Improvement of reference evapotranspiration by considering time lag effect in back propagation neural network model

Jialiang Huang[1,2+], Yuhong Guo[1,2+], Xiyu Zuo[1,2], Xin Wang[1,2], Yifei Yao[1,2], Zhitao Zhang[1,2], Youzhen Xiang[1,2], Junying Chen[1,2*]

(1. *Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas of Ministry of Education, Northwest A&F University, Yangling 712100, China*;
2. *Institute of Water-Saving Agriculture in Arid Areas of China, Northwest A&F University, Yangling 712100, China*)

Abstract: Accurate estimation of reference evapotranspiration (ET_0) is of great significance for agriculture and climate irrigation systems. However, the accuracy of estimated reference evapotranspiration using traditional methods are not ideal due to the presence of complicated nonlinear relationship between meteorological factors and reference evapotranspiration. Given that, three pre-processed techniques including Pearson, Kendall and Spearman were performed to explore time lag effect between meteorological factors and reference evapotranspiration. An imputation model for ET_0 was proposed in the context of time-lag effect using modeling method of network (BP) and multiple linear regression (MLR). Then, its estimation accuracy was compared with that from traditional model without considering time-lag effect to confirm influence of time-lag effect to estimated reference evapotranspiration. The results showed that (1) ET_0 lags behind the solar radiation (R_S) for 10~20 minutes, but in advance in the air temperature (Ta) 120~150 minutes, relative humidity (RH) 90~140 minutes and wind speed (u2) 10~50 minutes; (2) In the estimated monthly ET_0 BP model (R^2 = 0.91) is better than MLR model (R^2 = 0.86); (3) By considering time lag effect, MLR model and BP model can effectively be improved ET_0 simulation accuracy, increasing 4% to 6% and 16% to 22%, but BP model is more preferable. All our preliminary results throw light on time lag effect is one of the important variables of the application model simulation reference evapotranspiration.

Keywords: reference evapotranspiration; meteorological factors; back propagation neural network model; time lag effect
DOI: 10.33440/j.ijpaa.20220501.188

Citation: Huang J L, Guo Y H, Zuo X Y, Wang X, Yao Y F, Zhang Z T, Xiang Y Z, Chen J Y. Improvement of reference evapotranspiration by considering time lag effect in back propagation neural network model. Int J Precis Agric Aviat, 2022; 5(1): 10–20.

1 Introduction

Evapotranspiration is one of the most important hydrological processes of land water circulation[1], affects surface water and energy balancing processes, contact soil, biological, and atmosphere[2]. Meanwhile, Evapotranspiration is also an important part of the climate system, including crop demand estimates, farmland irrigation water utilization efficiency, ecological process analysis, regional dry and wet assessment, surface ecological environment. The formation and evolution of surface ecological environment is based on evapotranspiration[3-4]. Therefore, for the study of evapotranspiration, it has always been one of the hot problems of domestic and foreign subject research.

Due to its importance, there are currently a variety of evapotranspiration measuring methods, such as on-site direct measurement methods based on different types of steamer and evaporating dish evaporator[5-6], but the requirement of the experiment is too high time-consuming long, not suitable for large-scale research; depending on the meteorological measurement method, such as energy balance, the vortex is also used to measure the actual evapotranspiration, but these methods are very expensive and complicated, which limits their wide applicability[7]. Therefore, various analytical formulas for indirect estimation reference crops are still the preferred method.

The Penman-Monteith equation recommended by Food and Agriculture Organization of the United Nations (FAO) is a evaporator estimate model with high precision, high applicability and high reliability[8]. However, this method for calculating ET_0 is needed meteorological data requiring numerous parameters[9], some parameters are not regularly measured, often needed to estimate or assume[10]. In addition, most of the evaporate estimation models are measured by the month or week[11-12], some areas of wind speed, dew point or cloud change, the monthly or weekly average value cannot be truly on behalf of the environmental evapotranspiration capability of each time period. Therefore, it is easy to cause errors in ET_0 estimation[13].

The hysteresis between transpiration (or liquid flow) and meteorological factors have been widely reported[14]. These hysteresis occurred in a single plant or even smaller units (such as branches)[15]. ET_0 is defined as the evapotranspiration during

Received date: 2022-10-12 **Accepted date:** 2022-12-03
Biography: Jialiang Huang, Master student, research interests: Irrigation agriculture, Email: huangjialiang@nwafu.edu.cn; **Yuhong Guo**, Master student, research interests: Agricultural Technology & Equipment, Email: 2020050977@nwafu.edu.cn; **Xiyu Zuo**, Master student, research interests: Agricultural Technology & Equipment, Email: zuoxiyu123@126.com; **Xin Wang**, Master student, research interests: Efficient utilization of crop water and fertilizer, Email: 1123714182@qq.com; **Yifei Yao**, PhD, research interests: Soil moisture assimilation and monitoring, Email: yifeiyao@nwsuaf.edu.cn; **Zhitao Zhang**, PhD, research interests: Regional scale drought monitoring and evapotranspiration estimation based on satellite and UAV remote sensing, Email: zhitaozhang@126.com; **Youzhen Xiang**, PhD, research interests: Regional scale drought monitoring and evapotranspiration estimation based on satellite and UAV remote sensing, Email: youzhenxiang@nwsuaf.edu.cn;
***Corresponding author: Junying Chen**, PhD, research interests: Remote sensing monitoring and prediction of soil water and salt information, Email: junyingchen@nwafu.edu.cn.

