ISSN : 0974 - 7435 Volume 10 Issue 20

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(20), 2014 [11847-11851]

Application of BP network in calculation of non-point source pollution

Liang dangling[1] Dong Wencai[1], Wang Kexin[2], Cao Chengpeng[3], Lichong[4], Fu Qiang*

1,2,3,*: College of Water Conservancy & Architecture, Northeast Agricultural University, Harbin, Heilongjiang 150030, (CHINA)

[4]HLJ Post Yi Tong Info-net Co., Ltd., 150030, Harbin, (CHINA)

E-mail : fuqiang@neau.edu.com

ABSTRACT

The formative process of non-point source pollution (NSP) is complicated, and can be hardly simulated with a traditional mathematical method or determined with overseas existing model parameters. This problem can be solved by calculation with the BP network. In the paper, the BP network is applied in simulation and calculation of the total N load in NSP of Heilongjiang. The application indicates that the BP network is feasible.

KEYWORDS

NSP; BP network; Load.

© Trade Science Inc.

INTRODUCTION

The artificial neural networks system is a machine designed to simulate the work mode of human brain. As a parallel distributed processor with numerous connections, it obtains knowledge and problem-solving capability by learning. The BP network is a type of artificial neural network, also called "back propagation network". NSP is influenced by multiple factors, and its formative process is influenced by climate, geology, soil type, land-use type, vegetation coverage and precipitation process. Required parameters can be hardly determined with the models AGNPS and SWAT, whose applicability is poor. In an artificial neural network model, what is needed is no more than sample training, and samples can be acquired by measurement without need of considering parameters. Although the formative mechanism of NSP is complicated, there's a specific relation of rigorousness between its factors and pollutants. The neural network can approximately simulate such relation, followed by calculation, which offers a high accuracy. Given there's no measured data about output layer-total N in the studied region, in the paper, output layer trained samples were calculated with the output factor method (Jones output factor model). In the calculation, cultivated land, natural land, urban land, population, large domestic animal, pig, sheep and poultry were taken as output layers. The output factor method was put forward during the study on land use-nutrient load-lake eutrophication relation in USA and Canada, at the beginning of the 1970s. [1]

DESIGN STEPS OF BP ALGORITHM

(1)Initialization: assign random numbers to w_{ij} and T_{li} of each connection weight and θ_i and θ_l of each threshold, from the section (-1, +1);

(2) Randomly select a mode pair x_j, t_l, offer it to the network;

(3) Calculate the input s_i of each unit on the intermediate layer with the connection weight w_{ij} and the threshold θ_j of the input mode x_j, calculate the output y_i of each unit on the intermediate layer (hidden layer) with the function S, as the formula below:

$$y_i = f\left(s_i\right) = f\left(\sum_j w_{ij} X_j - \theta_i\right)$$

(4) Calculate the input L_l oof each unit on the output layer with the output y_i, the connection weight T_{li} and the threshold θ_l of the hidden layer, calculate the output o_l of each unit on the output layer with the function S, as the formula below;

$$o_l = f\left(L_l\right) = f\left(\sum_i T_{li} y_i - \theta_l\right)$$

(5) Calculate the generalized error δ_l of each unit on the output layer with the expected output mode t_l and the actual network output o_l, as the formula below;

$$\delta_l = \left(t_l - o_l\right) \times o_l \times \left(1 - o_l\right)$$

(6) Calculate the generalized error δ_i' of each unit on the intermediate layer with the connection weight T_{li}, the generalized error of output layer e_k and the output of intermediate layer y_i, as the formula below;

$$\delta' = y_i \left(1 - y_i\right) \sum_l \delta_l T_{li}$$

(7) Correct the connection weight T_{li} and the threshold θ_l with the generalized error δ_i of each unit on the output layer and the output y_i of each unit on the intermediate layer, as the formulas below;

$$T_{li}\left(k + l\right) = T_{li}\left(k\right) + \eta \delta_i y_i$$

$$\theta_l\left(k + l\right) = \theta_l\left(k\right) + \eta \delta_l$$

k : number of iterations

(8) Correct the connection weight w_{ij} and the threshold θ_i with the generalized error δ_i' of each unit on the intermediate layer and the input x_j of each unit on the input layer, as the formulas below; $w_{ij}(k+l) = w_{ij}(k) + \eta'\delta_i' x_j$

$$\theta_i(k+l) = \theta_i(k) + \eta'\delta_i'$$

(9) Randomly select the next learning mode pair, offer it to the network, return to the step 3 until all sample mode pairs are trained;

