Horticulture, Environment, and Biotechnology (2018) 59:45–50
https://doi.org/10.1007/s13580-018-0015-1

RESEARCH REPORT

Estimation of greenhouse CO_2 concentration via an artificial neural network that uses environmental factors

Tae Won Moon[1] · Dae Ho Jung[1] · Se Hong Chang[2] · Jung Eek Son[1]

Received: 14 March 2017 / Revised: 27 May 2017 / Accepted: 17 July 2017 / Published online: 12 February 2018

Abstract

In order to improve photosynthesis efficiency and crop growth, it is important to predict CO_2 concentration as well as CO_2 consumption in greenhouses. The objective of this study was to predict greenhouse CO_2 concentration via an artificial neural network (ANN) that incorporated environmental factors. Temperature, relative humidity, atmospheric pressure, solar radiation, and CO_2 concentration were measured every 10 min over a 6-month period in a greenhouse located in Boryeong, Korea. Measured environmental data were used to train the ANN. Among the 14,866 data points used in the experiment, 10,000 and 4866 data points were used for training and testing, respectively. An ANN with an input layer with input neurons, two hidden layers with 32–2048 neurons, and an output later with one neuron was selected. A rectified linear unit was used as the activation function in each node of the ANN. An ANN structure that included 256 neurons in the hidden layers showed the highest test accuracy ($R^2 = 0.97$) was selected from all the structures, while multivariate linear regression showed lower test accuracy than the ANN ($R^2 = 0.78$). The ANN accurately estimated CO_2 concentration in the greenhouse using big data for changing patterns of the inside environmental factors without vent position data. Furthermore, it is possible to estimate crop CO_2 consumption in greenhouses with this ANN using the change in greenhouse CO_2 concentration.

Keywords Black box modeling · Machine learning · Mango · Solar radiation · Temperature

1 Introduction

Greenhouses allow farmers to actively control growth environmental conditions such as temperature, light, relative humidity, and CO_2 concentration. By controlling these environments, crops can be produced year-round regardless of the climate. Based on the benefits of environmental control, greenhouse use and size in agriculture is continuously increasing (Guo et al. 2012).

To maximize the benefits of greenhouse cultivation, it is necessary to control growth environments efficiently. Research that predicts environmental factors has been steadily conducted in order to better control greenhouse environments (Ehret et al. 2001; Sonneveld et al. 2005; Min et al. 2012; Cha et al. 2016; Yu et al. 2016). However, since greenhouses are not completely isolated from the outside, environmental changes within the greenhouse are affected by external factors. Therefore, it is not easy to predict and control the environmental changes within greenhouses. The concentration of CO_2 is an important environmental factor in greenhouses and has a major influence on crop growth (McMurtrie and Wang 1993). Previous studies have reported the importance of CO_2 concentration during crop growth and have attempted to estimate CO_2 concentration (Critten 1991). Furthermore, there have been attempts to estimate and control CO_2 concentration via artificial neural networks (ANNs) using overall greenhouse environments for a short period (Linker et al. 1998).

Recently, eco-friendly greenhouses have been studied with regard to environmental conservation (Cuce et al. 2016). In order to reduce CO_2 emissions that are accelerating global warming, greenhouses that utilize surplus resources from power plants have been being studied. It is necessary to estimate and accurately control CO_2 concentration within these greenhouses to effectively reduce CO_2 emissions. However, CO_2 concentration is affected not only

✉ Jung Eek Son
sjeenv@snu.ac.kr

[1] Department of Plant Science and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea

[2] Korea Electronics Technology Institute, Seongnam 13509, Korea

by various environmental factors, but also by photosynthesis and respiration of the plants grown within the greenhouse itself. Therefore, although greenhouse environments can be controlled in many aspects, CO_2 concentration has a complex nonlinear relationship with environmental variables.

