African Journal of Agricultural Research Vol. 6(27), pp. 5817-5821, 19 November, 2011
Available online at http://www.academicjournals.org/AJAR
DOI: 10.5897/AJAR10.859
ISSN 1991-637X 2011 Academic Journals

Full Length Research Paper

Predicting photosynthetic rate of sunflowers using back propagation neural network based on uniform design

W. Y. Chen[1,2], J. Yang[1]*, Z. Y. Chen[1], F. Y. Luo[1], C. L. Wu[1] and Z. S. Peng[1]

[1]Key Laboratory of Southwest China Wildlife Resources Conservation ,China West Normal University, Nanchong, Sichuan 637009, People's Republic of China.
[2]Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan province, 610041, People's Republic of China.

Accepted 12 May, 2011

The new model to predict photosynthetic rate (Pn) using back propagation neural network (BPNN) based on uniform design (UD) was studied. Four parameters of BPNN were designed at six levels individually by UD experiment to optimize the architecture of the BPNN model. The optimal parameters were used to construct an intelligent, feasible BPNN model which could more accurately predict the photosynthetic rate of sunflowers response to environmental factors. The constructed BPNN model had three layers namely input layer, hidden layer with nine neurons and an output layer. Four environment factors including photosynthetic active radiation (PAR), temperature (T), carbon dioxide level (CO_2) and relative humidity (RH) were input layers, and photosynthetic rate (Pn) as an output layer. Results showed that the predicted values and actual values of Pn fitted very well, with mean absolute percentage error (MAPE) of 3%, mean square error (MSE) of 0.75 μmol CO_2 $m^{-2}s^{-1}$ and mean absolute error (MAE) of 0.72 μmolCO_2 $m^{-2}s^{-1}$. There was no significant difference using significant test between the actual values obtained from portable photosynthetic system and predicted value calculated by models. The conclusion was that the model established by BPNN based on UD was more accurate than stepwise regression to predict Pn of sunflowers giving the environmental factors (PAR, T, CO_2 and RH).

Key words: Back propagation neural network (BPNN), model, photosynthetic rate (Pn), sunflower, uniform design (UD), stepwise regression.

INTRODUCTION

Sunflower (*Helianthus annuus* L.), belonging to Asteraceae family is one of the major oil seed crops, as a foodstuff and also known as an ornamental plant (David et al., 2008). The success of wide plantations depends on their productivity determined by the interaction of physiological factors with the environment and physiological processes, such as photosynthesis (Saraswathi and Paliwal, 2008). Considering the time, labor and costly-consuming factors to monitor the photosynthesis of plants directly, it is necessary to establish an intelligent, feasible neural network model, for the important relativities of photosynthetic rate with cultivation and productivity (Nagel and Griffin, 2004). Predicting photosynthetic rate of lettuce using statistical regression and neural network has been reported by Frick et al. (1998); however, the errors between their predicted values and actual values were 14.3 and 24.6% by statistical regression and neural network models, respectively. Meanwhile, the parameters for constructing network models were empirical (Kaur and Raghava, 2004). In fact, the theoretically optimal architecture of neural network was the optimal combination of parameters, but it is difficult to be selected by empirical method. To gain a more precise architecture, it was necessary

*Corresponding author. E-mail: yangjunlz@tom.com. Tel: +86-817-2568352. Fax: +86-817-2568352.

Abbrevations: BPNN, Back propagation neural network; **UD,** uniform design; **Pn,** photosynthetic rate; **PAR,** photosynthetic active radiation; **T,** temperature; **CO_2,** carbon dioxide level; **RH,** relative humidity; **MAPE,** mean absolute percentage error; **MSE,** mean square error; **MAE,** mean absolute error.

Table 1. Parameters and their levels for UD.

Parameter	Level					
	1	2	3	4	5	6
X_1	4	6	7	9	11	13
X_2	0.01	0.02	0.03	0.04	0.05	0.06
X_3	0.2	0.4	0.6	0.7	0.8	0.9
X_4	2000	3000	4000	5000	6000	7000

Symbols x_1, x_2, x_3 and x_4 represented factors for numbers of neurons in hidden layer, learning rate, momentum parameter and iteration numbers respectively. Number 1 to 6 represents the level of each parameter.

to quest an optimal network for predicting photosynthetic rate.

