Ecological Modelling 221 (2010) 575–581

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Identification of important factors for water vapor flux and CO_2 exchange in a cropland

Zhong Qin[a], Gao-li Su[b,c], Jia-en Zhang[a], Ying Ouyang[d,**], Qiang Yu[e,f,*], Jun Li[e]

[a] Key Laboratory of Ecological Agriculture of Ministry of Agriculture of the People's Republic of China, Key Laboratory of Agro-ecology and Rural Environment of Guangdong Regular Higher Education Institutions, South China Agricultural University, Guangzhou 510642, China
[b] Centre of Climatology, Zhejiang Meteorological Bureau, Hangzhou 310029, China
[c] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101, China
[d] Department of Water Resources, St. Johns River Water Management District, Palatka, FL 32178, USA
[e] Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[f] Plant Functional Biology & Climate Change Cluster, and Department of Environmental Sciences, University of Technology, Sydney, PO Box 123 Broadway, NSW 2007, Australia

ARTICLE INFO

Article history:
Received 15 May 2008
Received in revised form 15 October 2009
Accepted 5 November 2009
Available online 16 December 2009

Keywords:
Artificial neural network
Water vapor and CO_2 flux
Cropland
Automatic relevance determination

ABSTRACT

Water vapor flux and carbon dioxide (CO_2) exchange in croplands are crucial to water and carbon cycle research as well as to global warming evaluation. In this study, a standard three-layer feed-forward back propagation neural network technique associated with the Bayesian technique of automatic relevance determination (ARD) was employed to investigate water vapor and CO_2 exchange between the canopy of summer maize and atmosphere in responses to variations of environmental and physiological factors. These factors, namely the photosynthetically active radiation (PAR), air temperature (T), vapor pressure deficient (VPD), leaf-area index (LAI), soil water content in root zone (W), and friction velocity (U^*), were used as inputs in neural network analysis. Results showed that PAR, VPD, T and LAI were the primary factors regulating both water vapor and CO_2 fluxes with VPD and W more critical to water vapor flux and PAR and T more crucial to CO_2 exchange. Furthermore, two time variables "day of the year (DOY)" and "time of the day (TOD)" could also improve the simulation results of neural network analysis. The important factors identified by the neural network technique used in this study were in the order of PAR > T > VPD > LAI > U^* > TOD for water vapor flux and in the order of VPD > W > LAI > T > PAR > DOY for CO_2 exchange. This study suggests that neural network technique associated with ARD could be a useful tool for identifying important factors regulating water vapor and CO_2 fluxes in terrestrial ecosystem.

1. Introduction

Responses of water vapor flux and carbon dioxide (CO_2) exchange between cropland and atmosphere to environmental regulations are important to global warming and terrestrial carbon and water cycle researches. Generally, the eddy-covariance flux measurements and process-based biophysical models (Schelde et al., 1997; Baldocchi and Wilson, 2001) are used to quantify water vapor flux and CO_2 exchange at different space and time scales in biosphere and to identify the primary environmental factors that govern the seasonal patterns of such flux and exchange. These measurements and models have provided useful insights into the research subject. However, the major limitation of these models is that many specific parameters for the selected response functions must be optimized or be defined with detailed physiological data at a species level (Wijk and Bouten, 1999). In addition, extensive process-related data including evaporation, photosynthesis, transport of nutrients and carbon, plant growth, CO_2 fixation, etc. must often be defined in such models (Baldocchi and Wilson, 2001; Wijk et al., 2002). These complex data requirements make it difficult to apply the models (Amthor, 1994; Lek et al., 1996; Lek and Guegan, 1999). Therefore, a need exists to apply a new approach in such modeling researches. To this end, the artificial neural network (ANN) technique was employed to circumvent the obstacles in this study.

The ANN is a powerful tool for the purpose of this study since it can process information in a non-linear manner (Bruntz, 1989; Beale and Jackson, 1990; Raiche, 1991), enable completely unconstrained optimization and estimate input–output responses (Kosko, 1992; Demuth and Beale, 1995; Schulz and Härtling, 2003). This property makes the ANN a promising method for ecological and environmental modeling. Some example applications of the ANN in ecological and environmental sciences were reported in

* Corresponding author. Tel.: +86 10 64856515; fax: +86 02 9514 4142.
** Co-corresponding author. Tel: +1 386 312 2320; fax: +1 386 329 4125/4329.
E-mail addresses: ouyangy@ufl.edu (Y. Ouyang), yuq@igsnrr.ac.cn (Q. Yu).

0304-3800/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2009.11.007

early 1990s (Bolte, 1989; Zhuang and Engel, 1990; Muttiah and Engel, 1991; Chao and Anderson, 1994) and have since gained increasing popularity (Brey et al., 1996; Scardi, 1996; Francl and Panigrahi, 1997; Giske et al., 1998; Tjoelker et al., 1999; Werner and Obachb, 2001; Dekker et al., 2001; Wijk et al., 2002; Moisen and Frescino, 2002; Dedecker et al., 2004). Most of these studies showed that ANN has high accuracy in ecological modeling as compared to those classical methods (Lek et al., 1996; Huntingford and Cox, 1997).