the growth period of reference crops with sufficient water supply, consistent plant height and complete coverage in large scale space. Among them, the reference crop is an ideal crop, representing a specific surface form, commonly used alfalfa, ryegrass, grass and so on[16]. Therefore, the response of ET_0 to meteorological factors should also hysteresis. However, as far as we know, ET_0 has less research on the response of meteorological factors[17].

ET_0 was calculated based on meteorological data from October 2020 to May 2021, and the nonlinear relationship between ET_0 and meteorological factors was established by using multiple linear regression model and BP neural network model. Time lag was clearly considered as an important variable. Specifically, our research objectives are as follows: i) use univariate analysis to determine the correlation between ET_0 and meteorological factors; ii) Analyze the time lag effect of meteorological factors and ET_0 and determine the lag time; iii) Multiple linear regression model and BP neural network model are used to simulate the dynamic process of ET_0, and the importance of time lag effect in modeling is considered quantitatively.

2 Materials and Methods

2.1 Site description

The experimental site was located in the Institute of Water Saving Agriculture in Arid Areas of China (108°4'20"E, 34°17'42.17"N), Northwest A&F University, Yangling, Shaanxi Province. It is in the middle of Guanzhong Plain and has a warm temperate semi-humid and semi-arid climate, with an average annual temperature of 12.9°C, frost-free period of 211 days, average annual sunshine of 2163.8 hours and annual total solar radiation of 114.8 kcal cm^{-2}. The average annual precipitation is 635.1 mm.

2.2 Experiment Design

The total area of the experimental site was 32.5 m×10.5 m, which was divided into 12 plots (4 m× 4 m) (Figure 1), and a 1.5-m-wide guard row was set between the plots to eliminate the marginal effect. The wheat (Xiaoyan 22, a wheat variety in Shaanxi Province) was sown on September 25, 2020 and harvested on June 12, 2021. In the whole growth period, only 20 kg compound fertilizer was used as base fertilizer when the wheat was sown. During the experiment, the irrigation in the plot was strictly controlled and consistent so that the wheat was not affected by drought stress. The plot layout is shown in Figure 1.

2.3 Data Acquisition

As is shown in Figure 2, The experimental site was equipped with a meteorological station with a net radiation monitoring system (Apogee SN-500), an air temperature-humidity sensor (Rotronic HC2AS3) and a soil heat flux sensor (180M), which can synchronously monitor the net radiation (Rn, W/m^2), atmospheric temperature (Ta, °C), relative humidity (RH, %) at 2 m above the ground surface and soil heat flux (SHF, W/m^2) at the depth of 0.05 m. The average values of Tc and its corresponding environmental factors were measured and recorded every 10 min. At the same time, long-term rainfall data were obtained from Yangling National General Weather Station near the experimental site. The experimental data were acquired from October 1, 2020 through May 31, 2021. Therefore, a total of 33,840 sets of valid data were collected.

Figure 1 Overview of the experimental site

Figure 2 Meteorological station

2.4 Data Processing

2.4.1 Reference evapotranspiration

The reference evapotranspiration (ET_0) is calculated according to the Penman-Monteith formula after normalization in FAO:

$$ET_0 = \frac{0.408 \times \Delta(R_n - G) + \gamma \times \frac{900}{T+273} \times u_2(e_s - e_a)}{\Delta + \gamma \times (1 + 0.34 \times u_2)} \quad (1)$$

where: ET_0 refers to the cube evapotranspiration amount (mm/day); Δ refers to the slope of the tangent line of the temperature-saturated vapor pressure relationship curve at the mean temperature (kPa/°C); R_n refers to the net solar radiation (W/m^2); G refers to soil heat flux (W/m^2); γ refers to the thermometer constant (kPa·°C); u_2 refers to a wind speed of 2 meters (m/s); e_s refers to the saturation vapor pressure (kPa); e_a refers to the actual vapor pressure (kPa); T refers to the average temperature (°C).

