Research Article

Effects of Environmental Factors on Ozone Flux over a Wheat Field Modeled with an Artificial Neural Network

Zhilin Zhu1,2

1Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
2College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

Correspondence should be addressed to Zhilin Zhu; zhuzl@igsnrr.ac.cn

Received 24 January 2019; Revised 11 May 2019; Accepted 22 May 2019; Published 4 June 2019

Academic Editor: Ilan Levy

Copyright © 2019 Zhilin Zhu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ozone (O3) flux-based indices are considered better than O3 concentration-based indices in assessing the effects of ground O3 on ecosystem and crop yields. However, O3 flux (Fo) measurements are often lacking due to technical reasons and environmental conditions. This hampers the calculation of flux-based indices. In this paper, an artificial neural network (ANN) method was attempted to simulate the relationships between Fo and environmental factors measured over a wheat field in Yucheng, China. The results show that the ANN-modeled Fo values were in good agreement with the measured Fo values. The R^2 of an ANN model with 6 routine independent environmental variables exceeded 0.8 for training datasets, and the RMSE and MAE were 3.074 nmol·m^−2·s and 2.276 nmol·m^−2·s for test dataset, respectively. CO2 flux and water vapor flux have strong correlations with Fo and could improve the fitness of ANN models. Besides the combinations of included variables and selection of training data, the number of neurons is also a source of uncertainties in an ANN model. The fitness of the modeled Fo was sensitive to the neuron number when it ranged from 1 to 10. The ANN model consists of complex arithmetic expressions between Fo and independent variables, and the response analysis shows that the model can reflect their basic physical relationships and importance. O3 concentration, global radiation, and wind speed are the important factors affecting O3 deposition. ANN methods exhibit significant value for filling the gaps of Fo measured with micrometeorological methods.

1. Introduction

In an atmospheric column, approximately 90% of total ozone (O3) exists in the stratospheric O3 layer and protects Earth’s surface from excessive UV radiation. This category of O3 is therefore referred to as “good” ozone [1]. However, the troposphere O3 can negatively affect vegetation tissues, photosynthesis, and crop yields [2]. The tropospheric O3 mainly results from the in situ photochemical production, and a small portion is transported from the stratosphere [3]. O3 destruction involves chemical decomposition and deposition on Earth’s surface. The total O3 deposition can be quantitatively reflected by measured O3 flux (Fo), and its magnitude and variations can facilitate understanding of the O3 deposition processes.

The quantitative assessment of the effects of ground O3 on ecosystem and crop yield loss in a region requires selection of an assessment model and calculation of the corresponding assessment index, and the calculation of assessment indices requires continuous measurement of O3 concentration or flux [4, 5]. The concentration-based indices do not consider the status of vegetation and ecosystems (e.g., stomatal conductance and leaf area index). Many studies have shown that O3-reduced yield loss is more closely related to the stomatal O3 flux [5–7].

Gerosa et al. [8] presented a method of estimating O3 stomatal flux by partitioning total Fo. Briefly, O3 stomatal resistance was obtained by converting vapor stomatal resistance (Penman–Monteith approach), and the O3 concentration near the canopy was calculated using the resistance...
model, which needs the total O$_3$ flux. So, measuring total $F_o$ is the first step in estimating stomatal O$_3$ flux [9]. $F_o$ is usually measured by micrometeorological methods, and the eddy covariance (EC) method is considered the most direct flux-measuring method [8, 10–12]. However, due to a lack of high performance fast-response O$_3$ analyzers (e.g., it needs to change O$_3$-sensitive dye often and its sensitivity is unstable), EC-measured O$_3$ flux has many temporal gaps in the associated datasets [8, 13, 14]. This leads to difficulties in calculating the value of a O$_3$ stomatal flux-based index using the EC O$_3$ flux since the effect of O$_3$ on crop yields results from the accumulation of O$_3$ damage. Therefore, it is necessary to find a method of filling the gaps in O$_3$ flux datasets.