Recently, great attentions have been given to design, test and apply models for computing rates of biosphere–atmosphere trace gas exchange for better understanding on how trace gas fluxes may respond to environmental perturbations. Since these responses are the highly non-linear processes, several attempts have been devoted to probing the ANN's capacity in modeling trace gas fluxes in ecosystem. Wijk and Bouten (1999) applied ANN to simulate water vapor and CO_2 fluxes in coniferous forest ecosystems by using a minimal set of input variables. These authors found that ANN models could predict water vapor and CO_2 fluxes independently of tree species and without the need for detailed physiological or site-specific information. More recently, Ryan et al. (2004) used ANN to simulate nitrous oxide emissions from an intensive grassland ecosystem as a function of six input variables (e.g., daily rainfall, soil moisture content, temperature, and soil nitrate). They demonstrated that the ANN model was a potential useful tool to simulate complex biological systems in soils without the need of complex parameters as have been used in the traditional models such as mechanistic models. These modeling studies have provided very valuable insights into complex biological systems. However, the capability of ANN in modeling water vapor and CO_2 fluxes between the cropland and the atmosphere has not yet been investigated. Furthermore, no effort has been devoted to identifying and ranking the impacts of the key input variables upon these fluxes.

The goal of this study was to investigate environmental and physiological constraints on water vapor and CO_2 fluxes in a summer maize field in North China Plain, using both the field measurements and the ANN technique. The specific objective was to identify the relatively important input variables or factors regulating water vapor and CO_2 fluxes from the maize field using the Bayesian technique of automatic relevance determination (ARD) associated with the ANN analysis. As an algorithm for training ANN within the Bayesian framework, application of the ARD technique would eliminate the need for detailed physiological information such as light interception, soil water and nutrient contents, photosynthesis, and stomatal conductance characteristics. A successful application of the ARD technique would also lead to an improvement in ANN modeling for agricultural water and carbon cycle research, especially in the semi-arid area.

2. Material and methods

2.1. Study site and data collection

The experimental site was located at the Yucheng Experimental Station (36°57′N, 116°36′E, 20 m ASL) in North China Plain, Shandong province, China, which is a semi-humid area with a monsoon climate. Mean annual precipitation, temperature, and solar radiation at the station over the past three decades are 528 mm, 13.1 °C, and 5225 MJ m^{-2}, respectively. Winter wheat-summer maize rotation is the conventional culture system practiced in this region. Generally, winter wheat, Gaoyou No. 503, is sown at the rate of 150 kg ha^{-1} with 20 cm wide per row by hand. Maize, Yandan No. 21, is sown at a rate of 60 kg ha^{-1} per plot after the winter wheat growth season. The growth season for winter wheat is from early December to mid-June and for maize is from early-June to later September.

The experimental site used in this study has loamy soil and is rich with nutrients and organic matter (Table 1). The eddy-covariance system, mounted at 2 m above the canopy, was used in this study. This system consisted of a three-axis sonic anemometer (model CSAT3, Campbell Sci., Logan, UT) for measuring wind speed and sonic virtual temperature, an open path, infrared absorption gas analyzer (CS-7500, Campbell Scientific Inc.) for measuring water vapor flux and CO_2 concentrations, and a suite of software (Loggernet 2.0) for real-time and post processing analysis. The fluxes were measured using a data logger (model CR23X, Campbell Sci., Logan, UT) for a 10 min interval and then averaged to a 30 min interval. Correction for density effects was conducted to the water vapor flux.

The above-canopy net radiation was measured with net radiometers (CNR-1, Kipp & Zonen Inc., Saskatoon, Saskatchewan, Canada). Incident photosynthetically active radiation (PAR) was measured with radiation sensors (LI-190SZ, LI-COR Inc., Lincoln, NE). Air temperature and relative humidity were measured with a thermistor and capacitive RH sensor probe (model HMP45C, Vaisala, Helsinki, Finland). Wind speed was measured with cup anemometers (034A-L and 014A, R.M. Young Co., Traverse, MI, USA). Two soil heat flux plates (model HFT-3, Campbell Scientific Inc.) were embedded inter-rows and inter-plants at a depth of 0.05 m to determine heat fluxes. Soil temperature was measured with copper–constantan thermocouples. Soil water content was measured with time-domain reflectometry sensors (CS-615, Campbell Scientific Inc.) at the soil depth of 0.05, 0.20, and 0.50 m. Rainfall was measured at 0.7 m above the ground with a tipping bucket (TE525MM, Campbell Scientific Inc.). Irrigation and fertilizer were applied with same frequency and amount as the local farmland. Leaf-area index was measured with an electronic leaf-area meter (LAI-2000, LI-COR, Lincoln, NE) every 5 d throughout the crop growth season.

2.2. Data processing

Experimental data collected during the summer maize growth period, which began from the day of sowing (day of year or DOY 165) to harvest (DOY275) in 2003, were used for the ANN analysis. All 30-min eddy-covariance measurement data were validated vigorously for anomalous turbulence and for sensor malfunction (Hollinger et al., 1995; Falge et al., 2001a). About 5.67% of the missing data due to instrument maintenance and calibration was removed. About 0.06%, 0.48%, and 0.73% of unreasonable data for CO_2 flux (F_{CO_2}), latent heat flux (LE), and sensible heat flux (Hs), respectively, were discarded. About 3.06% and 3.41% of soil heat flux

Table 1
Characteristics and physical parameters of soil profiles at Yucheng Station.