2.4.2 Time-lagged Correlation Analysis

The time lag relationship between the ET_0 and the meteorological factor is analyzed by the displaced correlation method. Specifically, ET_0 is similar to the diurnal variation of meteorological factors. The ET_0 time series are fixed, and the time series of meteorological data are moved by 10 minutes, and the correlation coefficient of the two sequences is considered to be the lag time when the correlation coefficient is maximized. Lag time Calculation formula:

$$R = \max(R_{-n}, R_{-(n-1)}, \cdots, R_{-1}, R_0, R_1, \cdots, R_{(n-1)}, R_{-n}) \quad (2)$$

Take the maximum value as a lag correlation coefficient. The time lag between ET_0 and the environmental factor is 10* n minutes. In this paper, positive lag time between ET_0 and meteorological factors means that ET_0 appears or reaches its peak

earlier than meteorological factors in a day, while negative lag time between ET_0 and meteorological factors means that ET_0 appears or reaches its peak later than meteorological factors in a day.

2.4.3 Related coefficient calculation method

Three commonly used methods to determine the correlation are the Pearson coefficient method, Kendall coefficient method and Spearman coefficient method. Pearson correlation coefficient is a statistical method that can quantitatively measure the linear relationship between random variables[18]. Its output range is generally in [–1, 1], 1 indicates the relationship between x and y in which all data points fall on a straight line and y increases as x increases; conversely, if the coefficient is –1, it means that all data points fall on the straight line, and y decreases with the increase of x. 0 indicates that there is no linear relationship between x and y variables. Here, we select the feature vector and meteorological factor as the independent variable. Pearson correlation coefficient is used to analyze the feature quantity with large absolute value of linear relationship with meteorological factor. If the output value is larger, the linear relationship of correlation features will be stronger, as shown in the following formula:

$$\rho(x,y)=\frac{\sum_i(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i(x_i-\bar{x})^2}\sqrt{\sum_i(y_i-\bar{y})^2}} \tag{3}$$

where, $\rho(x, y)$ is the correlation coefficient of the random variable x and y; $\sum_i(x_i-\bar{x})(y_i-\bar{y})$ is the variance of the random variable x and y ; $\sum_i(x_i-\bar{x})^2$ is a variance of the random variable x and $\sqrt{\sum_i(y_i-\bar{y})^2}$ is a variance of the random variable y.

Kendall ranking relationship is used to measure the rank correlation between the two observation indicators, proposed by Maurice Kendall[19], can be used to measure the correlation between the two attribute variable levels under the conditions of the pairing design, usually used Express τ. Essentially, the more similar the rank between the two variables is, the greater the Kendall correlation coefficient, therefore is Kendall correlation coefficient is a general nonparametric statistic, which does not depend on the distribution of two variables. The value range of Kendall's rank correlation coefficient is [–1, 1], and the calculation formula is as follows:

$$\tau=\frac{n_c-n_d}{\sqrt{(n_0-n_1)}\sqrt{(n_0-n_2)}} \tag{4}$$

where, $n_0=n(n-1)/2$, $n_1=\sum t_i(t_i-1)/2$, $n_2=\sum \mu_i(\mu_i-1)/2$, n is the number of samples, n_c represents the number of consistent pairs of changes, n_d represents the number of inconsistent pairs of changes, t_i and μ_i represent the values of two variables x and y, respectively.

Spearman rank correlation is defined as a Pearson correlation coefficient between level variables, which is a general non-parameter statistic proposed by Charles Spearman, usually represented r_s[20]. Similar to the above relationship coefficient, the Spearman rank correlation is [–1, 1]. Rank can be understood achievement is a sequence or sort, then it is solved according to the sort position of the original data. This characterization does not have the restriction requirements when conjunction of Pearson correlation factors, so the Spearman rank correlation coefficient can be used the continuous random variables can also be used for discrete ordered random variables. The calculation formula is:

$$r_s=1-\frac{6\sum d_i^2}{n(n^2-1)} \tag{5}$$

where, d_i indicates that the two variables of the variable are sorted separately, and n is the number of samples.

2.5 Establishment of the model

2.5.1 Multiple linear regression model

Multiple linear regression model (MLR) is a statistical method for analyzing data[21]. The purpose is to understand whether two or more variables are correlated, their correlation direction and intensity, and to establish a mathematical model to observe a particular variable. More specifically, multiple regression analysis can help people understand the amount of variables due to only one variable change. In general, by multiple regression analysis we can estimate the conditions of variables due to the requirements of variables. Multiple regression analysis is a model that establishes the relationship between variable y and the self-variable x.

$$h(x_0,x_1,...,x_n)=\sum_{i=0}^{n}\theta_i x_i \tag{6}$$

where, $h(x_0, x, x_2, \ldots, x_n)$ is feature parameters; x_i is feature variable.