The gap-filling methods for ecosystem CO$_2$ flux have been summarized in the literature [15, 16]. However, there have not been investigations into methods for filling gaps in $F_o$ data. As the controlling factors for O$_3$ flux are more complex, it is difficult to produce physically based equations to reflect the $F_o$ and environmental variables under any conditions. Although some mechanism-based models (typically a resistance-in-series) have been developed, many parameters must be empirically estimated based on the specific site or type of vegetation [17, 18].

Artificial neural networks (ANN), as an artificial intelligence using machine learning techniques, have been applied for approximation, prediction, classification, modeling, and data gap-filling in many scientific disciplines [16, 19–21]. ANN is able to learn from training data and to represent the nonlinearity between the dependent and independent variables. It is suitable for simulating empirical datasets with complex and nonlinear relationships between variables and dependent variable(s). It is particularly useful when it is difficult to express these relationships using one or more simple equations. Compared to physically based models, ANNs have more success in simulating complex processes with a high precision without any analytical models [22–24]. These attributes make ANNs a good candidate for attempts at filling gaps in O$_3$ flux data. In addition, a sensitivity analysis can be conducted by analyzing the response curves of the different input variables that are created by fixing all other input variables at a certain value [25, 26]. For gap filling of $F_o$, ANN modeling is also a practical and simple method [16].

ANN methods have been used widely to model and predict O$_3$ concentration [27, 28] and CO$_2$ fluxes [21, 29, 30]. However, the ANN approach has not been used to simulate O$_3$ flux variation with other factor fluxes or meteorological factors. The objectives of this study are to (1) develop a series of ANN models to simulate the relationship between $F_o$ and environmental variables; (2) determine the optimal ANN model to simulate $F_o$ for interpolation of measured $F_o$; and (3) analyze the responses and relative importance of different environmental variables to O$_3$ flux.

2. Materials and Methods

2.1. Site and Observations. The observations were carried out over a winter wheat field at the Yucheng Comprehensive Experiment Station of the Chinese Academy of Sciences (36°50′N, 116°34′E; 28 m a.s.l.; Shandong Province, China). The site is located in the Northwest-Shandong plain, characterized by loamy soil texture and semiarid and warm temperate climate. The experimental site is very flat, and the fetch requirements for EC measurements are met. The study period of the field experiment covered entire growing season of wheat (from 2 March to 6 June 2012).

The ambient O$_3$ concentration was measured with a slow-response UV absorption-based O$_3$ analyzer (Model 205, 2B Technologies Inc., USA). The $F_o$ flux was measured with the eddy covariance method in combination with observations of CO$_2$, H$_2$O, and sensible heat fluxes. These variables were measured with a 3D sonic anemometer (CSAT3, Campbell Scientific Instrument, USA) and an open-path CO$_2$/H$_2$O gas analyzer (LI-7500, LI-COR Biosciences, USA). The O$_3$ fluctuation was measured with a fast-response O$_3$ analyzer developed by Enviscope GmbH [31] (hereafter referred to as ENVI). Micrometeorological and radiation variables measurements include air temperature and relative humidity (HMP45C; Vaisala Co., Finland), wind speed (A100R; Vector Instruments, UK), net radiation (CNRI; Kipp and Zonen, the Netherlands), and photosynthetically active radiation (LI-190SB; LI-COR, USA). All sensors were installed at 2.2 m height.

Because of the continuous consumption of organic dye, the sensitivity of the ENVI slowly decreased with time. To maintain high sensitivity in the ENVI, we replaced the organic dye disc every 3 to 4 days. Raw data from the EC system were recorded at 10 Hz, and 30-minute mean data were recorded by a data logger (CR3000, Campbell Sci., USA).

2.2. O$_3$ Flux Calculation. Since the fast-response O$_3$ analyzer’s sensitivity is not constant [14], the sensitivity must be calibrated by a slow-response O$_3$ concentration analyzer. However, it has been shown that the sensitivity can be treated as constant when using a 30 min averaging interval [32]. This means that the ENVI’s output $X$ (in mV) is proportional to the absolute ambient O$_3$ concentration over a 30 min period.