Soil layer depth (cm)	Soil texture	Bulk density (g cm^{-3})	Total nitrogen (g kg^{-1})	Organic matter (g kg^{-1})	Wilting coefficient	Field capacity (m^3 m^{-3})	Saturated water content (m^3 m^{-3})	pH
0–20	Sandy loam	1.28	0.64	9.56	0.08	0.29	0.4	8.26
20–65	Sandy loam	1.39	0.36	5.3	0.08	0.33	0.41	8.33
65–97	Medium loam	1.4	0.21	2.66	0.28	0.37	0.42	8.52
97–104	Light clay	1.42	0.17	2.32	0.28	0.37	0.44	8.53
104–150	Sandy clay loam	1.39	0.17	1.86	0.28	0.37	0.46	8.55

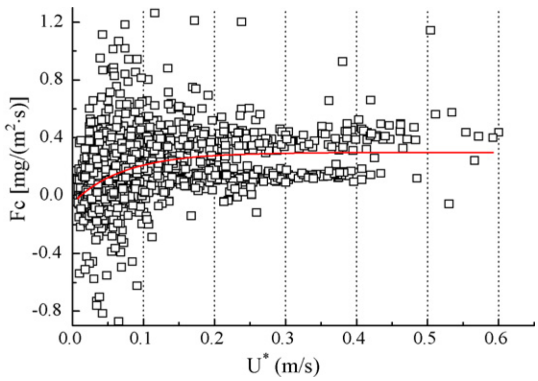

Fig. 1. Measured nighttime CO_2 flux (Fc) vs. nighttime turbulence (U^*).

data recorded inter- or intra-row were eliminated because of malfunction of the sensors and supporting equipment. Measurements within 24 h after a rain event were removed from the database as well because the sensor optics obscured (Wilson and Baldocchi, 2000).

Nighttime fluxes are normally underestimated by the eddy-covariance system during the stable condition because the CO_2 storage in the layer is below the eddy flux system (Villalobos, 1997). In this study, a wind friction velocity (U^*) threshold ($U^* > 0.12\ \mathrm{m\,s^{-1}}$) was determined (Falge et al., 2001b; Anthoni et al., 2004) and flux measurements when U^* was smaller than the threshold were removed from the data set to minimize problems related to insufficient turbulent mixing (Fig. 1).

Energy closure, expressed as $\mathrm{Rn} - G = \mathrm{Hs} + \mathrm{LE}$, where Rn ($\mathrm{W\,m^{-2}}$), is the net radiation and G ($\mathrm{W\,m^{-2}}$) is the soil heat flux, and LE and Hs are latent and sensible heat fluxes density ($\mathrm{W\,m^{-2}}$) on a 30 min basis, respectively. The LE is associated with evaporation of water at the surface and subsequent condensation of water vapor in the troposphere and therefore is used as an indicator of water vapor flux. Linear regression indicated that the agreement between the sum of the turbulent fluxes (LE + Hs) and the available energy (Ra) was generally good (Fig. 2).

In an ANN analysis, the training set is a set of data used for learning, which is to fit the parameters (i.e., weights) of the classifier. The validating set is a set of data used to tune the parameters of a classifier. The testing set is a set of data used only to assess the performance (generalization) of a fully specified classifier (Ripley, 1996).

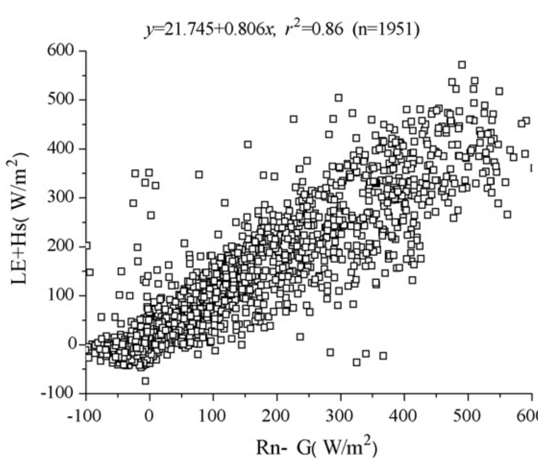

Fig. 2. Half-hourly sums of LE + Hs against available energy (Rn − G) during the summer maize growth period.

2.3. Input variables selection

The optimal selection of input variables is always crucial in ANN-based modeling. A basic understanding of the governing system for the outputs is required for an appropriate set of input variables during initial ANN development. In our study, crop surface water vapor and CO_2 exchange were taken as outputs. Temporal variations of these two variables are controlled by meteor-biological driving forces such as solar radiation, wind speed, air temperature, humidity, soil moisture, temperature, and biological processes (Kaimal, 1972; Anderson et al., 1986; Jarvis et al., 1997; Baldocchi, 1997). Solar radiation is known as a primary energy source for all of the physiological functions of vegetations. Of which, about 50% of the energy, defined as photosynthetically active radiation (PAR), is absorbed by plants for their growth. It is assumed that PAR is a major parameter controlling many biological and physical processes related to the evolution of plant canopies, agricultural and environmental fields (López et al., 2001). Furthermore, many of the exchange processes between the vegetation canopies and the atmosphere, as well as dry matter yield, are regulated by photosynthesis, which is related to the amount of absorbed PAR (Hanan and Bégué, 1995; Li et al., 1997). Therefore, PAR is the most likely input candidate in constructing an ANN model for modeling water and carbon exchange between the crop and the atmosphere. Air temperature (T) is another critical factor for canopy photosynthesis, development, growth and biomass partition of the crop. In addition, when modeling mass and energy transfer process, the environmental driving factors such as temperature, photosynthesis, and evapotranspiration should also be included.