For linear regression models, we often use mean square errors as a loss function. The algebraic method of the loss function is shown below:

$$J(\theta_0,\theta_1,\cdots,\theta_n)=\sum_{i=0}^{n}(h(x_0,x_1,\cdots,x_n)-y_i)^2 \tag{7}$$

There are two commonly used methods to find θ_j parameters when the loss function is minimized: the gradient descent method and the least square method. Here we use the least square method. Then the formula of θ_j is as follows:

$$\beta_n=\frac{n\sum x_i y_i-\sum x_i\sum x_i}{n\sum x_i^2-(\sum x_i)^2} \tag{8}$$

2.5.2 BP Neural Network Model

The BP neural network model is a one-way multi-layer feedforward network, which utilizes the error reverse propagation algorithm, and the parameter adjustment is implemented according to the gradient decline[22]. The nonlinear mapping ability of the model is strong, which can adjust the internal parameters of the system according to the error between the output results and the measured values, and carry out the operation of signal forward propagation and error back propagation repeatedly according to the accuracy requirements, and finally get the output results. The topology structure of the model is composed of input layer, hidden layer and output layer. The input layer consists of 4 environmental factors; The number of network layers of the hidden layer is set to 10, and the hyperbolic tangent activation function is adopted. The output layer is ET_0 and uses the identity activation function. The number of training is 1000 times, the learning rate is 0.001, and the target error is 1e–10.

There are two parts in the BP neural network model: firstly, the signal is propagated from the forward propagation input to the output, and the second part is the error of the neural network model output value and the desired value from the output to the input reverse adjustment. In the forward propagation process, the input signal is input from the input layer, and after the implied layer, the output layer generates an output signal. At the output layer, the output value, and the expected value, if the difference does not satisfy the setting condition, then this error requires reverse propagation to adjust the adjustment value threshold. In the reverse adjustment phase, the error is forwardly propagated by the output, and the weight and threshold of the implicit layer to the

output layer, the input layer to the weight and threshold value of the hidden layer are adjusted.

The specific steps of the training of neural networks are shown below.

1. Signal forward propagation

The output of the hidden layer is shown in the formula:

$$H_j = g(\sum_{i=1}^{n}(\omega_{ij}x_i + \alpha_j)) \tag{9}$$

where, x_i is the environmental factors sequence after pretreatment; ω_{ij} is the weight from input layer to hidden layer; α_j is the bias from input layer to hidden layer, and the excitation function $g(x)$ is Sigmoid function: $g(x) = \frac{1}{1+e^{-x}}$.

The output of the output layer O_k, is shown in the formula:

$$\begin{cases} l = \sqrt{n+m} + \xi \\ O_k = \sum_{i=1}^{n}(\omega_{jk}H_j + \beta_j) \end{cases} \tag{10}$$

where, ω_{jk} is the weight from the hidden layer to the output layer; β_j is the bias from hidden layer to output layer; ξ is a constant of 1-10.

The error E is shown in the formula:

$$E = \frac{1}{2}\sum_{k=1}^{m}(y_k - O_k)^2 \tag{11}$$

1. Error back propagation

To minimize the error, Gradient descent method is used in this paper. The iterative updating method of weights from hidden layer to output layer and the iterative updating method of weights from input layer to hidden layer is shown in the formula:

$$\begin{cases} \omega_{jk}^{t+1} = \omega_{jk}^{t} + \eta H_k^t(1-y_k^t)e_k^t \\ \omega_{ij}^{t+1} = \omega_{ij}^{t} + \eta H_j^t(1-H_j^t)x_i^t\sum_{k=1}^{m}\omega_{jk}^t y_k^t(1-y_k^t)e_k^t \end{cases} \tag{12}$$

2.6 Evaluation indicator

In order to compare the BP neural network model and the MLR model simulation ET_0 and ET_0 calculated using the PM formula, we use the decision coefficient (R^2) and average absolute error (MAE) three parameters to evaluate the model performance. The modeling accuracy of the model is used to determine the coefficient (R^2) and the relative error (RMSE), the model's verification accuracy employs a relative error (RMSE) and average absolute value error (MAE).

Accuracy of training set is evaluated by two indicators: the decision coefficient (R^2) and the root mean square error (RMSE). Among them, the value of R^2 ranges from 0 to 0-1; The effect is excellent. The RMSE characterizes the deviation of the predicted value and the measured value, the closer the value, indicating that the higher the predicted accuracy of the model, the stronger the prediction ability. R^2 and RMSE calculation formulas are:

$$R^2 = \frac{\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{13}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\tilde{y}_i - y_i)^2}{n}} \tag{14}$$

where, $\tilde{y}_i$ is the predicted value of ET_0; y_i is measured value of ET_0; $\bar{y}$ is the average value of ET_0.

The root mean square error (RMSE) and mean absolute value error (MAE) were used to evaluate the errors in the model simulation. When indicating the simulation accuracy, the smaller the value of the evaluation index is, the higher the simulation accuracy is. RMSE calculation method is the same as the above, and the MAE calculation formula is as follows:

$$\mathrm{MAE} = \frac{\sum_{1}^{n}|\tilde{y}_i - y_i|}{n} \tag{15}$$

The overall framework of this paper is shown in Figure 3.

