Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2013, 5(9):355-361

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Age distributions of two tree species by simulation of BP neural network system

Mengtao Zhang[1], Xingang Kang[1*], Shuo Cai[1] and Lixin Zhang[2]

1. Corresponding author *Beijing Forestry University, Beijing, P. R. of China kxg 520512@163.com 0086-010-62338133*
[2]Yunnan Forest Institute, Kunming, P. R. of China

ABSTRACT

In order to investigate the relationship between tree age and the correlated tree's DBH (diameter at breast height) we simulated it with the method of artificial neural network, in which a BP (i.e. back propagation) was applied with the data from forest inventory measured in 1986-1988. The trees were two dominant tree species in northeastern China, Pinus koraiensis Sieb. et Zucc. and Abies nephrolepis (Trautv.) Maxim. The simulation was undertaken by taking the matrix of trees' DBH as input vector and their correlated ages as output vector. The results showed that the accuracy from the model was as high as 0.9, indicating the method of BP neural network might be supplemental method to traditional one that making use of stand growth of trees.

Keywords: artificial neural network; back propagation; age distribution; simulation

---

INTRODUCTION

The age of trees in forest is an important parameter in evaluating the quality of natural forests and the stages of their succession and notably, in decision making in forest management (Bondarev, 1997; Haugo and Halpern, 2010; Karenlampi, 2011). In the study of forest stand modeling, it is essential so many studies have focused on the estimation of age distribution of tree or tree age structure to explore the role of it in forest ecosystem (Kalliovirta and Tokola, 2005; Garet *et al.*, 2012).

Natural forests are intrinsically nonlinear in its growth and substantial complexity as shown in the relationship of tree age and DBH (Kang, 2011). In the fitting of models, some conditions may be ignored or simplified due to technical difficulties in obtaining data; therefore the models were limited in describing this nonlinear relationship in tree age and DBH (Sui, 1994; Wang *et al.*, 2004). To an extent, an artificial neural network (ANN) was able to evade such difficulties and provide a simple learning algorithm for it. ANN is ubiquitously featured with characterization of self-organization, self-adapting and self-learning, which have been used widely in many areas. It is noteworthy that ANN handles small sample well. In forest research, Guan and Gertner used ANN by simulating tree mortality with logistic regression and exponential functions then comparing it to the actual data obtained empirically. The results showed satisfactory matched (Guan and Gertner, 1991). Cai et.al researched diameter distributions of three broadleaved tree species, who used back propagation neural network, and the results showed the precision of simulations higher than 99% (Cai *et al.*, 2012). We have found few studies in exploring the relationship of tree age and DBH with the method of ANN. In this paper we used data from two dominant tree species of mixed forest of *Pinus koraiensis Sieb. et Zucc.* and *Abies nephrolepis (Trautv.) Maxim.* to exemplify with it.

ARTIFICIAL NEURAL NETWORKS AND BACK PROPAGATION

Artificial neural network is a mathematical model of information processing, which is similar to the structure of the synapses of the brain, and it is constituted with a large number of interconnected nodes (or neurons) (Chaoui *et al.*, 2009; Ghosal and Chaki, 2010). Each node represents a specific output function, known as the activation function. Connection between two nodes represents a weighted value of the connected signal, which is equivalent to the memory of the artificial neural network (Sivanandam, 2009). Currently the most widely used ANN model is the BP (back Propagation) model. The BP structure is shown in Figure1.

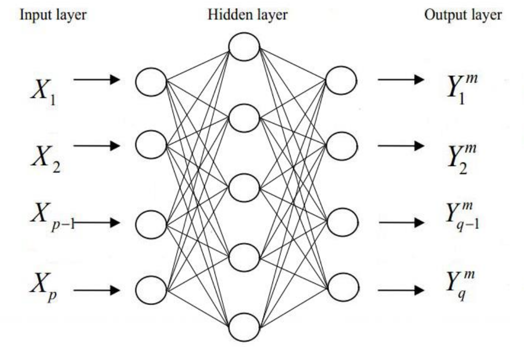

Fig. 1 Structure of the back propagation (BP) neural network

Theory has been proved a three-tier model of BP network to achieve any continuous map; it is by changing the connection points on the unit to connect the weight to achieve the mapping (Cai *et al.*, 2010). BP model is characterized by one-way signal transmission from the input layer to output layer, with a layer of neurons to transmit information between each other, each neuron with adjacent layers connected to all neurons, the connection weights use symbols W_{ij} to represent. The role of neuronal function is Sigmoid function, set the input layer with *P* nodes, the output layer has one node, *k*-1 level of any node with *i* express, *k* layer of any node with *j* express, *k*+1 layer of any node with l express. W_{ij} is the weight which connected *k*-1 layer neurons *i* to *k* layers neurons *j*. *k*-1 layer node *i* output is $Q_{(k-1)i}$.