Vapor pressure deficit (VPD) is the difference between the actual water vapor pressure and the saturation of water vapor pressure at a particular temperature, which has been proved to have a simple nearly straight-line relationship to the rate of evapotranspiration and other measures of evaporation. VPD is being used to predict crop water needs especially in some commercial irrigation systems. The impacts of VPD will serve as a basis in building a monitoring model for the amount of water requirement and the rate of CO_2 assimilation.

Wind speed above the crop canopy is important for determining the formation and transfer of momentum, heat, and moisture processes. As a major turbulence driving force in the CO_2 and water vapor exchange processes, wind speed and its direction, influenced by the circumfluence of the planetary boundary layer and the condition of underlying vegetation, changed dramatically over time, but its friction velocity varied in the rhythm that it becomes high during the day and low during the night. Consequently, wind friction velocity (U^*) was selected as an input for the ANN model.

Soil water content (W) in the root zone controls crop water use, soil aeration conditions, and CO_2 production from respirations, which makes it highly correlated with field-scale heat flux and CO_2 transport. Therefore, soil water content associated with water vapor flux and CO_2 exchange between the crop and the atmosphere should be included in the ANN modeling.

Information on leaf-area index (LAI) is paramount since it determines the population of biologically active material that is exchanging gas and energy with the atmosphere (Baldocchi and Meyers, 1998). Since it is highly related to a variety of canopy processes, such as water interception, evapotranspiration, photosynthesis, etc., LAI is used as a key parameter for global and regional models of biosphere/atmosphere exchange. LAI will be included preferentially in constructing ANN model for depicting the exchange of water vapor and CO_2 between crop field surface and atmosphere. It should be noted that all these input variables mentioned above have different distribution patterns with time scales. Two time factors "day of year" (DOY) and "time of day" (TOD, expressed in digital form) should be included in the ANN analysis

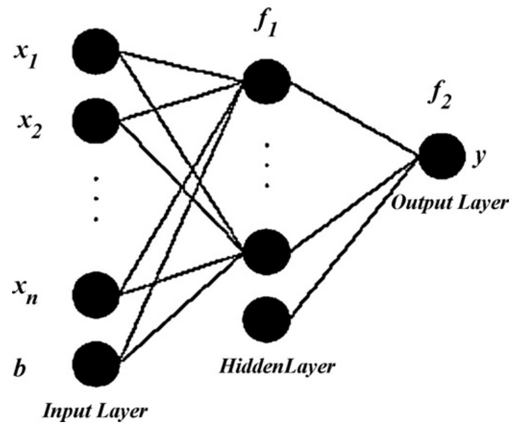

Fig. 3. Topology of a three-layer feed-forward neural network. Redrawn after Lek et al. (1996).

(Wijk and Bouten, 1999) for better predictions of water vapor and CO_2 fluxes.

2.4. ANN model development

A feed-forward back propagation neural network with an input layer, hidden layer and output layer was employed for a general non-linear mapping between the driving meteorological variables and the measured water vapor and CO_2 fluxes. The number of input and output nodes corresponded to the number of input and output variables, while the number of the hidden nodes depended on the complexity of the relations between input and output variables. Topology of a three-layer feed-forward neural network (Lek et al., 1996) used in our study was shown in Fig. 3. Determination of the number of the hidden nodes is not an easy task because using a larger number of hidden nodes can potentially improve the accuracy and convergence of the neural network at the cost of increasing the computational time. In this study, the optimal number of hidden neurons has been determined empirically as the minimum number of neurons for which estimation performance on a testing set is satisfied. The activation function used in the neural network was a sigmoidal function:

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

As mentioned above, six input parameters, namely the photosynthetically active radiation (PAR), air temperature (T), vapor pressure deficient (VPD), leaf-area index (LAI), soil water content in root zone (W), and friction velocity (U^*), were used in this study. Each parameter had 1742 data points. These data were randomly divided into three data sets (i.e., training, validating, and testing sets) using the function "rand" the ANN package of the MATLAB. The percentages (i.e., 70% for training, 20% for testing, and 10% for validating) used in this study are based on the literature reports (Matthew et al., 2004).

To minimize the training time by eliminating the possibility of reaching the saturation regions of the sigmoid transfer function during training, both the input and output values were linearly scaled to lie within the range between 0 and 1 using:

$$x_{\mathrm{scaled}} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \tag{2}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum recorded values for each input variable, respectively.

An appropriate structure for the back propagation algorithm, including the number of hidden layers and the number of nodes in the hidden layer, is selected with a trial-and-error process. Training of the network was accomplished by presenting the network with the extracted training and testing data sets from which the network could learn from and calibrate itself. Once the optimal net architecture was found, the network was applied to the whole data set for obtaining an indication of the general fit of the model to the data. The network was then applied to validate the accuracy of predicting water vapor and CO_2 fluxes using the validation data set. The optimization method used in the validation phase was the Levenberg–Marquardt method (Marquardt, 1963; Demuth and Beale, 1995), which minimizes the total sum of squared errors (SSE) between measured and modeled values by tuning the ANN parameters (e.g., scaling factors and inter-neuron connection weights). All these computations were accomplished using Neural Network Toolbox 2.0 in Matlab (Ver.6.5).