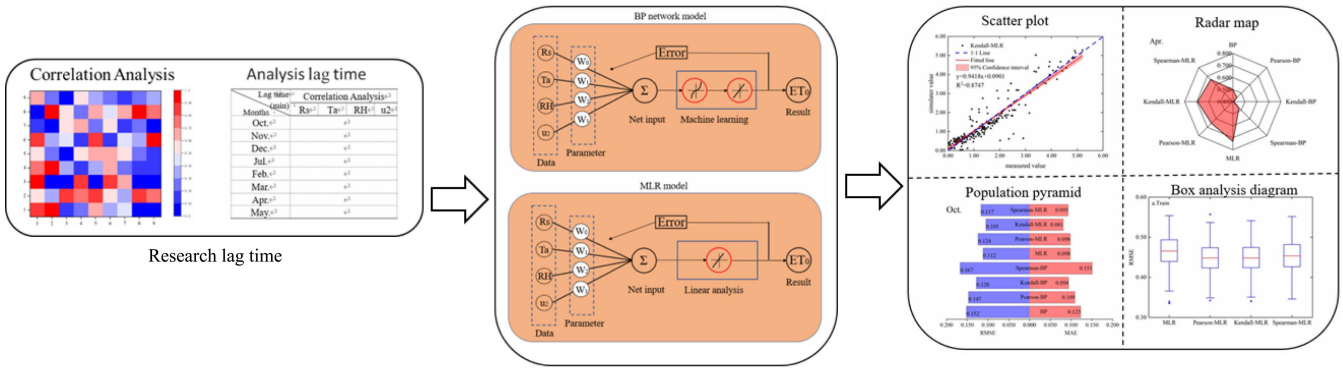

Figure 3 The flow diagram

3 Results

3.1 Analysis of the correlation between meteorological factors and ET_0

The following figures (Figure 4) were the heat map of correlation between ET_0 and meteorological factors solved by Pearson, Kendall and Spearman methods. As a result Pearson, Kendall and Spearman methods all showed that among the correlation between ET_0 and meteorological factors, the correlation between ET_0 and Rs was the best, which was above 0.783, 0.626 and 0.814, respectively. The Pearson method indicated that RH, Ta and u2 were second (R= –0.786~ –0.629), third factors (R = 0.555~0.679) and fourth (R = 0.474~0.675) factors of affecting ET_0. However, both Kendall and Spearman methods indicated that Ta, u2 and RH respectively were second (R = 0.165 ~0.446 and 0.275 ~0.645), third (R = 0.425~0.513 and 0.606~0.725) and fourth (R= –0.332~–0.54 and –0.418~–0.756) factors of affecting ET_0.

3.2 Analysis of time lag effect between environmental factors and ET_0

Transpiration was mainly driven by meteorological factors, solar radiation, temperature, humidity, and wind speed as the main meteorological factor. A large number of studies had shown that the daily change process of the transpiration and meteorological factor is inconsistent, and there was a peak of peak between the two and time lag effect between environmental factors and ET_0. The meteorological factor and ET_0 time lag were calculated using the misplaced correlation method. The result was shown in Figure 5.

It can be found in Figure 4 that the correlation between single factor meteorological factors and ET_0 was improved by the three correlation coefficient methods, among which the correlation

coefficient between ET_0 and Ta was increased the most, and the correlation were increased by 0.14, 0.16 and 0.19 on average respectively, followed by RH were increased 0.11. Possibly because evapotranspiration was more sensitive to Ta and RH.

It was shown that time lag effect can effectively improve the correlation between environmental factors and ET_0, especially atmospheric temperature and relative humidity. Considering time lag effect, all three methods showed that solar radiation was the most important factor affecting ET_0, followed by relative humidity and atmospheric temperature, and finally wind speed. Time lag effect between plant transpiration and environmental factors was caused by the unsynchronization between transpiration and meteorological factors. In other words, the level of transpiration may be regulated by the previous meteorological factors, and the current meteorological factors may also significantly affect the subsequent transpiration.