ANNs have been used in recent studies to derive meaningful results from complex nonlinear data (Vaidyanathan 2015; Wang et al. 2016). ANN is gaining popularity over other algorithms because it can achieve high-level abstraction from raw data (LeCun et al. 2015). From the 1980s to the early 2000s, ANNs made simple estimations using small ANN structures. Since 2009, ANNs have been applied to various fields with the emergence of big data and hardware that has superior computational power compared to first-generation technology. The purpose of this study was to estimate CO_2 concentration in greenhouses with an ANN that incorporates environmental factor data.

2 Materials and methods

2.1 Greenhouse and cultivation conditions

A double-span arch-type plastic house (34.4 W × 30.0 L × 5.7 H, m), 1,032 m^2] located at Boryeong, Korea (36°23′34″N 126°29′12″E) was used for the experiment (Fig. 1). Polyolefin films with a thickness of 0.15 mm and a light transmittance of approximately 92% were used as a greenhouse covering material. The inside temperature was maintained at 25 ± 1 °C using a hot-water heating system. The ventilation system was automatically opened at a set point of 27 °C. One hundred 3-year-old Irwin mangoes (*Mangifera indica* L. Irwin) were planted in pots 0.8 m in diameter with a planting density of 6.25 m^2 in the greenhouse. Organic content of the soil ranged from 38 to 120 g kg^{-1}. Stem training and pruning were conducted periodically to induce vegetative growth of crops and to fix tree structure. A drip irrigation system was used for watering.

Fig. 1 A double-span arch-type plastic house (34.4 W × 30.0 L × 5.7 H, m) used for Irwin mango cultivation located at Boryeong, Korea

2.2 Data collection and preprocessing

Environmental factors such as temperature, relative humidity, light intensity, atmospheric pressure, and CO_2 concentration were measured using a complex sensor module developed by Korea Electronics Technology Institute (Seongnam, Korea). The sensor modules were placed at nine locations throughout the greenhouse. Environmental data from the greenhouse were measured every 10 min from July 27 to December 9 2016 and the mean value from the nine locations was used. Weather data such as temperature, relative humidity, wind direction, wind velocity, and atmospheric pressure measured at Boryeong Meteorological Station were used. The time of measuring environmental factors was also used for training. The outside CO_2 concentration was constant at approximately 410 µmol mol^{-1}. Table 1 shows the ranges of the environmental factors measured. In order to improve the training efficiency of the ANN, environmental data was normalized from 0 to 1. A total of 14,866 data points was used for estimating the CO_2 concentration via the ANN.

2.3 Artificial neural network (ANN)

The ANN consisted of an input layer, hidden layers, and an output layer, each of which had neurons. In this study, 10 neurons of the input layer corresponded to environmental data, two hidden layers, and one neuron of the output layer for CO_2 concentration were selected (Fig. 2). In the hidden layers, the value received from the input layer was multiplied by weight and the input information was transmitted through the activation function. A Rectified Linear Unit (ReLU) function was used as the activation function $f(x)$,

Table 1 Ranges of the measured environmental factors

Environmental factor	Range
Inside temperature (°C)	14.2–39.0
Inside relative humidity (%)	25.4–93.1
Inside atmospheric pressure (hPa)	992.6–1036.0
Inside PPFD (µmol m^{-2} s − 1)	0.0–1213.6
Outside temperature (°C)	− 0.1–36.1
Outside relative humidity (%)	0.0–95.0
Outside atmospheric pressure (hPa)	989.9–1033.3
Wind direction (°)	0.0–360.0
Wind velocity (ms^{-1})	0.0–23.0
Inside CO_2 concentration (µmol mol^{-1})	337.0–794.5