2.5. Inputs ranking using ARD

Previous studies have shown that ANN was well suited for simulation of water and carbon exchange between biosphere and atmosphere (Wijk and Bouten, 1999). However, no information was provided on: (1) What were the key factors that regulate water and carbon fluxes? (2) How many key factors were needed to efficiently encode information for surface-atmosphere flux exchange? (3) What were the states of the key factors for which input combinations were most relevant to the water vapor and carbon fluxes? With the aim to answer these questions, we applied the Bayesian technique of automatic relevance determination (ARD) in this study to identify and rank the relative importance of different input variables in responses to the water vapor and CO_2 fluxes from the summer maize field. The ARD technique selects the most relevant input parameters and discards those that do not contribute significantly to the dynamics of a system being modeled. In this method, a hyperparameter is associated with each input. The hyperparameter for an input corresponds to an inverse variance of the weights on connections from that input, and represents the relevance of that input to the task of predicting the measured output. Thus, the smallest hyperparameter will indicate the input which accounts for the largest part of the variability and hence is the most relevant input, and the hyperparameter with the next increase in value indicates the next most relevant input and so on (López et al., 2001). Detailed illustrations of the ARD technique for input variable ranking can be found in Neal (1992, 1996) and MacKay (1992, 1994, 1995).

Two steps were adopted when constructing the ANN models for simulating water vapor and CO_2 fluxes. In the first step, six meteor-biological driving variables (e.g., PAR, T, and LAI) along with two time variables "DOY" and "TOD" were used as inputs to build the model. In the second step, the variable combinations which achieved the best simulation results were selected plus two time variables "DOY" or "TOD" as the new inputs for the network to re-establish the model. By checking the modeling results, the possibilities of DOY or TOD in improving the results of neural networks could be determined.

Table 2
Results for the automatic relevance determination (ARD).

Input	Hyperparameter	
	Fc[a]	LE[b]
PAR	0.97120	1.08686
LAI	5.00055	0.93179
T	1.76553	1.00712
VPD	1.92014	0.56611
U^*	8.76210	1.54045
W	17.56479	0.69038
Random	1406.01692	73,173.12907

[a] Fc represents CO_2 flux.
[b] LE represents latent heat flux.

Table 3
Results of ANN approximations for CO_2 flux (Fc).

Model	Input variables	Topology structure	Slope of the linear regression	Intercept of the linear regression	R^2	RMSE
1	PAR, *T*	2-5-1	0.5040	−0.0902	0.5038	0.2860
2	PAR, *T*, VPD	3-5-1	0.6706	−0.0568	0.6596	0.2754
3	PAR, *T*, VPD, LAI	4-7-1	0.8942	−0.0219	0.8951	0.1750
4	PAR, *T*, VPD, LAI, *U**	5-9-1	0.9113	−0.0240	0.9054	0.1684
5	PAR, *T*, VPD, LAI, *U**, *W*	6-9-1	0.8991	−0.0238	0.8974	0.1738
6	PAR, *T*, VPD, LAI, *U**, TOD	6-9-1	0.9141	−0.0201	0.9102	0.1642
7	PAR, *T*, VPD, LAI, *U**, DOY	6-9-1	0.9072	−0.0218	0.9062	0.1669

Table 4
Results of the ANN approximations for water vapor flux (LE).

Model	Input variables	Topology structure	Slope of the linear regression	Intercept of the linear regression	R^2	RMSE
8	VPD, *W*	2-5-1	0.4933	44.51	0.4989	46.35
9	VPD, *W*, LAI	3-7-1	0.5832	36.48	0.5957	44.94
10	VPD, *W*, LAI, *T*	4-7-1	0.6155	33.79	0.6259	44.61
11	VPD, *W*, LAI, *T*, PAR	5-8-1	0.9153	7.779	0.9153	26.11
12	VPD, *W*, LAI, *T*, PAR,*U**	6-8-1	0.9359	5.547	0.9352	23.09
13	VPD, *W*, LAI, *T*, PAR,DOY	6-8-1	0.9422	5.185	0.9373	22.48
14	VPD, *W*, LAI, *T*, PAR,TOD	6-8-1	0.9282	6.489	0.9314	23.22

3. Results and discussion

During the processes of neural modeling, the ARD technique was implemented to determine which input variables are most relevant to water vapor and CO_2 fluxes. Responses of fluxes to input variables were given in Table 2. This table shows that the artificial random input variable had the largest hyperparameter value, indicating that this variable was highly un-relevant to the CO_2 exchange, while the input variable PAR had the smallest hyperparameter value, signifying that this variable was mostly correlated to CO_2 exchange. The relevant order for the rest of input variables in response to CO_2 exchange was in the following order: $T > \mathrm{VPD} > \mathrm{LAI} > U^*$. It should be noted that the variable *W*, the least relevance to CO_2 exchange, could be excluded from the neural network model since the inclusion of this variable might lead to a slight degradation of the model fitness. Table 2 further reveals that the variable VPD was most relevant to water vapor flux. The relevant order for the rest of input variables in response to water vapor flux was in the following order: $W > \mathrm{LAI} > T > \mathrm{PAR} > U^*$. It should be emphasized that the variable U^*, a least relevant input variable, could be removed in case of a worse model fitness.