Based on this assumption, the O$_3$ deposition velocity ($V_d$), defined as the O$_3$ flux divided by O$_3$ concentration, can be calculated by the following equation [33]:

$$V_d = \frac{F_o}{\rho_o} = -\frac{\omega V_G}{V_G},$$

where $\rho_o$ is the O$_3$ mass density (or concentration; $\mu$g m$^{-3}$), $\omega$ is the vertical wind speed (m s$^{-1}$), and $V_G$ is the output signal (mV) of the fast-response O$_3$ analyzer. The overbar denotes the time average, and the primes indicate the temporal fluctuation of each variable. The O$_3$ vertical turbulent flux $F_o$ (µg s$^{-1}$ m$^{-2}$) between atmosphere and surface is calculated by the following equation:

$$F_o = -\frac{\rho_o V_d}{T} = \frac{P M_{o3}}{R} \frac{x_o}{V_G} \omega V_G,$$
2.3. Artificial Neural Network Modeling. In a general model, variable(s) and the independent variable can be expressed as one or more analytical expression(s). In contrast, artificial neural networks (ANNs) are capable of recognizing and simulating complex relations without any analytical expression(s) of variables and outputs. The relation is produced by the ANN that has been trained. Although there are many types of ANN (e.g., Hopfield networks and Kohonen networks) and many algorithms to be used to train ANN, the feedforward network (FFN) with backpropagation (BP) algorithm is the most popular and has been used in more than other types of neural networks for a wide variety of problems [26, 37–40]. The feedforward BP networks have a simple structure for simulating complex systems, and its structure is sufficiently robust to simulate many nonlinear systems [27, 41]. Therefore, the BP network was used in the study. Figure 1 shows a schematic of the three-layer BP network. The nonlinear elements or neurons are arranged in successive layers, and the information flows from input layer to output layer through the hidden layers. The number of inputs depends on the model performance, and theoretically, more inputs increase model performance. The hidden layers are another important parameter in a BP network.

Because of the complexity of the ANN calculation process, the ANN core algorithms used in this study are adopted from the MATLAB software package. However, the ANN requires a prepared dataset and the determination of parameters. The general steps for the setup of the BP network and parameters are described as follows:

1. The determination of independent variables (Xᵢ, 1, 2, ..., n). Although ANNs do not require the specification of dependent variables and independent variables, variables should have a certain relationship in terms of mechanism(s) or variation patterns. Based on the previous studies and observed items, we attempt to build a series of ANN models. These models have different combinations of input variables to determine the most relevant input variables for modeling O₃ flux. The input variables include micrometeorological variables, radiation values, and fluxes.

2. The determination of the number of nodes or “neurons” in the hidden layer. This is related to experience or examinations. In general, this number should be larger than the number of Xᵢ.

3. The generation of a network by training. The training dataset and parameters are the keys to generate a good BP network. The training dataset must be representative and of high quality, and the parameters should be suitable. The training process will determine a set of optimal weights and biases. Although there are different transfer functions, such as linear function and a threshold function [38], sigmoid transfer function is the most popular choice for many studies [40]. This transfer function was also used for the neurons in the hidden layer. Using the MATLAB software package, the weights and biases in the ANN are obtained in an iterative calibration process based on the Levenberg-Marquardt (LM) algorithm. The root-mean-squared error is used to monitor the performance of modeling. Normally, the validation error decreases during the initial phase of training and begins to rise when the network begins to overfit the data. To avoid overfitting, the weights and biases of the network can be automatically saved when the validation error reaches a minimum.

4. The evaluation of the ANN model using training and test datasets. In this study, three statistical parameters, i.e., root-mean-squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R²), were used to evaluate the performance of models. RMSE and MAE present information about the degree of accuracy of an ANN model, and R² measures the degree of the linear relationship between measured and modeled O₃ fluxes.