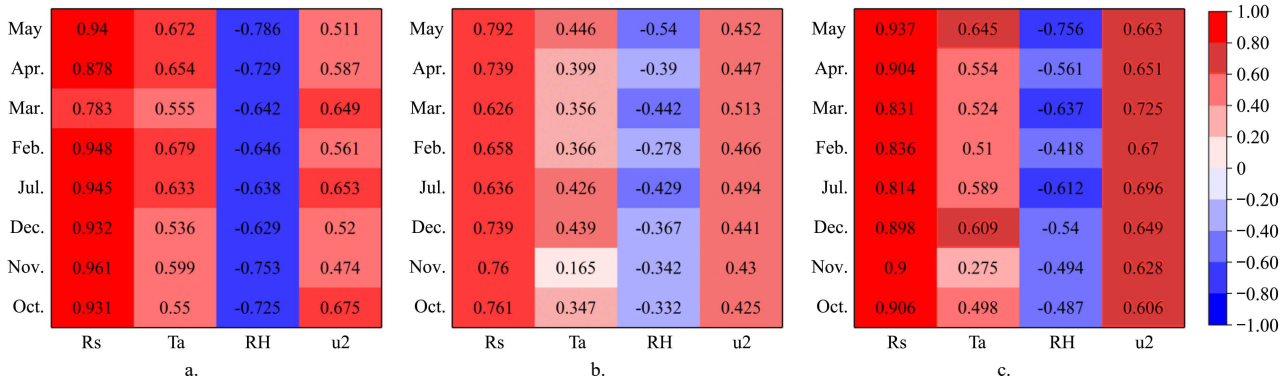

Fig.4a The heat map of the correlation between ET_0 and meteorological factors solved by the Pearson method; Fig.4b The heat map of the correlation between ET_0 and meteorological factors solved by the Kendall method; Fig.4c The heat map of the correlation between ET_0 and meteorological factors solved by the Spearman method

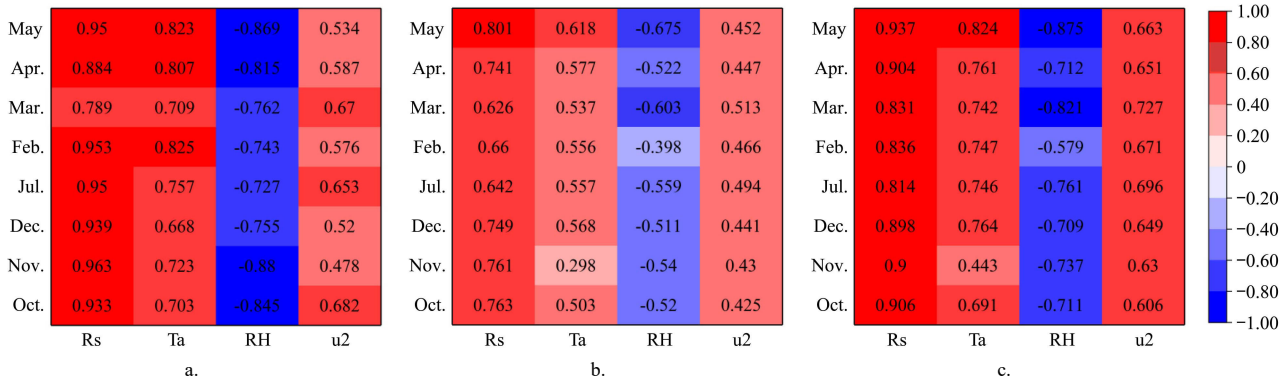

Fig.5a The heat map of the correlation between ET_0 and meteorological factors solved by the Pearson method with considering time lag effect; Fig.5b The heat map of the correlation between ET_0 and meteorological factors solved by the Kendall method with considering time lag effect; Fig.5c The heat map of the correlation between ET_0 and meteorological factors solved by the Spearman method with considering time lag effect

Table 1 Calculate the time lag of ET_0 and weather factors

Lag/min / Months	Pearson				Kendall				Spearman			
	Rs	Ta	RH	u2	Rs	Ta	RH	u2	Rs	Ta	RH	u2
Oct.	–10	120	90	20	–10	130	130	0	10	130	140	0
Nov.	–10	120	100	20	–10	140	140	0	10	140	140	30
Dec.	–10	120	110	0	–10	110	130	0	–10	120	130	0
Jul.	–10	110	100	0	–10	110	110	0	–10	120	120	0
Feb.	–10	120	120	40	10	140	140	0	20	150	140	0
Mar.	–10	140	120	50	0	150	130	0	10	150	130	30
Apr.	–10	140	100	10	–10	140	120	0	0	140	110	0
May	–10	130	90	30	–10	130	90	0	–10	130	100	0

It can be found that the three correlation coefficients calculate the lag time is basically consistent by Table 1. The Pearson method showed that ET_0 was lagging in Rs for 10 minutes, in advance Ta, RH and u2 of about 125±15, 105±15 and 0± 10 minutes; the Kendall method showed that ET_0 is lagging behind Rs about 0±10 minutes, in advance Ta, RH and u2 of about 130± 20, 125±25 and 0 minutes; the Spearman method showed that ET_0 is lagging behind Rs of about 5±15 minutes. In advance Ta, RH and u2 of approximately 135±15, 120±20 and wind speed 0± 30 minutes.

3.3 Simulated ET_0 Process

In order to identify the influence of time lag effect on simulated ET_0, four schemes were established according to Pearson, Kendall and Spearman methods to identify time lag effect: 1) in scheme 1 Rs, u2, Ta and RH were used as input variables without considering time lag effect to establish MLR and

BP models; 2) in scheme 2, Pearson's Rs, u2, Ta and RH were used as input variables with considering time lag effect to establish MLR and BP models; 3) in scheme 3, Kendall's Rs, u2, Ta and RH were used as input variables with considering time lag effect to establish MLR and BP models; 4) in scheme 4, Spearman's Rs, u2, Ta and RH were used as input variables with considering time lag effect to establish multiple regression and BP neural network models.

