738	0.773	0.949	0.870
ANN9	6-5-5-2	0.739	0.767	0.949	0.822
MLR		0.454	0.392	0.564	0.426

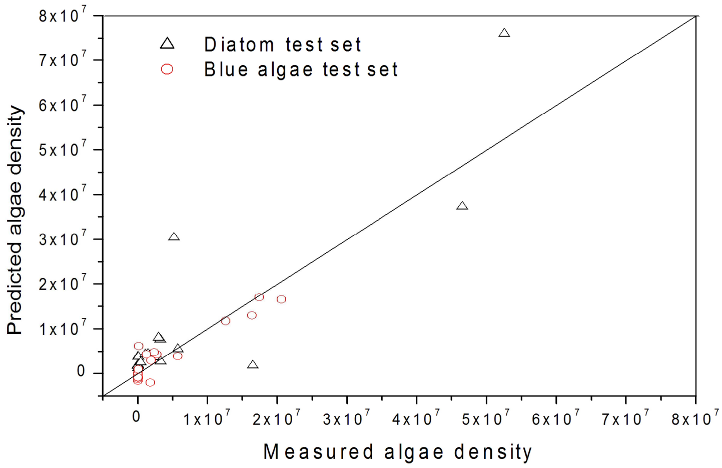


Fig. 3 Measured values versus predicted values for algae density using optimal ANN model

CONCLUSION

The artificial neural network model of various algae density in Xiangxi Bay of Three Gorges Reservoir was established. Using stepwise MLR method, the important environmental factors (NO_3, NH_4, TP, PO_4, SiO_4 and WT) were selected as input variables in ANN model. The ANN model has been shown to perform well for simulating the algae density of diatom and blue algae simultaneously, in which the training set R^2 values are 0.738 (diatom) and 0.949 (blue algae), the test set R^2 values are 0.773 (diatom) and 0.870 (blue algae), respectively.

Acknowledgments
This work was granted by Science Foundation of China Three Gorges University (No. 1112046) and National Natural Science Foundation of China (No. 51179095).

REFERENCES

[1] JG Wu; JH Huang; XG Han; ZQ Xie; XM Gao. *Science*, **2003**, 300, 1239–1240.
[2] T Tang; DF Li; WB Pan; XD Qu; QH Cai. *Chin. J Appl. Ecol.*, **2004**, 15(1), 141–144.
[3] HY Wang. *Resour. Environ. Yangtze Basin*, **2005**, 14(2), 233–237.
[4] HJ Wu; ZY Lin; SL Guo. *Resour. Environ. Yangtze Basin*, **2000**, 9(9), 237–241.
[5] CW Wen; CS Lee. *Water Resour. Res*., **1998**, 34, 427–436.

[6] Y Kuo; C Liu; KH Lin. *Water Res.*, **2004**, 38, 148–158.
[7] J Kuo; M Hsieh; W Lung; N She. *Ecol. Model*., **2007**, 200, 171–177.
[8] S Lek; M Delacoste; P Baran; I Dimopoulos; J Lauga; S Aulagnier. *Ecol. Model*., **1996**, 90, 39–52.
[9] ZG Niu; HW Zhang; HB Liu. *J. Tianjin Polytechnic Univ*., **2006**, 25, 89–92.
[10] J Shu. *North Environ*., **2006**, 31, 44–46.
[11] XC Jin; QY Tu. *Criterion of eutrophiction survey on lakes*, 2nd ed, Beijing: Environmental Sci Press, **1990**.
[12] AJ Lewitus; ET Koepfler; JT Morris. *Limnol Oceanogr*, **1998**, 43, 636–646.
[13] RS Govindaraju. *J. Hydrol. Eng*., **2000**, 5, 124–137.
[14] KP Singh; A Basant; A Malik; G Jain. *Ecol. Model.*, **2009**, 220, 888–895.
[15] N Karunanithi; WJ Grenney; D Whitley; K Bovee. *J. Comput. Civil Eng*., **1994**, 8, 210–220.