Both input and output of node *j* in the *k* layer are expressed by:

$$NET_{kj} = \sum_i W_{ij} Q_{(k-1)i} \tag{1}$$

$$Q_{kj} = f(NET_{kj}) \tag{2}$$

In the forward propagation process, the information is transmitted to the input layer at first, and then it is calculated in the hidden layer, after that, the processed information is exported in the output layer. In this process, each layer neurons affects only the state of the next layer of neurons. If we cannot get the desired output information in the output layer, the error variation value is calculated in the output layer, and then the error information turned to reverse spread, Error signal through the network back propagation along the connection path to modify the weight of layers of neurons until they reach the desired goals.

MATERIALS AND METHODS

Study area and collection of materials: The experiment was carried out in Jin Gouling forest farm located in Changbai Mountain in northeastern PR China (43˚22′ N, 130˚10′ E). The tree species were *Pinus koraiensis Sieb. et Zucc.* and *Abies nephrolepis (Trautv.) Maxim.*, and the number of them were 82 and 164, respectively, which were cutting down from 1986 to 1988. The tree age was determined according to the tree ring. In addition, DBH, tree height and crown width were also measured (Table 1).

Table 1 The characteristics of two dominant tree species of natural spruce-fir mixed forest

Species	No. of trees	Age			Standard deviation	Variation coefficient
		Minimum	Maximum	Mean		
P. koraiensis	82	17	214	80.74	46.81	57.97%
A. nephrolepis	164	26	147	67.77	28.62	42.23%

Model construction: BP neural network is a multi-layer feed forward neural network, the transfer function of the neurons is S-shaped function, and the output is continuous quantity between 0 and 1. BP neural network can achieve arbitrary nonlinear mapping from input to output. Nevertheless, some shortcomings exist in BP neural network, such as slow convergence and local minima. In addition, it is difficult to determine the number of nodes in the hidden layer (Wilamowski and Yu, 2010). We attempted to use training algorithm based on numerical optimization theory, which was called the Levenberg-Marquardt training method of BP network.

In order to do simulation to obtain an ideal age distribution, the matrix of DBH (D) was taken as the input vector, and tree age (A) was the output vector. Therefore, BP network structure built for 1: S: 1, where the leftmost "1" referred to the number of nodes of the input layer, while the rightmost "1" referred to the number of nodes of the output layer, and the middle of "S" was the number of hidden nodes. In general, the number of nodes in the hidden layer is essential to BP network. If the number of nodes is too few, the content of learning is limited, and this leads to all rules of the training sample cannot be stored enough; by contrast, overmuch nodes will not only increase network training time, but also store irregular content of samples, this will reduce the generalization ability (Wilamowski, 2009). In this paper, an empirical formula was used to calculate the value interval, and then the same sample was trained and identified network minimum error corresponding to the hidden layer nodes. The empirical formula is as followed:

$$m = \sqrt{n+1} + a \tag{3}$$

Where m is number of nodes in hidden layer, n is the input node, 1 is the output node, and a is regulation constant, which in the range between 1-10. Though several simulations, 10 was chosen to the optimum number of nodes in hidden layer. The structure of BP network, therefore, was 1:10:1 (Fig.2), and entitled *A_net*.