Tables 3 and 4 show the statistics obtained with the ANN model for estimating half-hourly water vapor flux and CO_2 exchange between the crop canopy surface and atmosphere. The fitness of the ANN models was evaluated based on the regression analysis of estimated versus measured values, in terms of the intercept and slope of the linear fit and the determination coefficient, R^2. Root mean square error (RMSE) was also provided in the table. During the ANN training, several numerical experiments were conducted to find the combination of number of hidden nodes that gave the

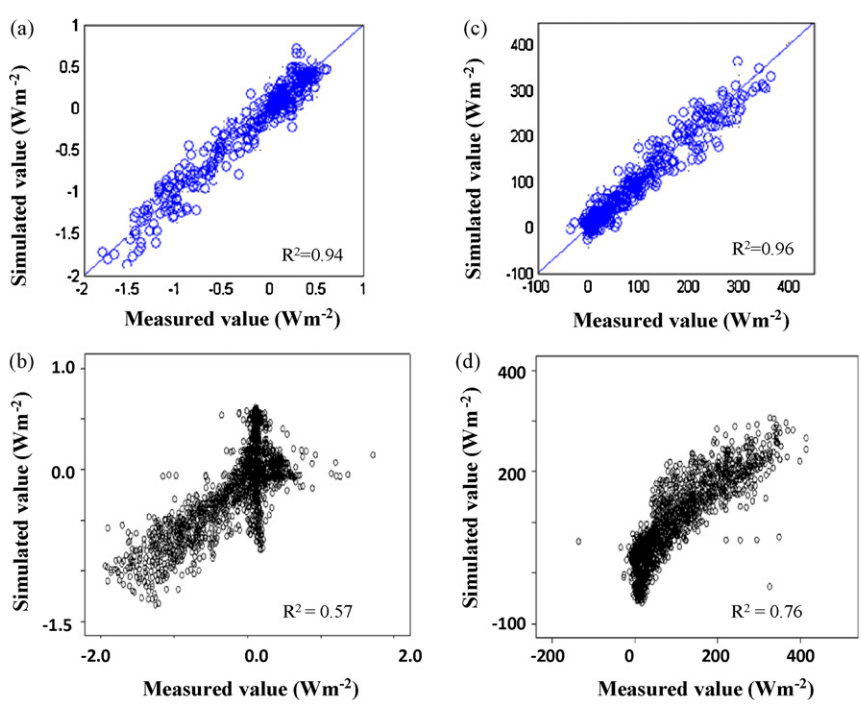

Fig. 4. Comparison of the ANN approximations with stepwise linear regressions for CO_2 and water vapor fluxes. (a) ANN approximations for CO_2 flux; (b) stepwise linear regressions for CO_2 flux; (c) ANN approximations for water vapor flux; (d) stepwise linear regressions for water vapor flux.

greatest fitness in predicting the validation data set. As the number of hidden nodes increased, the R^2 value for water vapor and CO_2 fluxes increased until the number of hidden units reached the values of 8 and 9, respectively. For values higher than 8 and 9, the number of hidden units did not seem to improve the estimates for the LE and Fc.

As shown in Tables 3 and 4, simulation results for water vapor and CO_2 fluxes were improved after combining DOY or TOD with the meteor-biological variables. Among several different inputs combinations, the combinations of "PAR–*T*–VPD–LAI–U^*–TOD" and "VPD–*W*–LAI–*T*–PAR–DOY" were the optimum topology for Fc and LE, respectively, because they had the highest R^2 and the lowest RMSE. Furthermore, the input variables PAR, *T* and VPD could enhance the performance of the neural network analysis for both carbon and water vapor fluxes although their relevance orders were different in water vapor flux and CO_2 exchange simulations.

A successful description of ecological processes requires an appropriate definition of the control structure (i.e., selection of system output, input and disturbance variables) and an efficient model on which the design, analysis and evaluation can be carried out. Thus, the confidence in the obtained results depends on the validity of the control structure and of the model used. For fluxes exchange process under the regulation of climatic control and environmental factors, the standard linear regression methods is typically used to relate an outcome (or dependent variable or response) to several independent variables. To evaluate the use of the stepwise regression in fluxes data analysis, we compared the performance of the neural networks and predictions of standard methods from statistics (Fig. 4). Results showed that the ANN approximation predicted the correlation for CO_2 flux and water vapor flux well as compared to those of the stepwise regression method, especially for CO_2 flux. The determination coefficient (R^2) had significantly increased from 0.57 to 0.94.

In this study, we presented an attractive approach which enables solution of highly non-linear and noisy black-box modeling problems with reference to the position or ordination of the inputs by introducing the neural network technique associated with ARD. It must be stressed, however, that we cannot deduce information about the possible errors of the model, like in statistical models. Except for the generalization error, i.e. the RSME over all test inputs, neural networks have not much to offer for measuring the quality of the net prediction. In particular, when there is only a small number of measurements available, a typical situation in ecology, it is essential to have some information how reliable the network prediction is (Werner and Obachb, 2001).