PPFD photosynthetic photon flux density

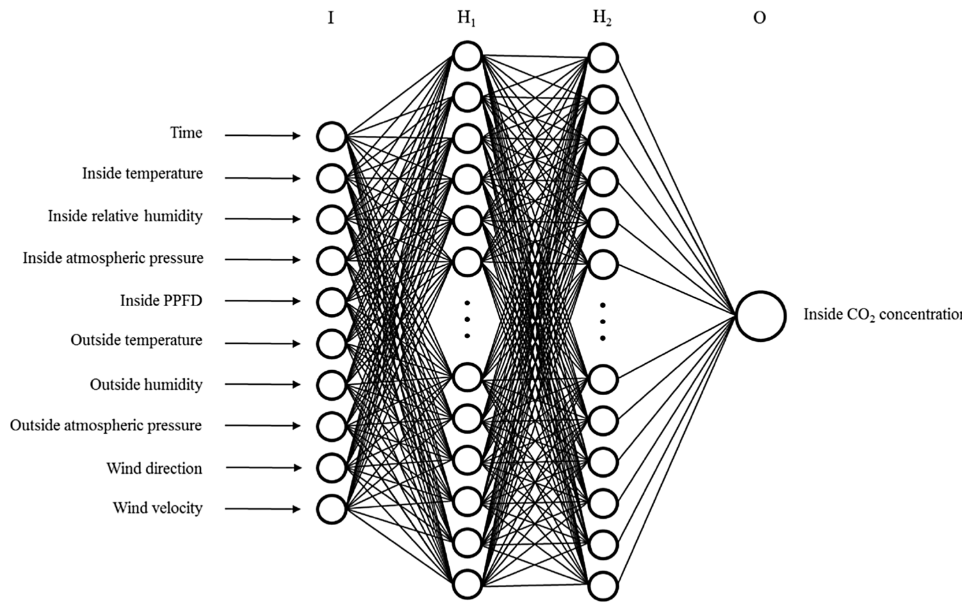

Fig. 2 Structure of an artificial neural network including the input layer (I), hidden layer I (H_1), hidden layer II (H_2), and output layers (O). Inside and outside environmental factors that were used as inputs for I are listed

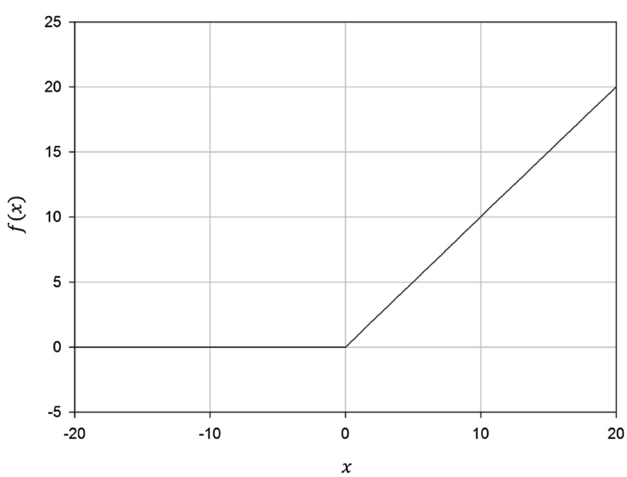

Fig. 3 A rectified linear unit (described in Eq. 1) was used as the activation function in each node of the artificial neural network

where x corresponds to the input value given to each neuron (Eq. 1, Fig. 3).

$$f(x) = max(0, x) \tag{1}$$

Training and testing of the ANN were conducted after classifying 14,866 data randomly obtained from the collection process into 10,000 training data and 4866 test data. The training data helped to adjust and generalize the ANN according to the differences between estimated and measured values. The test data were used to confirm the accuracy of the trained ANN. In addition, data sequences were randomly mixed so that they would not be skewed to a certain period of time. The experiments were performed using Tensorflow (v. 0.11, Python Deep Learning Library, Google, Menlo Park, CA, USA). During the adjustment of the ANN, root mean square error (RMSE) was used for its optimization (Eq. 2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - P_i\right)^2}{n}} \tag{2}$$

where Y_i and P_i are the CO_2 concentrations measured by the sensor and estimated by the ANN, respectively. n refers to the total number of training data and i represents each training data point. The coefficient of determination (R^2) was used for training accuracy and test accuracy to verify model robustness.