The sample size was randomly divided into a modeling set and validation set in a ratio of 5:5. As mentioned above, while applying the climate data model, the 10-minute ET_0 was converted into the daily average ET_0. It meant that the cumulative value was calculated every 144 ET_0 values to estimate the ET_0 of the day. Then the fitting effect was evaluated according to the monthly scale. The following figure showed the ET_0 estimation results of different schemes.

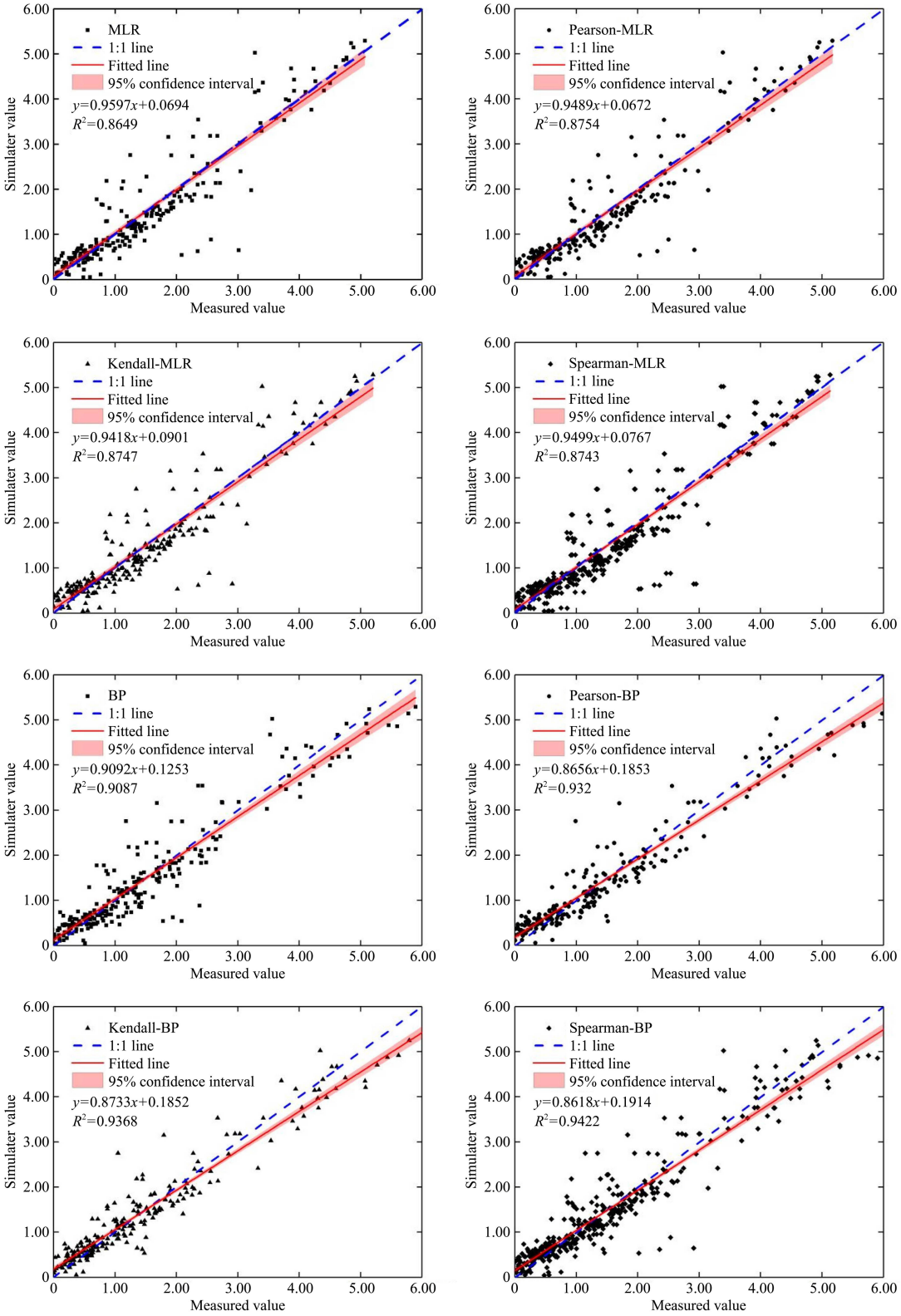

Figure 6 Scenarios simulation values and true prodigal points

To illustrate the fitting performance of the four schemes, the scatter plot (Figure 6) was shown to describe the measurement and simulation value of ET_0. It can be clearly indicated that the fit effect of BP model is better than MLR model, regardless of considering time lag effect, indicating that BP model was capable of well described in the training stage. The decision coefficient (R^2) of the four BP models above was 0.90 or more, which was much higher than MLR model. This was because of the nonlinear relationship between ET_0 and meteorological factors, and the BP model was a lot great when constructing complex and nonlinear relationships.