(10) Randomly re-select a mode pair from original mode samples, return to the step 3 until the global error function E gets lower than the preset minimum (namely the network converges) or the times of learning are higher than the present value (namely the network can't converge);

CASE OF APPLICATION

(1) Take the total N in NSP load of Heilongjiang for example. Given there's no measured data about the total N in the studied region, output layer data of trained samples (i.e. the total N) was calculated with the output factor method. The data of 2002~2011 was taken as the trained sample, and the data of 2012 and 2013 was compared. The basic data is shown in TABLE 1, extracted from *Statistical Yearbook of Heilongjiang.* [2] [3]

TABLE 1 : Basic Information about Land Area, Population, Livestock and Poultry in the Studied Region

Year	Cultivated land (×104 /hm^2)	Natural land (×104/hm^2)	Urban land (×10^4/hm^2)	Fertilizer (t×10^4)	Population (×10^4)	Large domestic animal (×10^4 pieces)	Pig (×10^4 pieces)	Sheep (×10^4 pieces)	Poultry (×10^4 pieces)
2002	880	3139.9	133.3	88.5	3526.2	406	763.2	308.6	10300
2003	925.3	3134	129.7	100.2	3640.0	441	793.1	327.5	11284.4
2004	927.5	3133.9	133.8	108.5	3672.0	519	909.4	399.7	13302.5
2005	931.4	3134.5	134.2	108.9	3701.0	632.6	1140.3	494.6	16500
2006	934.2	3211.5	134.7	115.1	3728.0	742.2	1281.2	567.5	18000
2007	934.6	3214.3	135.07	121.8	3751.0	650.1	1325.2	627.3	23500
2008	932	3216.9	135.38	125.8	3762.0	438.1	958.1	462.8	12088
2009	932	3211.5	135.96	127.3	3782.5	550.3	1006.3	483.5	12878
2010	961.7	3217.8	114.3	121.6	3807	547.7	1085.4	507.4	13144.4
2011	960.1	3264.6	114.54	123.2	3714.7	558.3	1123	567.8	13739.4
2012	951.2	3222.4	114.43	129.7	3732.6	603.4	1163	749.1	14783.4
2013	969	3313.3	114.64	125.7	3723.5	690.3	1326.4	1029.5	15987.7

Jones output factor model:

$$L = \sum_{i=1}^{n} E_i \left[A_i(I_i) \right] + p$$

L: loss of nutrient; E_i: output factor of the i th nutrient source; A_i: area of the i th type of land, quantity of the i th type of livestock, poultry or population; I_i: output of the i th type of nutrient source; p: nutrient input of precipitation;

E_i refers to nutrient output rates of different land use types in the basin. For livestock, it stands for the percent of excrement directly into the receiving water body with volatilization of ammonia during manure collection and manure storage considered. For human, it reflects use of detergent by local people, nutrition of local diet and treatment of domestic sewage, calculated as the formula below.

$$E_i = D_{ca} \times H \times 365 \times M \times B \times R_S \times C$$

E_i: annual N and P output of population (kg/a); D_{ca}: daily nutrient output per capita (kg/d); H: population in the basin; M: factor of mechanical nutrient removal during pollution treatment; B: factor of biological nutrient removal during sewage treatment; R_S: nutrient retention factor; C: dephosphorization factor;

The nutrient input P produced by precipitation can be indicated as,

$$P = caQ$$

c: nutrient concentration in rainwater (g/m^3); a: annual precipitation (m^3); Q: runoff coefficient (given the precipitation output factor is undetermined, the pollution caused by precipitation is not considered in the paper); [5]All of output Factors of Nutrient Sources are shown in TABLE 2.

TABLE 2 : Output Factors of Nutrient Sources

Type	**Natural land (kg/hm^2·/a)**	**Cultivated land (kg/hm^2/a)**	**Urban land (kg/hm^2/a)**	**Livestock and poultry (kg/hm^2/a)**				**Human (kg/hm^2/a)**
				Large domestic animal	**Pig**	**Sheep**	**Poultry**	
Quant.	3.76	13.49	9	10.21	0.74	0.4	0.04	2.14

The calculated results of the total N load in NSP of the basin are shown in TABLE 3.