For comparison with ANN, multivariate linear regression was conducted with the same data set using Eq. 3. The model was analyzed using the statistical programming language R (The University of Auckland, Auckland, New Zealand).

$$y = \Sigma a_i x_i + b \tag{3}$$

where y, x, a, and b are the dependent variable (CO_2 concentration), independent variable (environmental factor), regression coefficient, and intercept, respectively. The subscript of i indicates the number of independent variables.

In order to train the ANN, the AdamOptimizer was used, which is a method widely used for optimization (Kingma and Ba 2014). The parameters for the AdamOptimizer were set to the commonly used values (Table 2). The number of training was 5000 times in total, where the case of using entire data is called 1 time. In order to confirm the optimal ANN structure, the number of neurons in the hidden layer was changed to 32, 64, 128, 256, 512, 1024, and 2048. The two hidden layers consisted of the same number of neurons.

Table 2 Parameters for the AdamOptimizer artificial neural network

Parameter	Value	Description
Minibatch size	5000	Number of training cases over which each stochastic gradient descent update is computed
Learning rate	0.001	Learning rate used by the AdamOptimizer
β_1	0.9	Exponential mass decay rate for the moment estimates
β_2	0.999	Exponential velocity decay rate for the moment estimates
ε	1e−0.8	A constant for numerical stability
Dropout probability	0.7	Probability of dropping out units in the neural network
Training epoch	5000	Number of training iterations
Training data size	10,000	Size of data set used for training
Test data size	4881	Size of data set used for test

Dropout was set to 1.0 in the test to use the entire neural network

Table 3 Training and test accuracies of the artificial neural network according to the number of neurons in the hidden layers

Number of neurons	Training accuracy (R^2)	Test accuracy (R^2)	RMSE
32	0.924	0.916	34.219
64	0.944	0.927	27.620
128	0.967	0.956	23.275
256	0.979	0.968	19.904
512	0.978	0.966	19.827
1024	0.980	0.966	21.414
2048	0.959	0.953	22.085

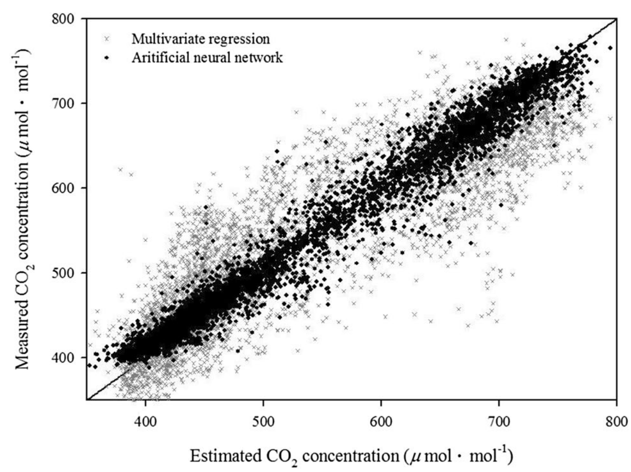

Fig. 4 Comparison of estimated and measured CO_2 concentrations in the greenhouse when using 256 neurons in the hidden layers of the artificial neural network and multivariate linear regression (Table 3)

3 Results and discussion

3.1 Accuracy of the artificial neural network

The maximum test accuracy (R^2) and RMSE were 0.97 and 19.90, respectively, using 256 neurons in the hidden layer of the ANN structures (Table 3). When an ANN has an excessive number of neurons compared to the training data, the ANN cannot generalize the data and is adjusted to fit only the trained data, which is called overfitting (Tetko et al. 1995). When the number of neurons exceeded 256, the training accuracy increased while the test accuracy decreased because of overfitting. Therefore, increasing the ANN structure will not increase the estimation accuracy infinitely.