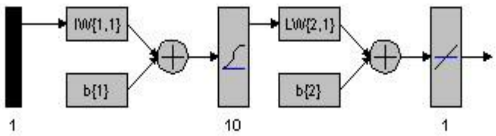

Fig.2 BP neural network model of *Pinus* and *Abies*

In the Figure 2, "S" symbol of input layer box indicated that the transfer function of the hidden layer neurons was an asymmetric sigmoid function, which meant the *log*sig function; "/" symbol of output layer box indicated that the neurons transfer function was a linear function, which meant the *purelin* function. The formula of *log*sig function is:

$$f(x) = \frac{1}{1+e^{-x}} \tag{4}$$

BP neural network model which showed in Fig. 2 was expressed in the transfer function of MATLAB as the following formulas:

$$H_i = \log sig(\sum_{i=1}^{s} W_{i,1}^1 D + b_i^1) \tag{5}$$

$$A = purelin(\sum_{i=1}^{s} W_{1,i}^2 H_i + b^2) \tag{6}$$

Where $W_{i,1}^1$ is the connection weight between *i*th neuron of the hidden layer and the input node, i =1, 2…*S*; the superscript "1" is the first neuron of the hidden layer; $W_{1,i}^2$ is the connection weight between *i*th neuron of the hidden layer and the neuron of output layer; the superscript "2" is the second neuron of the output layer; b_i^1 is threshold value of the *i*th neuron; b^2 is threshold value of the output neuron; H_i is output transference of the *i*th neuron of the hidden layer, *i*=1, 2, …S.

Model training: For any artificial neural networks, pretreatment is important to training sample. It can not only affect the network training time, even network paralysis may happen if mishandle the training sample. In order to improve the efficiency of network training, input and output variables were normalized in this paper and the formula is:

$$V' = \frac{(V_i - V_{\min})}{(V_{\max} - V_{\min})} \tag{7}$$

Where V' is the normalized value; V_i is the actual value; $V_{\max}$ and $V_{\min}$ are maximum and minimum values, respectively. In this paper, both DBH and tree age were normalized. In addition, each input weight is indicated by *IW* {1, 1}; the layer weight is indicated by *LW* {2, 1}; the bias vector *b* {1} indicated the threshold matrix of the hidden layer; *b* {2} indicated the threshold matrix of the output layer. The formula was as followed and vector values showed in Table 2.

$$IW\{1,1\} = \begin{Bmatrix} w_{1,1}^1 \\ w_{2,1}^1 \\ w_{3,1}^1 \\ \vdots \\ w_{9,1}^1 \\ w_{10,1}^1 \end{Bmatrix}, b\{1\} = \begin{Bmatrix} b_1^1 \\ b_2^1 \\ b_3^1 \\ \vdots \\ b_9^1 \\ b_{10}^1 \end{Bmatrix}, b\{2\} = \{b_1^2\}, \quad LW\{2,1\} = \{w_{1,1}^2, w_{1,2}^2, w_{1,3}^2, \cdots, w_{1,9}^2, w_{1,10}^2\} \tag{8}$$

Table 2 Simulation results of the age distribution of two dominant species

Pinus koraiensis				*Abies nephrolepis*			
IW{1,1}	LW{2,1}	b{1}	b{2}	IW{1,1}	LW{2,1}	b{1}	b{2}
65.6731	10.5121	-59.5609	-1.0656	-47.9789	-5.7211	65.166	0.3366
55.0978	2.8732	-52.493		37.9443	-0.1122	-38.2913	
-57.0233	-1.1148	47.3582		57.1979	0.1727	-41.6794	
45.5663	-14.0947	-41.4728		48.3911	-6.0757	-54.3225	
55.7563	0.0725	-32.9191		163.9359	0.1403	-95.1966	
-56.3566	-0.1956	25.6494		152.8794	0.0971	-72.3685	
-56.188	0.0168	18.2065		-106.3078	-0.118	37.6741	
57.0358	0.0577	-14.4557		216.4473	0.0485	-52.045	
58.1871	0.0667	-5.4352		-305.9256	-0.0271	36.6885	
56.6782	1.0115	3.9984		190.9316	7.0028	5.61	

RESULTS

Analysis of simulation and precision: The training error curve of age distributions of two dominant tree species was observed in Fig.3. We could find that the error was closed to a stable point with epoch increased. After the neural network training, the corresponding theoretical output value could be fitted in the MATLAB SIM function, and the formula was as follow:

$$A' = SIM(A\_net, [D]) \tag{9}$$

Where A_net is trained network object, and it defines and stores all network parameters of the threshold and weight. The formula (9) can operate in MATLAB; A' is the normalized age value, while the normalized result should be converting to the fitted value of tree age. The formula is as follow:

$$A'' = A_{\min} + A'(A_{\max} - A_{\min}) \tag{10}$$

Where A'' is the anti-normalized value of tree age.