TABLE 3 : Total N Load in NSP of the Basin, 2002~2013 (unit: t)

Year	**Cultivated land**	**Urban land**	**Natural land**	**Human**	**Livestock and poultry**
2002	118712	11997	118061.4	75460.68	70101.73
2003	12483	11673	117838.4	77896	70079.99
2004	125119.8	12042	117834.6	78580.8	58992.15
2005	125645.9	12078	117857.2	79201.4	85290.89
2006	126023.6	12123	120752.4	79779.2	94729.5
2007	126073.5	12159	120857.3	80271.4	84405.08
2008	125720.1	12184.2	120958.1	80506.8	66153.46
2009	125726.8	12212.1	120752.4	80956.2	57356.24
2010	129733.3	10287	120989.3	81469.8	53592.44
2011	129517.5	10308.6	122749	79501	73079.59
2012	128316.9	10298.7	121162.2	79864.8	79123.1
2013	130718.1	10317.6	124580.1	79693.6	90806.07

(2) The basic information about land area, precipitation, population, livestock and poultry of the studied region from 2002 to 2013 was taken as the input sample, and the total N load in NSP of the basin from 2002 to 2013 was taken as the output sample, for training. The data of 2012 and 2013 was verified. The results are shown in TABLE 4.

TABLE 4 : Calculated Results

Year	**2012**			**2013**		
Type	**Output factor method (t)**	**BP network (t)**	**Error (%)**	**Output factor method (t)**	**BP network (t)**	**Error (%)**
Cultivated land	128316.9	125160	2.5	130718.1	125230	4
Natural land	121162.2	119850	1	124580.1	119930	3.7
Human	79864.8	79380	0.6	79693.6	79430	0.3
Livestock and poultry	79123.1	71320	9.9	90806.07	72060	20.6
Urban land	10298.7	11710	13.7	10317.6	11580	12.2
Total	418765.7	407420	2.7	436115.5	408230	6.4

COMPARATIVE ANALYSIS

Learnt from the comparison, the accuracy of urban land expectation of 2012 and that of livestock & poultry expectation of 2013 are not high. Seen from the formative factors of NSP, the low accuracy of urban land expectation of 2012 is attributed to significant change in land area from 2009 to 2010, and the low accuracy of livestock & poultry expectation of 2013 is attributed to significant change in quantity of livestock & poultry from 2012 to 2013. Another important cause for the low accuracies is the small quantity of trained samples so that the BP network can’t fully simulate output process of pollution load. If there is a large quantity of trained samples, the expectation accuracy can be improved. On the whole, the expected values basically coincide with the calculated values. Given this, the BP network can be used to calculate NSP load. [6] [7]

CONCLUSION

In the paper, applying the BP network in calculating NSP load simplifies the calculation, and avoids difficulties in parameter determination with the models AGNPS and SWAT. BOD, COD, TN, TP and SS can be simultaneously taken as output layers. [8]The application also has a shortcoming, namely a large quantity of samples are needed for training. If there are sufficient samples for training, the BP network will be effective for calculation of NSP.

REFERENCES

[1] Fu Qiang, Liu Jianyu, Wang Likun, Feng Jiang; Based on ANN Model to Forecast the Groundwater Level on the Area of Well Irrigation Rice, Journal of Northeast Agricultural University, **2**, 152-159 **(2002.06.33).**

[2] Cai Ming, Li Huaien, Zhuang Yongtao, Wang Qinghua; Application of Modified Export Coefficient Method in Polluting Load Estimation of Non-point Source Pollution, Journal of Hydraulic Engineering, **(2004.07).**

[3] Heilongjiang General Soil Survey Office, Heilongjiang Land Administration, Soil of Heilongjiang, China Agriculture Press, **(2001.12).**

[4] Yan Pingfan, Zhang Changshui; Artificial Neutral Networks and Evolutionary Computing, Tsinghua University Press, **(2000.11).**

[5] Liang Bo, Wang Xiao Yan, Cao Liping; Water Environment Non- point Source Pollution Loading Estimation Methods in China, Journal of Jilin Normal University, **(2004.08).**

[6] Zhang Shuilong, Zhuang Jiping; Study on Distributed Model of Non-point Sources Pollution in Agriculture on Watershed Scale, Journal of Arid Land Resources & Environment.

[7] Zhang Keqiang, Gao Huaiyou; Treatment and Disposal of Pollutants from Livestock and Poultry Industry, Chemical Industry Press.

[8] Tang Keli, et.a., Soil and Water Conservation in China, Science Press, **(2004.04)**.