In addition, the R^2 and RMSE of the multivariate regression model were 0.78 and 54.70, respectively (Eq. 4), indicating that the ANN estimated the CO_2 more accurately than the multivariate linear regression (Fig. 4).

$$\begin{aligned} C_i = &-55.87 * t + 11.28 * T_i + 1.760 * RH_i - 241.7 * P_i \\ &- 0.05812 * L_i - 9.114 * T_o + 1.010 * RH_o \\ &+ 0.04140 * D_o - 0.7791 * v_o + 243.8 * P_o - 1237 \end{aligned} \tag{4}$$

where C, t, T, RH, P, L, D, and v are the CO_2 concentration, time, temperature, relative humidity, atmospheric pressure, light intensity, wind direction, and wind velocity, respectively. Subscripts of i and o mean inside and outside of the greenhouse.

In the ANN, the accuracy was lower at about 500–600 μmol mol^{-1} CO_2 concentrations. The measured CO_2 concentration ranged from 337.0 to 794.5 μmol mol^{-1}, but CO_2 concentration data were lacking at approximately 500–600 μmol mol^{-1}. Therefore, the ANN might not accurately estimate CO_2 at these concentrations due to insufficient data at lower and higher concentrations.

3.2 Validation of CO_2 concentration in the greenhouse

In general, the CO_2 concentrations estimated by the ANN showed better agreement with those measured in the greenhouse than those estimated by the multivariate linear

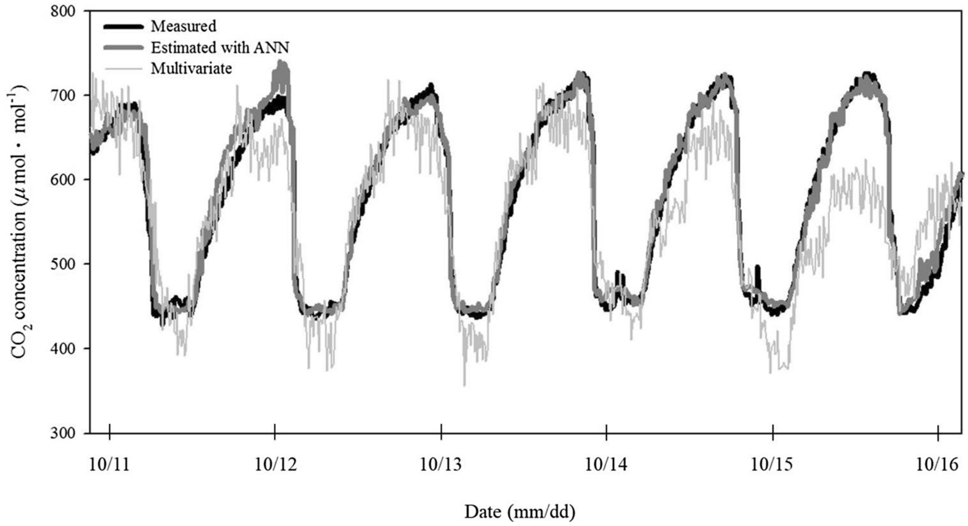

Fig. 5 Comparison of CO_2 concentrations estimated using the artificial neural network (ANN), multivariate linear regression, and measured values in a single greenhouse between October 10–16, 2016