UD has wide use on experiments design. The aim of UD is to choose a set of *n* experimental points that are scattered uniformly on all experiments. Thus, the advantage of UD is to obtain experiments that are uniformly scattered in the domain. So, UD is a good candidate for some experiments design because of its good features, such as its functional agility in arranging experimental runs and its robustness against model uncertainty (Zhang and Fang, 2006). Therefore, in comparison with empirical method, the larruping trait of UD is to achieve results with a great number of factors and factor levels but fewer numbers of experiments (Wang et al., 2008).

BPNN has been proved to be a powerful tool for many fields such as prediction on β-turn types in proteins (Kaur and Raghava, 2004), predicting protein secondary structure (Wood and Hirst, 2004) and predicting harvest dates of greenhouse-grown sweet peppers (Lin and Hill, 2007). However, up to date, there is no report on prediction photosynthetic rate using BPNN based on UD. In this research, BPNN model based on UD was used to determine a more precise protocol for predicting photosynthetic rate of sunflower. Many factors can directly or indirectly effect photosynthetic rate of plants, but the main environmental factors are PAR, CO_2, T and RH. Therefore, it is very necessary to establish an intelligent, feasible model to predict Pn under natural environment only with PAR, CO_2, T and RH.

MATERIALS AND METHODS

Data collection for BPNN

Seeds of sunflowers were sown in the trial field of China West Normal University (N: 30.812°, E: 106.067°) on the 3th of March, 2008. Water and nutrients were managed normally during the whole growth period. Using LI-6400 portable photosynthesis system *LI-6400* (Li-Cor, Lincoln, NE, USA) to measure the daily variation of photosynthesis of flowering sunflowers in natural habitats from 7:00 am to 19:00 pm on 20th, 21st and 22nd of July 2008, and corresponding parameters of photosynthetic active radiation (PAR), temperature (T), carbon dioxide level (CO_2), relative humidity (RH), stomatal conductance (Cond), intercellular carbon dioxide level (Ci), transpiration (Tr) and photosynthetic rate (Pn) were measured by LI-6400 portable photosynthesis system. We measured three sunflowers leaves once an hour and every sunflower had three replicates; the average values of three sunflowers were obtained. So, twelve groups of data were obtained daily and thirty-six groups of data were obtained after three days. The daily variations of sunflowers' photosynthesis were measured on sunny and cloudless days. Because of the appearance of small cloud at 16:00 pm on the 21st July, 2008, eleven groups of data were analyzed. Therefore, thirty-five groups of data were obtained to establish BPNN models. All measurements were under natural conditions. Every groups of data included these parameters: photosynthetic active radiation (PAR), temperature (T), carbon dioxide level (CO_2), relative humidity (RH) and photosynthetic rate (Pn).

Experimental design and optimizing BPNN

Four parameters, numbers of neuron in hidden layer (x_1), learning rate (x_2), momentum parameter (x_3) and iterations number (x_4), were divided into six levels individually and processed by U_6 (6^4) (Table 1). Then six representative experiments were conducted by DPS software and the optimal combination of parameters with least error was employed to establish BPNN model. The obtained thirty-five groups of data were divided randomly, so that twenty-seven groups were used as training samples and eight groups as predicting samples to establish BPNN model and stepwise regression model.

To compare the results of training and testing qualitatively, the following three errors were defined (Zhang and Fang, 2006):

Mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{|y_t|} \tag{1}$$

This is the average error size as a percentage of the mean of the variable being forecasted.

Mean square error (MSE)

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \tag{2}$$

This is the average of squared forecast errors.

Table 2. The U_6 (6^4) table and mean absolute errors.

Number of experiment	Variable level				Mean square errors
	X_1	X_2	X_3	X_4	
N1	3	6	6	2	0.90
N2	4	5	1	1	0.75
N3	6	1	3	4	2.06
N4	2	3	2	6	1.27
N5	5	5	5	5	1.43
N6	1	6	4	3	1.98

Every model was repeated thirty times and the optimal model (combination of network parameters by N2) with least mean square errors (MSE) were chosen and saved for predicting.