In addition, mean squared error (MSE) were used for precision analysis:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \tag{11}$$

Where $\hat{Y}$ (A'') is the fitted value; Y_i is actual value; n is the number of sample. In this paper, the fitted value and the actual value can be given directly by the MATLAB, which named the error matrix:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(net\_errors)^2 \quad (12)$$

After calculation, results were showed that: MSE were 0.00389 and 0.00875, and the fitted precision were 99.6% and 99.1%, respectively (Fig.3).

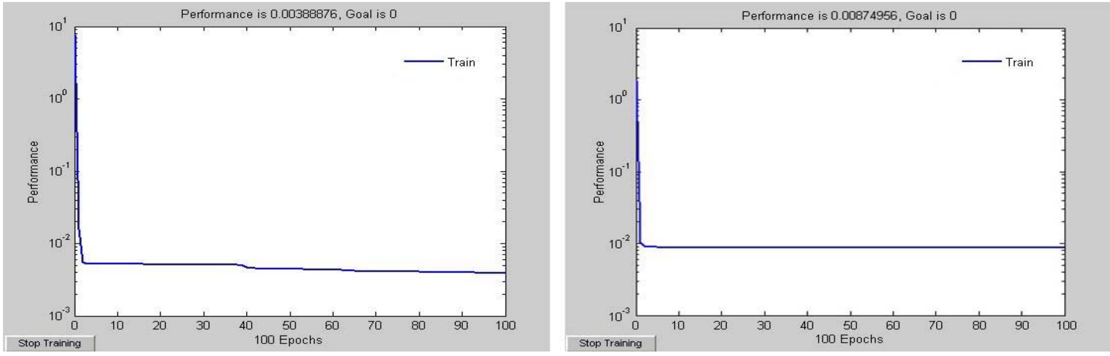

Fig.3 Error curves of training of the age distributions of *Pinus* and *Abies*.

Through self-learning of the network, the network output results obtained from Eq. (1) and (2) were given in Table 3. The simulation results indicated that the degree of agreement was with satisfactory.

Table 3 Simulation results of the age distribution of two dominant tree species

Age Class	*Pinus koraiensis*			*Abies nephrolepis*		
	No. of trees	Normalization Value	Training Value	No. of trees	Normalization Value	Training Value
0-20	2	0.033	0.051	-	-	-
20-40	18	0.157	0.105	33	0.056	0.084
40-60	19	0.269	0.153	48	0.200	0.218
60-80	4	0.445	0.261	24	0.359	0.390
80-100	16	0.657	0.402	26	0.525	0.512
100-120	6	0.714	0.438	29	0.667	0.623
120-140	8	0.836	0.605	3	0.845	0.654
140-160	4	0.927	0.741	1	0.99	0.672
160-180	2	0.907	0.676	-	-	-
180-200	1	0.929	0.769	-	-	-
200-220	2	0.992	0.825	-	-	-

Regression analysis: In order to test the correlation between predicted values and the measured values of the age distributions, the linear regression was used.

$$Y = aX + b \quad (13)$$

The accuracy of the model could be analyzed by parameter hypothesis (a=1, b=0) testing of the regression equation. In MATLAB, POSTREG function could be used for linear regression analysis of the measured values and the fitted values of the network model. POSTREG function call format was as follows:

$$[a, b, r] = POSTREG(Y, Y) \quad (14)$$

Where *a*, *b* are parameters of the linear regression equation, *r* is the correlation coefficient, $\hat{Y}$ is the fitted value; Y_i is actual value.