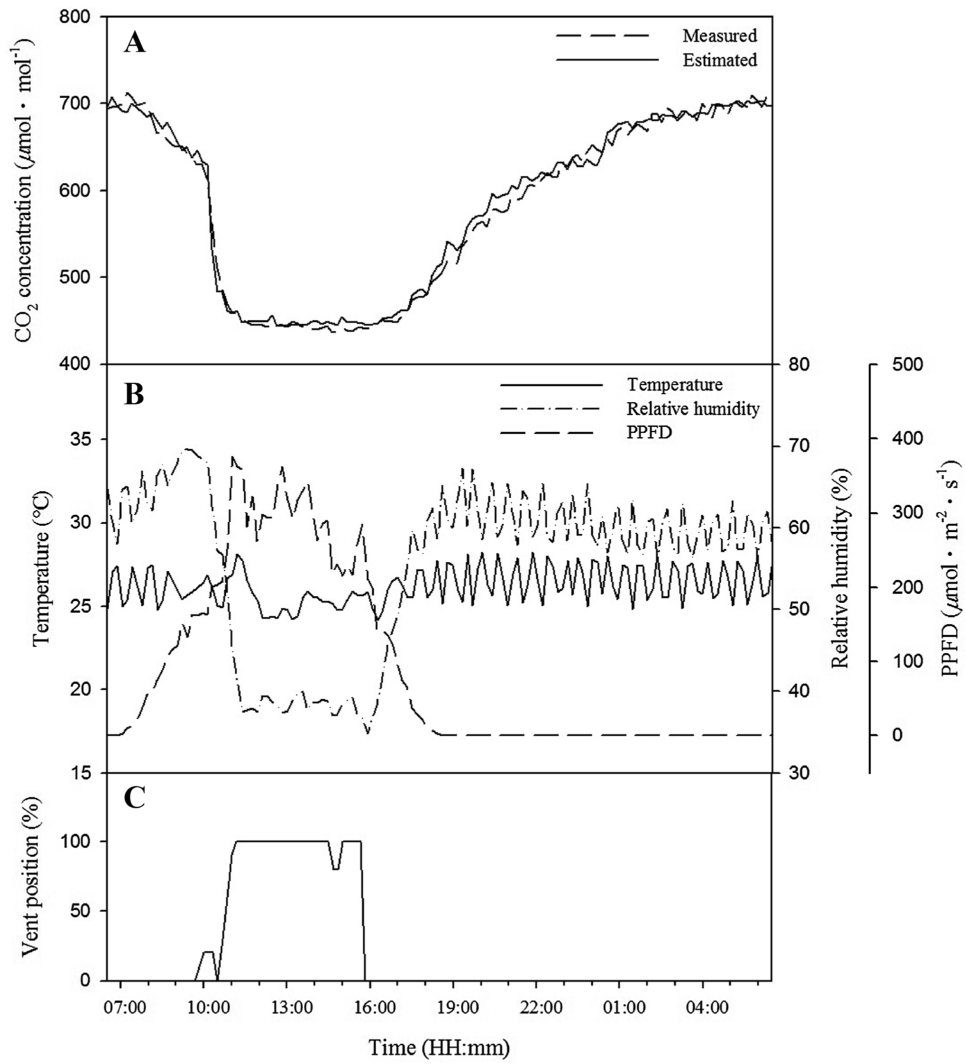

Fig. 6 Estimated and measured CO_2 concentrations (**a**); inside temperature, relative humidity and PPFD (photosynthetic photon flux density) (**b**); and vent position (**c**) over 24 h starting at 06:30 on October 13, 2016. Vent positions represent the opening ratio of the windows (0%—closed and 100%—fully open)

regression (Fig. 5). Compared to the ANN, the multivariate regression model inaccurately estimated the CO_2 concentrations with about 100 µmol mol^{-1} difference on days 15 and 16.

The ANN accurately estimated CO_2 concentrations in the greenhouse using big data for the changes in inside temperature, relative humidity, and CO_2 concentration without vent position data (Fig. 6). It was estimated that the inside CO_2 concentration could be calculated based on ventilation and outside CO_2 concentration after the ANN recognized the change in vent position from the sudden change in relative humidity. Due to the nature of black box modeling, it is difficult to determine exactly what environmental factors influenced the results. However, the ANN accurately estimated changes in CO_2 concentration even though the ventilation affected various environmental factors.