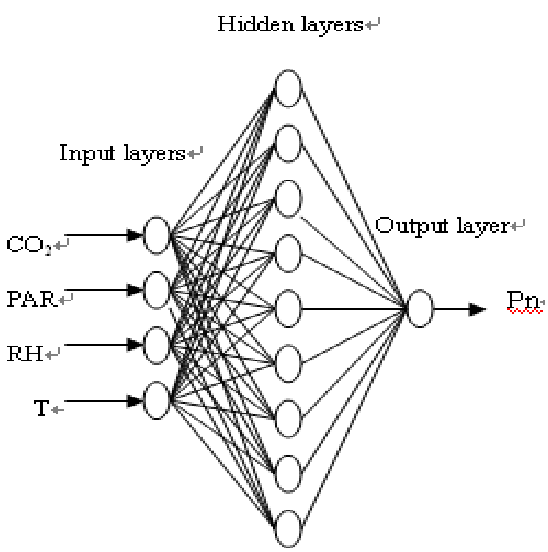

Figure 1. The topology of the back propagation neural network including three layers. The neuron numbers in each layers are input neurons, hidden neurons and output neuron in ratio 4: 9: 1. Input neurons are CO_2 (carbon dioxide level), PAR (photosynthetic active radiation), RH (relative humidity) and T (temperature). Output neuron is Pn (photosynthetic rate).

Mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{3}$$

This is the sum of the absolute values of the errors divided by the number of errors. In the formulae above, yt and $\hat{y}_t$ denotes the actual value and the predicted value, respectively.

DPS software was used for UD and data analysis (Tang and Feng, 2006). Matlab V6.5 was used for BPNN program.

RESULTS

BPNN model based on UD

From the results of UD design (Tables 1 and 2), an optimized architecture was accepted with least MSE of 0.75 µmol CO_2 $m^{-2}s^{-1}$ due to the network parameters: number of neurons in hidden layer (x_1) = 9, learning rate (x_2) = 0.05, momentum parameter (x_3) = 0.2, and iteration numbers (x_4) = 2000. To reduce overfitting, the training and testing error was selected as 0.001. Therefore, the optimal topological architecture of the back propagation neural network was "4 -9 -1", which was used to predict photosynthetic rate under four environmental factors as input layers (Figure1).

To test the aforementioned optimal BPNN model via eight predicting samples, the predicted values and actual values were observed (Table 3 and Figure 2), with the MPAE and MAE being 3% and 0.72 µmol CO_2 $m^{-2}s^{-1}$, respectively. There were no significant difference between the actual values and predicted values by significant test (P>0.05). Therefore, the model based on UD and BPNN could be considered usable for predicting photosynthetic rate of sunflowers.

Stepwise regression model

$$y = 85.0434 - 0.2724*T - 0.2601*CO_2 + 0.4734*RH + 0.0064*PAR \tag{4}$$

The value of the equation R^2 was record as 0.8793. Then the predicting sample data were used to test the stepwise regression model, with the MAPE, MAE and MSE being 7%, 1.53 and 1.76 µmol$CO_2$$m^{-2}s^{-1}$, respectively. There were no significant difference between the actual values and predicted values by significant test (P = 0.96). The results indicated that there were no agreements between the actual values and the predicted values as BPNN model.

DISCUSSION

Optimal architecture of BPNN by UD and stepwise regression analysis

The BPNN is the most successfully and widely used algorithm. A typical three-layer network, including an

Table 3. The results via the optimal model.

Number of samples	Actual value ($\mu molCO_2m^{-2}s^{-1}$)	Predicted value ($\mu molCO_2m^{-2}s^{-1}$)	Absolute error ($\mu molCO_2m^{-2}s^{-1}$)	Absolute percentage errors (%)
1	16.90	16.95	0.05	0.00
2	34.20	32.71	1.49	0.05
3	32.40	33.28	0.88	0.03
4	33.70	32.69	1.02	0.07
5	12.10	13.07	0.97	0.06
6	15.30	16.23	0.93	0.00
7	33.40	33.42	0.02	0.02
8	20.70	20.32	0.38	0.03
Mean	24.84	24.83	0.72	0.03