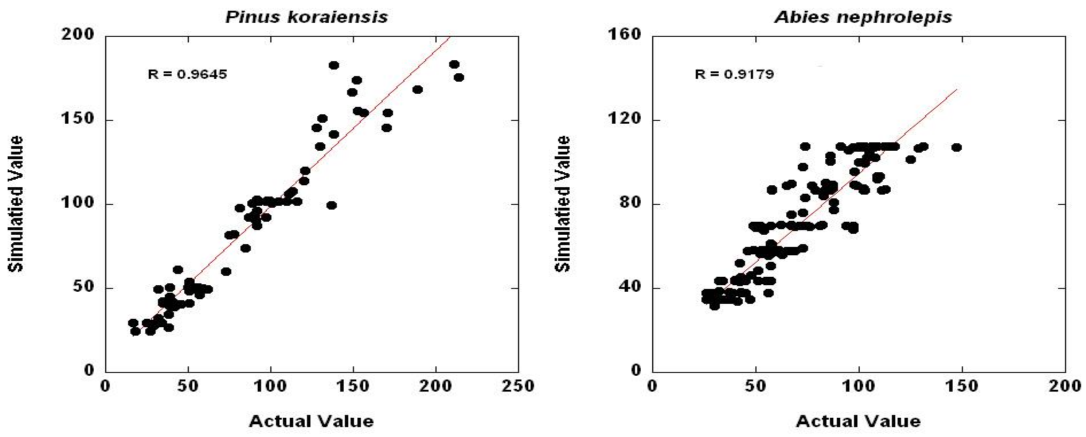

Fig.4 Regression analysis of age distributions of two dominant tree species

Table 4 Linear regression Functions of age distributions of two dominant tree species

Species	a	b	R	Function
Pinus koraiensis	5.6731	0.9302	0.9645	*Y*=5.6731*X*+0.930
Abies nephrolepis	10.664	0.8427	0.9179	*Y*=10.664*X*+0.842

According to linear regression between the fitted values and the actual values of the age distributions (Figure 4 and Table 4), we found that determination coefficients (R) were 0.9645 and 0.9179, respectively, which indicated the design tree age distribution of BP artificial neural network model was satisfied.

CONCLUSION

The major objective of our study was to simulate age distributions of two dominant tree species (*Pinus koraiensis* and *Abies nephrolepis*) of the mixed forest in Changbai Mountain, with BP neural network. We concluded that: the accuracy of simulated models was above 0.9, and the results indicated that artificial neural network might act as a good supplement for the traditional stand growth modeling methods. In order to achieve more accurate and convenient forecasts, more forest factors and management elements should be incorporated into artificial neural network in future study.

Acknowledgement
This study is supported by the planning subject of "the twelfth five-year-plan" in national science and technology for the rural development in China (NO. 2012BAD22B0203).

REFERENCES

[1] Bondarev, A., **1997**. *Forest Ecol Manag,* 93(3): 205-214.
[2] Cai, S., X.G. Kang, Z.W. Gong, P.T. Chen and L. Qin, **2010**. A model for forest resources management based on artificial neural network. In: Computer Application and System Modeling (ICCASM), 2010 International Conference on. IEEE: pp: V3-177-V173-182.
[3] Cai, S., X.G. Kang and L.X. Zhang, **2012**. *International Journal of Advancements in Computing Technology,* 4(15): 520-527.
[4] Chaoui, H., P. Sicard and W. Gueaieb, **2009**. *IEEE Transactions on,* 56(8): 3174-3187.
[5] Garet, J., F. Raulier, D. Pothier and S.G. Cumming, **2012**. *Ecol Indic,* 23: 202-210.
[6] Ghosal, S. and S. Chaki, **2010**. *The International Journal of Advanced Manufacturing Technology,* 47(9-12): 1149-1157.
[7] Guan, B.T. and G. Gertner, **1991**. *Forest Sci,* 37(3): 871-885.
[8] Haugo, R.D. and C.B. Halpern, **2010**. *Botany,* 88(5): 488-499.
[9] Kalliovirta, J. and T. Tokola, **2005**. *Silva Fennica,* 39(2): 227-248.
[10] Kang, X.G., **2011**. Forestry manegment (fourth ed). Beijing, China Forestry Press.
[11] Karenlampi, P.P., **2011**. *Journal of theoretical biology,* 270(1): 13-18.
[12] Sivanandam, M., **2009**. Introduction to artificial neural networks, 1e. Vikas Publishing House.
[13] Sui, D.Z., **1994**. *Canadian journal of remote sensing,* 20(4): 368-380.
[14] Wang, T., Y. Liang, H.B. Ren, D. Yu, J. Ni and K.P. Ma, **2004**. *Forest Ecol Manag,* 196(2-3): 267-274.
[15] Wilamowski, B.M., **2009**. *Industrial Electronics Magazine,* IEEE, 3(4): 56-63.

[16] Wilamowski, B.M. and H. Yu, **2010**. *Neural Networks, IEEE Transactions on*, 21(6): 930-937.