3.3 Limitations and possibilities

The estimates performed in this study were limited to data obtained from a single greenhouse. ANNs should be trained with data from various measurement sites to generalize all possible situations (Lopez et al. 2001). Previous studies with high accuracy had more data points or more inputs related to the factor being estimated (Trejo-Perea et al. 2009). If conditions are difficult to measure, virtual conditions could be modeled with simulation (Beltramo et al. 2016). In this study, CO_2 concentration could be estimated with a high coefficient of determination of 0.97 for the greenhouse located at Boryeong. To ensure that the ANN model used in this study is applicable to all greenhouses, it is necessary to verify the test accuracy using data from other greenhouses or simulation data. Despite experimental limitations, the ANN made significant estimations of the change in CO_2 concentration in the greenhouse. Therefore, the CO_2 concentration within the greenhouse could be estimated using an ANN that incorporated nine environmental factors. This suggests that CO_2 concentration in greenhouses can be estimated even in cases of CO_2 fertilization (Fernandez and Bailey 1992). Further studies are needed to estimate CO_2 consumption by plants in greenhouses using ANN systems.

Acknowledgements This research was supported by the Project of 'Development of Different Industry-Combined Smart System Test Bed for the Utilization of Power Plant-Hot Waste Water and CO_2' (20142020103570), Ministry of Trade, Industry and Energy, the Republic of Korea.

References

Beltramo T, Ranzan C, Hinrichs J, Hitzmann B (2016) Artificial neural network prediction of the biogas flow rate optimised with an ant colony algorithm. Biosyst Eng 143:68–78

Cha MK, Jeon YA, Son JE, Chung SO, Cho YY (2016) Development of a greenhouse environment monitoring system using low-cost microcontroller and open-source software. Korean J Hort Sci Technol 34:860–870

Critten DL (1991) Optimization of CO_2 concentration in greenhouses: a modelling analysis for the lettuce crop. J Plant Biochem Biotechnol 48:261–271

Cuce E, Harjunowibowo D, Cuce PM (2016) Renewable and sustainable energy saving strategies for greenhouse systems: a comprehensive review. Renew Sust Energ Rev 64:34–59

Ehret D, Lau A, Bittman S, Lin W, Shelford T (2001) Automated monitoring of greenhouse crops. Agronomie 21:403–414

Fernandez JE, Bailey BJ (1992) Measurement and prediction of greenhouse ventilation rates. Agric Forest Meteorol 58:229–245

Guo SR, Sun J, Shu S, Lu XM, Tian J, Wang JW (2012) Analysis of general situation, characteristics, existing problems and development trend of protected horticulture in China. China Veg 18:1–14

Kingma D, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980v9

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444

Linker R, Seginer I, Gutman PO (1998) Optimal CO_2 control in a greenhouse modeled with neural networks. Comput Electron Agric 19:289–310

Lopez G, Rubio MA, Martınez M, Batlles FJ (2001) Estimation of hourly global photosynthetically active radiation using artificial neural network models. Agric Forest Meteorol 107:279–291

McMurtrie RE, Wang YP (1993) Mathematical models of the photosynthetic response of tree stands to rising CO_2 concentrations and temperatures. Plant Cell Environ 16:1–13

Min J, Zhang H, Shi W (2012) Optimizing nitrogen input to reduce nitrate leaching loss in greenhouse vegetable production. Agric Water Manag 111:53–59

Sonneveld C, Van den Bos AL, Voogt W (2005) Modeling osmotic salinity effects on yield characteristics of substrate-grown greenhouse crops. J Plant Nutr 27:1931–1951

Tetko IV, Livingstone DJ, Luik AI (1995) Neural network studies. 1. Comparison of overfitting and overtraining. J Chem Inf Comput Sci 35:826–833

Trejo-Perea M, Herrera-Ruiz G, Rios-Moreno J, Miranda RC, Rivasaraiza E (2009) Greenhouse energy consumption prediction using neural networks models. Int J Agric Biol 11:1–6

Vaidyanathan S (2015) 3-Cells cellular neural network (CNN) attractor and its adaptive biological control. Int J PharmTech Res 8:632–640

Wang T, Gao H, Qiu J (2016) A combined adaptive neural network and nonlinear model predictive control for multirate networked industrial process control. IEEE Trans Neural Netw Learn Syst 27:416–425

Yu H, Chen Y, Hassan SG, Li D (2016) Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO. Comput Electron Agric 122:94–102