The three parameters were calculated with the following equations:

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Y_i - F_i)^2},
\]

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |Y_i - F_i|,
\]

\[
R^2 = \frac{\sum_{i=1}^{N} (Y_i - \overline{Y})(F_i - \overline{F})^2}{\sum_{i=1}^{N} (Y_i - \overline{Y})^2 \sum_{i=1}^{N} (F_i - \overline{F})^2},
\]

Figure 1: Schematic of a three-layer backpropagation artificial neural network.
where $Y_1$ and $F_1$ are the simulated and measured O$_3$ flux, respectively; the overbar denotes the average of each variable, and $N$ is the sample number.

2.4. Dataset Preparation. To create a satisfactory ANN model, a high-quality dataset is very important. Due to different reasons, all variables, especially the measured O$_3$ flux, presented many low quality or erroneous data. In this study, the selected data for training and testing satisfy the conditions as follows. Approximately 58.5% of the total data points in the set are low quality or erroneous, and they were filtered using the following metrics.

(1) Only daytime data were adopted. During nighttime, O$_3$ concentration is generally low and wheat stomata are closed, and the effect of O$_3$ on wheat can be ignored. Nighttime fluxes also exhibit a significant amount of inaccuracies because of weak turbulence [42].

(2) Ozone fluxes are between $-40$ nmol m$^{-2}$ s$^{-1}$ and 0 nmol m$^{-2}$ s$^{-1}$. As O$_3$ is always deposited downward on the surface, the positive O$_3$ fluxes represent errors. The initial threshold value of $-40$ nmol m$^{-2}$ s$^{-1}$ was empirically determined via the time series of O$_3$ flux. Moreover, the threshold was variable based on the wheat status. For example, the threshold was set to $-20$ nmol m$^{-2}$ s$^{-1}$ in the first stage of wheat. In addition, some rapid spikes of O$_3$ flux were deemed unlikely based on the underlying mechanisms. These data points were deleted even though they did not exceed the threshold value.

(3) The data of all inputs at a time are completed. Some input variables, especially the turbulent fluxes, have errors or gaps due to various reasons. To keep training data consistent, these data were also excluded.

Finally, 1402 data points (41.5% of total O$_3$ flux data) were used to train and validate all ANN models. Half of the high-quality dataset (701 data points) was selected for ANN training, and the other half of high-quality data were used for the test dataset. According to the MATLAB software manual, the training dataset is randomly divided into three subsets, i.e., the training set, the validation set, and the test set. Their ratios are 70%, 15%, and 15%, and the corresponding data points are 491, 105, and 105, respectively. The training set is used for computing the gradient as well as updating the network weights and biases. The validation set is used to validate the ANN performance, and the test set is used to compare different models.

3. Results and Discussion

3.1. Input Variables Choice and Its Performances. Although an ANN model is only dependent on the mathematical relations of inputs and outputs, variables that are irrelevant or lack a physical basis should be excluded. Based on equation (1), O$_3$ flux is the product of O$_3$ concentration and deposition velocity ($V_d$). For O$_3$ concentration, the production and decomposition result from a series of complex chemical and physical processes. The local O$_3$ concentration variation mainly depends on its precursors (e.g., VOCs and NO$_x$), long-range transport and a subset of environmental factors, such as radiation, temperature, and humidity [43–45]. $V_d$ is affected to a certain extent by atmospheric turbulence intensity and underlying surface conditions, such as wind speed, friction speed, soil moisture, O$_3$-reactive chemicals, and vegetation status [46–48].

Due to a lack of observations for O$_3$-reactive chemicals, only some micrometeorological and radiation variables were selected to model $F_o$ variation in this study. The initial input variables for attempting to model $F_o$ include O$_3$ concentration ($C_{O_3}$), global radiation ($Q$), air temperature ($T_a$), relative humidity (RH), wind speed ($U$), and soil water content ($S_w$). In addition, O$_3$ flux access to stomata is coupled with the CO$_2$ flux ($F_c$) and water vapor fluxes (LE). These two fluxes are also deployed in our attempts to model $F_o$. Because of the high correlation between the mean diurnal pattern of many variables and time [30], the effect of time ($T_m$) on the ANN model was also investigated in this study.