(MAE, mean absolute error) = 0.72 $\mu molCO_2m^{-2}s^{-1}$; (MAPE, mean absolute percentage error) = 3%. Significant difference between predicted values and actual values by significant test with P = 0.998 *($P > 0.05$).*

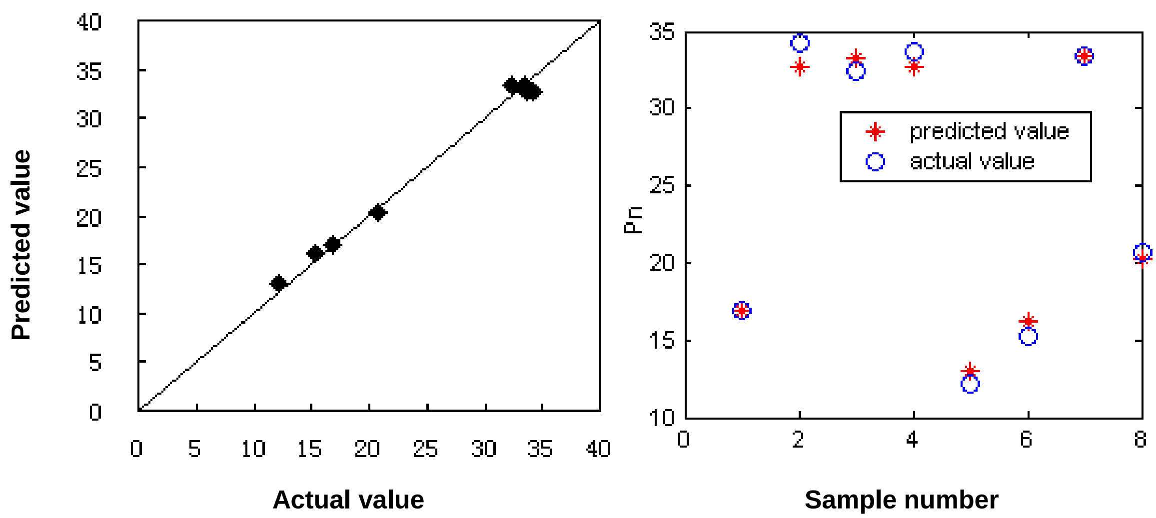

Figure 2. Comparison of actual values and predicted values by optimal BPNN model.

input layer, a hidden layer and an output layer was implemented in many applications, but still difficult to determine parameters such as numbers of hidden neuron and learning rate (Zhang and Fang, 2006). UD was employed here to optimize the architecture of BPNN for its most efficient experimental design and reducing the number of experiments extraordinarily. In this study, only six models were conducted, replacing a great many models with arbitrary combination of parameters.

Compared with previous report (Frick et al., 1998), this work supplied a better model to predict photosynthetic rate and the results also show that the MAPE and MSE by BPNN model were better than by stepwise regression. Therefore, it could be concluded that the model established here was much more accurate than statistical stepwise regression model. From the results, it was also indicated that the relationship between photosynthetic rate and impact factors was very complicated, and satisfactory results could be obtained by the powerful tool BPNN for prediction of nonlinearities, but not from linear regression models.

Optimal photosynthetic rate by improving environmental conditions

A higher photosynthetic rate should increase carbon gain and correspondingly increase oxygen accumulation of biomass, leading to an increase in fitness (Arntz et al., 2000). It is also commonly assumed that instantaneous photosynthetic rates of leaves are the consequence of environmental factors directly or indirectly (Zhang et al., 2005). To evaluate photosynthetic performance and to analyze photosynthetic acclimation to environmental factors such as high CO_2 concentrations, biochemical models of photosynthesis were used increasingly (Peterson et al., 1999). However, the simulation required representation of the intra-canopy or restriction environmental factors (Urban et al., 2003). After taking the aforementioned information into account, intelligent accurate prediction of photosynthetic rate under natural, even perturb conditions (Bunce, 2008; Kakani et al., 2008) was very necessary.