4. Conclusions

In this study, a Bayesian technique of automatic relevance determination (ARD) and a three-layer feed-forward back propagation neural network was employed to predict responses of water vapor and CO_2 exchange between a summer maize field and the atmosphere to environmental variables. The simulation results demonstrated that the VPD and *W* had most influential effects on surface water vapor flux, while the PAR and *T* were the key physical driving factors for CO_2 flux. Model performance could be improved when TOD and DOY were included as the input variables. Our study shows that the combinations of PAR–*T*–VPD–LAI–U^* with TOD and VPD–*W*–LAI–*T*–PAR with DOY yielded the optimum topology for Fc and LE, respectively. With the ARD feature selection scheme, input variables for surface fluxes exchange could be ranked in the order of their relevance. The Bayesian neural network offers a promising and viable alternative for cropland water and carbon fluxes modeling.

Acknowledgements

This study was supported by the National Basic Research Program of China State (No. 2005CB121106 and No. 2006CB100206), the National High Technology Research and Development Program of China (No. 2008AA10Z215), the National Natural Science Foundation for Youths (Grant No. 30800134). We are grateful to Yucheng Station, Institute of Geographic and Natural Research, Chinese Academy of Science for providing data sets. We thank Dafeng Hui for his enlighten advice on gap-filling missing data in eddy-covariance measurement. We are indebted to G. López for his suggestion on the development of neural network for this study.