To test the dependency of each variable on O$_3$ flux, the O$_3$ flux with one variable was first simulated, and then, O$_3$ flux with different variable combinations were simulated. Table 1 presents the statistics for each variable and variable combinations. For any single micrometeorological variable, the RMSE values exceed 5 ng m$^{-2}$ s$^{-1}$ and all $R^2$ values are less than 0.4. This means that any single variable cannot accurately model O$_3$ flux. Among these variables, O$_3$ concentration and radiation have relatively strong correlations. For other turbulent fluxes (LE, $F_c$), ANN-modeled $F_o$ values were strongly correlated with $F_c$ and LE, but modeled $F_o$ values were weakly correlated with sensible heat flux ($H$). This highlights the fact that O$_3$ flux is mainly affected by the stomata flux, as stomata are the main pathway of LE and $F_c$. Although these fluxes have similar diurnal patterns, $F_c$ and LE cannot be considered as the driving factors of O$_3$ deposition. The O$_3$ deposition process is complex, and changes in the process are mainly driven by many meteorological or environmental factors.

To obtain an optimal ANN model, we attempted to build more ANN models with different combinations of input variables. The combinations can be divided into two classes. Class 1 combinations only consist of routine micrometeorological and radiation factors ($T_a$, RH, $U$, $Q$, $S_w$, and $C_{O_3}$), i.e., models T11 to T15 in Table 1. Meanwhile, we also tested the effect of time ($T_m$). The variables in Class 1 are easy to observe and possess sufficiently high accuracy. The combinations can increase the fitness of observed and ANN-modeled $F_o$, even if the combination only uses meteorological variables (T12). Class 2 combinations add the turbulent fluxes ($H$, $F_c$, and LE) in addition to the variables in Class 1, i.e., models T17 to T20 in Table 1. Although the correlation levels increase slightly in Class 2 models when compared to the models of Class 1, the effect of these fluxes on $F_o$ gap filling is very limited. When there is a gap in $F_o$ data, other fluxes ($F_c$ and LE) often exhibit simultaneous gaps in their observational datasets. Meanwhile, the flux data have lower accuracies than the micrometeorological and
radiation variables. Among the two classes of combinations, we select T15 and T20 as examples to analyze their performances as well as the contributions and responses of each variable to O₃ flux.

Figure 2 shows a scatter plot of measured and modeled O₃ fluxes using ANN models T15 and T20 for the training dataset and test dataset. The modeled O₃ fluxes in these two models match the measured O₃ flux to a high degree. There are not obvious deviations from the 1:1 line. Although the first model contains only six routine variables (no turbulent fluxes), the model produced an $R^2$ of 0.826 and 0.767 for the training dataset and test dataset, respectively. The fitness of model T20 is marginally superior to the fitness of T15. It is clear that the ANN modeled $F_o$ each run are different, and they are around the measured $F_o$. To mitigate these uncertainties, we calculated the average of 10 runs with the same variables combinations and parameters to generate the modeled O₃ flux. Table 2 compares the statistics of ANN-modeled $F_o$ with T15. As shown, statistical indicators of 10-run-averaged are better than separate indicators. The RMSE and MAE are the smallest, and the $R^2$ is the largest.

Another problem that arises in ANN modeling is so-called overtraining or overfitting. An overtrained network cannot learn the general traits that exist in the training set, and the network will lose the capacity to generalize. Three important parameters can be used to avoid this phenomenon: the number of hidden layers, the number of epochs, and the number of neurons in each hidden layer. Although these parameters are crucial to a model, there are no definite rules on how to determine them. In the ANN package of the MATLAB software, the number of hidden layers and the number of neurons in each hidden layer can be assigned. To decrease the number of ANN parameters, a feedforward network with one hidden layer and a sufficient number of neurons was used in this study. This satisfies any finite input-output statistical modeling [41]. The optimal number of epochs (iterations) can be determined automatically by the software based on the best performances. Therefore, only the optimum neurons should be determined by testing the performance of an ANN model with different neurons.