The most important was that, the four key environmental factors chosen from photosynthesis parameters as input layers including carbon dioxide level (CO_2), photosynthetic active radiation (PAR), temperature (T) and relative humidity (RH) were feasible and convenient to observe, and thus to monitor the photosynthesis of sunflowers by the established intelligent model with less time, labor, and experimental cost consuming. Therefore, plant can grow well through monitoring timely and improving environmental conditions to enhance photosynthesis performance, such as alleviating the lunch break phenomenon effectively and replying to greenhouse effect.

Conclusion

In conclusion, the intelligent, feasible model using BPNN based on UD was powerful for predicting photosynthetic rate under four natural environmental factors. Such information would enhance the ability to predict sunflowers' function and competition between different species under changed environmental conditions. Although this method was only applied in sunflower, the results showed their potential applicability in other crops to grow well.

ACKNOWLEDGEMENTS

This work was financially assisted by Application and Basic Foundation, Science and Technology Bureau of Sichuan Province (2008JY0150), Key Program, Education Bureau of Sichuan Province (2003A098), and Sichuan Provincial Key Subject Program (SZD0420), People's Republic of China.

REFERENCES

Arntz AM, Delucia EH, Jordan N (2000). From fluorescence to fitness: variation in photosynthetic rate affects fecundity and survivorship. Ecology, 81: 2567-2576.

Bunce JA (2008). Acclimation of photosynthesis to temperature in Arabidopsis thaliana and *Brassica oleracea*. Photosynthetica. 46: 517-524.

David L, Mary DP, Alvarado JL (2008). Sunflower (*Helianthus annuus* L.) as a pre-Columbian domesticate in Mexico. PNAS, 17: 6232–6237.

Frick J, Precetti C, Mitchell CA (1998). Predicting lettuce canopy photosynthesis with statistical and neural network models. Am. Soc. Horticul. Sci., 123: 1076-1080.

Kakani VG, Surabhi GK, Reddy RK (2008). Photosynthesis and fluorescence responses of C_4 plant *Andropogon gerardii* acclimated to temperature and carbon dioxide. Photosynthetica, 46: 420-430.

Kaur H, Raghava GPS (2004). A neural network method for prediction of β-turn types in proteins using evolutionary information. Bioinformatics, 20: 2751-2758.

Lin WC, Hill BD (2007). Neural network modeling of fruit colour and crop variables to predict harvest dates of greenhouse-grown sweet peppers. Can. J. Plant Sci., 87: 137-143.

Nagel JM, Griffin KL (2004). Can gas-exchange characteristics help explain the invasive success of *Lythrum salicaria*?. Biol. Invasions, 6: 101-111.

Peterson AG, Ball JT, Luo Y (1999). The photosynthesis-leaf nitrogen relationship at ambient and elevated atmospheric carbon dioxide: A meta analysis. Glob. Change Biol., 5: 331-346.

Saraswathi SF, Paliwal K (2008). Diurnal and seasonal trends in photosynthetic performance of *Dalbergia sissoo* Roxb. and *Hardwickia binata* Roxb. from a semi-arid ecosystem. Photosynthetica, 46: 248-254.

Tang QY, Feng MG (2006). Data processing system: experiment design, statistical analysis and modeling. Science press, Beijing, pp. 58-66, 253-263.

Urban L, Roux XL, Sinoquet H, Jaffuel S, Jannoyer M (2003). A biochemical model of photosynthesis for mango leaves: evidence for the effect of fruit on photosynthetic capacity of nearby leaves. Tree Physiology, 23: 289-300.

Wang LQ, Yang J, Deng E, Wang GB (2008). Optimizing the shoot proliferation protocol of *Penthorum chinese* by axillary buds. Biotechnol Lett., 30: 2199-2203.

Wood MJ, Hirst JD (2004). Predicting protein secondary structure by cascade-correlation neural networks. Bioinformatics, 20: 419-420.

Zhang GY, Fang BS (2006). A uniform design-based back propagation neural network model for amino acid composition and optimal PH in G/11 xylanase. J. Chem. Technol. Biotech., 81: 1185-1189.

Zhang SB, Hu H, Zhou ZK, Xu K, Yan N, Li SY (2005). Photosynthesis in Relation to Reproductive Success of *Cypripedium flavum*. Ann. Bot-London, 96;43-49