References

Amthor, J.S., 1994. Scaling CO_2-photosynthesis relationships from the leaf to the canopy. Photosynth. Res. 39, 321–350.
Anderson, D.E., Verma, S.B., Clement, R.E., Baldocchi, D.D., Matt, D.R., 1986. Turbulence spectra of CO_2, water vapor, temperature and velocity over a deciduous forest. Agric. For. Meteorol. 38, 81–99.
Anthoni, P.M., Freibauer, A., Kolle, O., Schulze, E.D., 2004. Winter wheat carbon exchange in Thuringia, Germany. Agric. For. Meteorol. 121, 55–67.
Baldocchi, D.D., 1997. Flux footprints within and over forest canopies. Bound-Lay Meteorol. 85, 273–292.
Baldocchi, D.D., Meyers, T.P., 1998. On using eco-physiological, micrometeorological and biogeochemical theory to evaluate carbon dioxide, water vapor and trace gas fluxes over vegetation: a perspective. Agric. For. Meteorol. 90, 1–25.
Baldocchi, D.D., Wilson, K.B., 2001. Modeling CO_2 and water vapor exchange of a temperate broadleaved forest across hourly to decadal time scales. Ecol. Model. 142, 155–184.
Beale, R., Jackson, T., 1990. Neural Computing: An Introduction. Institute of Physics Publishing, Bristol, UK.
Bolte, J.P., 1989. Application of Neural Networks in Agriculture. ASAE Paper No. 89-7591. American Society of Agricultural Engineers, St. Joseph, MI.
Brey, T., Jarre-Teichmann, A., Borlich, O., 1996. Artificial neural network versus multiple linear regression: predicting P:B ratios from empirical data. Marine Ecol. Progr. 140, 251–256.
Bruntz, M., 1989. Use of neural network to do pattern recognition tasks in geoscience applications. In: Geological Society of America, Annual Meeting, St. Louis, Missouri, Abstract with programs, vol. 21, No. 6, p. 57.
Chao, K., Anderson, R., 1994. Neural-fuzzy Interface System for Daily Growth of Single Stem Roses. ASAE Paper No. 94-4015. St. Joseph, MI.
Dedecker, A.P., Goethals, P.L.M., Gabriels, W., De Pauw, N., 2004. Optimization of artificial neural network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). Ecol. Model. 174, 161–173.
Dekker, S.C., Bouten, W., Schaap, M.G., 2001. Analysing forest transpiration model errors with artificial neural networks. J. Hydrol. 246, 197–208.
Demuth, H., Beale, M., 1995. Neural Network Toolbox. For Use with Matlab. The MathWorks, Natick.
Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.O., Katul, G., Keronen, P., Kowalski, A., Lai, C.T., Law, B.E., Meyers, T., Moncrieff, J., Moors, E., William Munger, J., Pilegaard, K., Rannik, Ü., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001a. Gap filling strategies for defensible annual sums of net ecosystem exchange. Agric. For. Meteorol. 107, 43–69.
Falge, E., Baldocchi, D.D., Olson, R., Anthoni, P., Aubinet, M., et al., 2001b. Gap-filling strategies for long-term energy flux data sets. Agric. For. Meteorol. 107, 71–77.
Francl, L.J., Panigrahi, S., 1997. Artificial neural network models of wheat leaf wetness. Agric. For. Meteorol. 88, 57–65.
Giske, J., Huse, G., Fiksen, O., 1998. Modelling spatial dynamics of fish. Rev. Fish. Biol. Fish. 8, 57–91.
Hanan, N.P., Bégué, A., 1995. A method to estimate instantaneous and daily intercepted photosynthetically active radiation using a hemispherical sensor. Agric. For. Meteorol. 74, 155–168.
Hollinger, D.Y., Kelliher, F.M., Schulze, E.D., Vygodskaya, N.N., Varlargin, A., Milukova, I., Byers, J.N., Sogachov, A., Hunt, J.E., McSeveny, T.M., Kobak, K.I., Bauer, G., Arneth, A., 1995. Initial assessment of multi-scale measures of CO_2 and H_2O flux in Siberian taiga. J. Biogeogr. 22, 425–431.
Huntingford, C., Cox, P.M., 1997. Use of statistical and neural network techniques to detect how stomatal conductance responds to changes in the local environment. Ecol. Model. 97, 217–246.
Jarvis, P.G., Massheder, J., Hale, D., Moncrieff, J., Rayment, M., Scott, S., 1997. Seasonal variation of carbon dioxide, water vapor and energy exchanges of a boreal black spruce forest. J. Geophys. Res. 102, 28953–28967.
Kaimal, J.C., 1972. Turbulence spectra, length scales, and structure parameters in the stable surface layer. Bound-Lay Meteorol. 4, 289–309.
Kosko, B., 1992. Neural Networks and Fuzzy Systems. A Dynamical Systems Approach to Machine Intelligence. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, p. 449.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, L., Lauga, J., Stephane, A., 1996. Application of neural networks to modeling nonlinear relationships in ecology. Ecol. Model. 90, 39–52.
Lek, S., Guegan, J.F., 1999. Artificial neural networks as a tool in ecological modeling and introduction. Ecol. Model. 120, 65–73.
Li, Z., Moreau, L., Cihlar, J., 1997. Estimation of photosynthetically active radiation absorbed at the surface. J. Geophys. Res. 102, 29717–29727.
López, G., Rubio, M.A., Martinez, M., Batlles, F.J., 2001. Estimation of hourly global photosynthetically active radiation using artificial neural network models. Agric. For. Meteorol. 107, 279–291.
MacKay, D.J.C., 1992. Bayesian interpolation. Neural Comput. 3, 415–447.
MacKay, D.J.C., 1994. Hyperparameters: optimise or integrate out? In: Heidbreder, G. (Ed.), Maximum Entropy and Bayesian Methods, Santa Barbara 1993. Kluwer, Dordrecht.
MacKay, D.J.C., 1995. Bayesian non-linear modelling for the 1993 energy prediction competition. In: Heidbreder, G. (Ed.), Maximum Entropy and Bayesian Methods, Santa Barbara 1993. Kluwer, Dordrecht.
Marquardt, D.W., 1963. An algorithm for least-squares estimation of non-linear parameters. SIAM. J. Appl. Math. 11, 431–441.
Matthew, R., Muller, C., Di, H.J., Cameron, K.C., 2004. The use of artificial neural networks (ANNs) to simulate N_2O emissions from a temperate grassland ecosystem. Ecol. Model. 175, 189–194.
Moisen, G.G., Frescino, T.S., 2002. Comparing five modelling techniques for predicting forest characteristics. Ecol. Model. 157, 209–225.
Muttiah, R.S., Engel, B.A., 1991. Neural Network Methodology in Agriculture and Natural Resources. ASAE Paper No. 91-7018.
Neal, R.M., 1992. Bayesian Training of Back Propagation Networks by the Hybrid Monte Carlo Method. Technical Report CRG-TR-91-1. Department of Computer Science, University of Toronto, Canada.
Neal, R.M., 1996. Bayesian learning for neural networks. Lecture Notes in Statistic, vol. 118. Springer, New York.
Raiche, A., 1991. A pattern recognition approach to geophysical inversion using neural networks. Geophys. J. Int. 105, 629–648.
Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, New York.
Ryan, M., Müller, C., Di, H.J., Cameron, K.C., 2004. The use of artificial neural networks (ANNs) to simulate N_2O emissions from a temperate grassland ecosystem. Ecol. Model. 175, 189–194.
Scardi, M., 1996. Artificial neural networks as empirical models for estimating phytoplankton production. Mar. Ecol.-Prog. Ser. 139, 289–299.
Schelde, K., Kelliher, F.M., Massman, W.J., Jensen, K.H., 1997. Estimating sensible and latent heat fluxes from a temperate broad-leaved forest using the Simple Biosphere (SiB) model. Agric. For. Meteorol. 84, 285–295.
Schulz, H., Härtling, S., 2003. Vitality analysis of Scots pines using a multivariate approach. For. Ecol. Manag. 186, 73–84.
Tjoelker, M.G., Reich, P.B., Oleksyn, J., 1999. Changes in leaf nitrogen and carbohydrates underlie temperature and CO_2 acclimation of dark respiration in five boreal tree species. Plant Cell Environ. 7, 767–778.
Villalobos, F.J., 1997. Correction of eddy covariance water vapor flux using additional measurements of temperature. Agric. For. Meteorol. 88, 77–83.
Werner, H., Obachb, M., 2001. New neural network types estimating the accuracy of response for ecological modeling. Ecol. Model. 146, 289–298.
Wijk, M.T., Bouten, W., 1999. Water and carbon fluxes above European coniferous forests modeled with artificial neural networks. Ecol. Model. 120, 181–197.
Wijk, M.T., Bouten, W., Verstraten, J.M., 2002. Comparison of different modelling strategies for simulating gas exchange of a Douglas-fir forest. Ecol. Model. 158, 63–81.
Wilson, K.B., Baldocchi, D.D., 2000. Seasonal and interannual variability of energy fluxes over a broadleaved temperate deciduous forest in North America. Agric. For. Meteorol. 100, 1–18.
Zhuang, X., Engel, B.A., 1990. Neural Networks for Applications in Agriculture. ASAE Paper No. 90-7024. American Society of Agricultural Engineers, St. Joseph, MI.