Figure 4 shows the variations of MAE, RMSE, and $R^2$ for the models T15 and T20 as the number of neurons change. It can be seen that the MAE and RMSE of the two ANN models rapidly decrease when the number of neurons changes from 1 to 10, and then the variations are slower and more stable.
Figure 2: Observed versus simulated ozone flux with models T15 and T20, respectively. (a) Model T15 with the training dataset. (b) Model T15 with the test dataset. (c) Model T20 with the training dataset. (d) Model T20 with the test dataset.

Figure 3: Comparison of diurnal variations of ozone flux measured on May 12 with the EC method ($F_{O_{EC}}$) and 10-run ANN modeled with T15. Tra_mean is the mean of 10 runs in the ANN model of $F_{O}$. 
The minima for MAE and RMSE appeared when there were 17 neurons. The performance of $R^2$ is similar (Figures 4(b) and 4(d)). Therefore, the number 17 was determined to be the optimal neuron number for the hidden layer in our analysis.

### 3.3. Response of $F_o$ to Variables Based on ANN Models.

All parameters of an ANN model are fixed once training is finished. The responses of an ANN model can be evaluated by varying a single variable within its measured range and keeping other inputs fixed at their mean values [30, 37]. Figure 5 demonstrates the details of how $O_3$ flux varies across the range of all variables within models T15 and T20. The ranges (minimum and maximum) and mean values of each input variable in the training dataset are given in Table 3.

By excluding the effects of turbulent fluxes, the effect of each environmental variable on $F_o$ can be seen more clearly (Figure 5(a)). Positive responses between $O_3$ concentration, temperature, solar radiation, and relative humidity are shown clearly. After considering the effects of turbulent fluxes (Figure 5(b)), the responses of $F_c$ and LE are very obvious, whereas the responses of $T_a$, RH, and $S_{sw}$ are very weak. Moreover, $O_3$ flux shows a unidirectional variation with most variables, such as $O_3$ concentration, wind speed, temperature, and humidity. Nevertheless, the variation of $F_o$ with soil water content ($S_{sw}$) shows a different pattern (Figure 5(a)). $O_3$ deposition decreased when soil water
content was too high or too low. The results might be unreliable and should be explained carefully. One possible reason for this pattern is that the change in soil water content is slow over the course of a day and there is no clear diurnal variation. The response of $S_W$ and $F_o$ in the ANN models mainly reflects the long-term variation of $O_3$ flux. For other variables, the diurnal variation amplitude is larger than that of long-term variations. A simple expression that relates variables to each other is not given by the ANNs, as the response of $F_o$ to a single variable may depend on the values of the other inputs. However, the response curves of variables can basically reflect the effect of each variable on $O_3$ flux.

The response curve can also reveal the importance of each variable [41]. This method is called the “profile method” [25]. In this study, the importance of each variable is described by comparing the variation ranges of modeled values when a variable is varied from its minimum to maximum and other variables are fixed at their mean values given in Table 3. A larger variation in $F_o$ indicates a higher parameter importance. Excluding the effects of fluxes, the order of variable importance in model T15 is as follows: $C_{O_3} > U > Q > T_a > RH > S_W$ (Figure 5(a)). $O_3$ concentration, wind speed, and radiation are the important factors affecting ozone deposition over wheat fields. In addition, to quantitatively reflect the effect of each variable, the relative importance (RI) of each variable was calculated [39, 49, 50]. The RI of $C_{O_3}$, $U$, $Q$, $T_a$, RH, and $S_W$ are 26.7, 21.6, 20.6, 12.4, 11.3 and 7.4, respectively.

Using equation (2), it is easy to understand that the ozone concentration is the most important factor affecting ozone flux. Other factors controlling ozone deposition, such as solar radiation, temperature, humidity, and soil water content, have also been studied [47, 51]. Wind speed is another important factor affecting ozone deposition in this study. Although wind speed does not directly result in the increase of ozone deposition, it does enhance atmospheric turbulence. This causes more ozone in the air to enter crop stomata. In fact, radiation indirectly influences ozone deposition by influencing other model inputs. It results in the production of local ozone and leads to the changes in temperature, humidity, wind speed, and other environmental factors.

When considering the effects of fluxes, the order of variable importance in model T20 is as follows: $F_c > C_{O_3} > U > LE > Q > T_a$ (Figure 5(b)). As most $CO_2$ and water vapor are exchanged via the same pathway of stomata [52, 53], there are strong correlations between $F_o$ and other gas fluxes ($F_c$ and LE). Stomatal uptake might play an important role in ozone deposition. In model T20, the effect of $Q$ on ozone flux becomes small. This might be because the effects of radiation are replaced by these fluxes. In other words, when the effects of $CO_2$/$H_2O$ fluxes are taken into account, the correlations between $F_o$ and these fluxes are more significant than those between $F_o$ and radiation.

4. Conclusions

Based on the EC-measured $F_o$ and other environmental factors over a wheat field, an artificial neural network method was attempted to simulate the relationships between $F_o$ and environmental factors. Our conclusions can be summarized as follows:

1. Generally, more input variables can increase the performance of a model. By only using the routine...

---

Table 3: Input parameters and their ranges for different models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air temperature ($T_a$)</td>
<td>°C</td>
<td>-5.4</td>
<td>30.6</td>
<td>12.6</td>
</tr>
<tr>
<td>Relative humidity (RH)</td>
<td>%</td>
<td>14</td>
<td>100</td>
<td>57</td>
</tr>
<tr>
<td>Wind speed ($U$)</td>
<td>m·s⁻¹</td>
<td>0</td>
<td>7.2</td>
<td>3.6</td>
</tr>
<tr>
<td>Ozone concentration ($C_{O_3}$)</td>
<td>Ppb</td>
<td>1.0</td>
<td>124.2</td>
<td>62.6</td>
</tr>
<tr>
<td>Global radiation ($Q$)</td>
<td>W·m⁻²</td>
<td>0</td>
<td>924</td>
<td>462</td>
</tr>
<tr>
<td>Sensible heat flux ($H$)</td>
<td>W·m⁻²</td>
<td>-100</td>
<td>300</td>
<td>100</td>
</tr>
<tr>
<td>Latent heat flux ($LE$)</td>
<td>W·m⁻²</td>
<td>0</td>
<td>560</td>
<td>280</td>
</tr>
<tr>
<td>$CO_2$ flux ($F_c$)</td>
<td>mg·m⁻²·s⁻¹</td>
<td>-2.2</td>
<td>0.2</td>
<td>-1.0</td>
</tr>
</tbody>
</table>

---

Figure 5: Responses and relative importance of each variable in the ANN models T15 and T20.
micrometeorological variables of radiation and \( O_3 \) concentration, we can create an ANN model to simulate the variation of \( F_o \). The \( R^2 \) of a model with 6 routine independent variables can exceed 0.8 for the training dataset and test dataset. \( CO_2 \) flux and water vapor flux have strong corrections to \( F_o \), and they can improve the fitness of the ANN models when they are considered as independent variables.

(2) In addition to variable combinations and training data, the model parameters (e.g., the number of neurons) are also a source of uncertainties in an ANN model. The effect of neuron number on our model fitness (MAE, RMSE, and \( R^2 \)) is obvious when it changes from 1 to 10. Empirically, one hidden layer with 17 neurons produced the best results.

(3) The ANN model is composed of complex arithmetic expressions between \( F_o \) and independent variables, an optimal ANN model can roughly reflect their inner relationships and relative importance. Global radiation, \( O_3 \) concentration, and wind speed are the important factors that affect \( O_3 \) deposition. There are strong correlations between \( F_o \) and gas fluxes (\( F_g \) and LE). The ANN method represents a valuable route for filling the gaps of a time series of \( F_o \) during the experimental period with micrometeorological methods.

Data Availability

The MS-Excel data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant number 41675147) and the National Key R&D program of China (grant number 2017YFC0503801). The author would also like to thank Accdon for providing linguistic assistance during the preparation and revision of this manuscript.

References