As Figure 7 shown, by considering time lag effect, MLR and BP model fitting effect had improved except in October. RMSE of Pearson-MLR, Kendall-MLR and Spearman-MLR fell by 12.47%, 11.29%, and 10.62%, respectively with considering time lag effect. RMSE of Pearson-BP, Kendall-BP and Spearman-BP fell by 12.11%, 16.28%, and 12.86%, respectively with considering time lag effect. In the monthly ET_0, considering the time lag effect can effectively reduce the RMSE, indicating that the time lag effect was one of the important influencing factors of simulating evapotranspiration transpiration, and the lag should effectively improve the estimation of ET_0 model accuracy.

In order to identify the fitting performance of the test data, cross-verification analysis of measurement and simulation comparison was performed on the test data set (Figure 8). It can be found that most of RMSE and MAE of MLR and BP models by considering time lag effect had decreased. The simulation accuracy of ET_0 increased by 4%~6% and 16%~22%. These growths clearly indicate that considering time lag effects can effectively improve energy efficiency to improve ET_0 model accuracy.

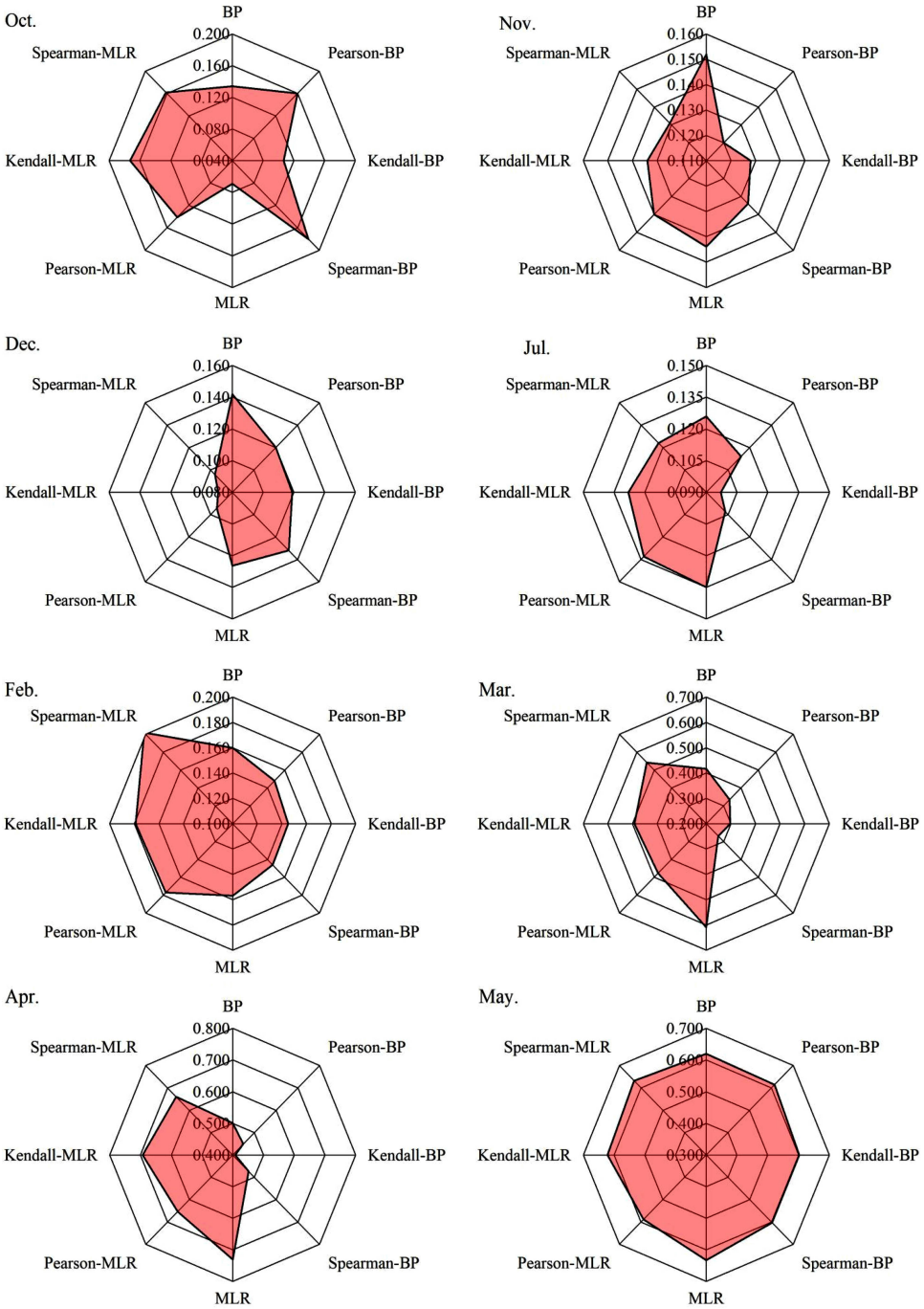

Figure 7 Radar map of ET_0 RMSE in the training phase

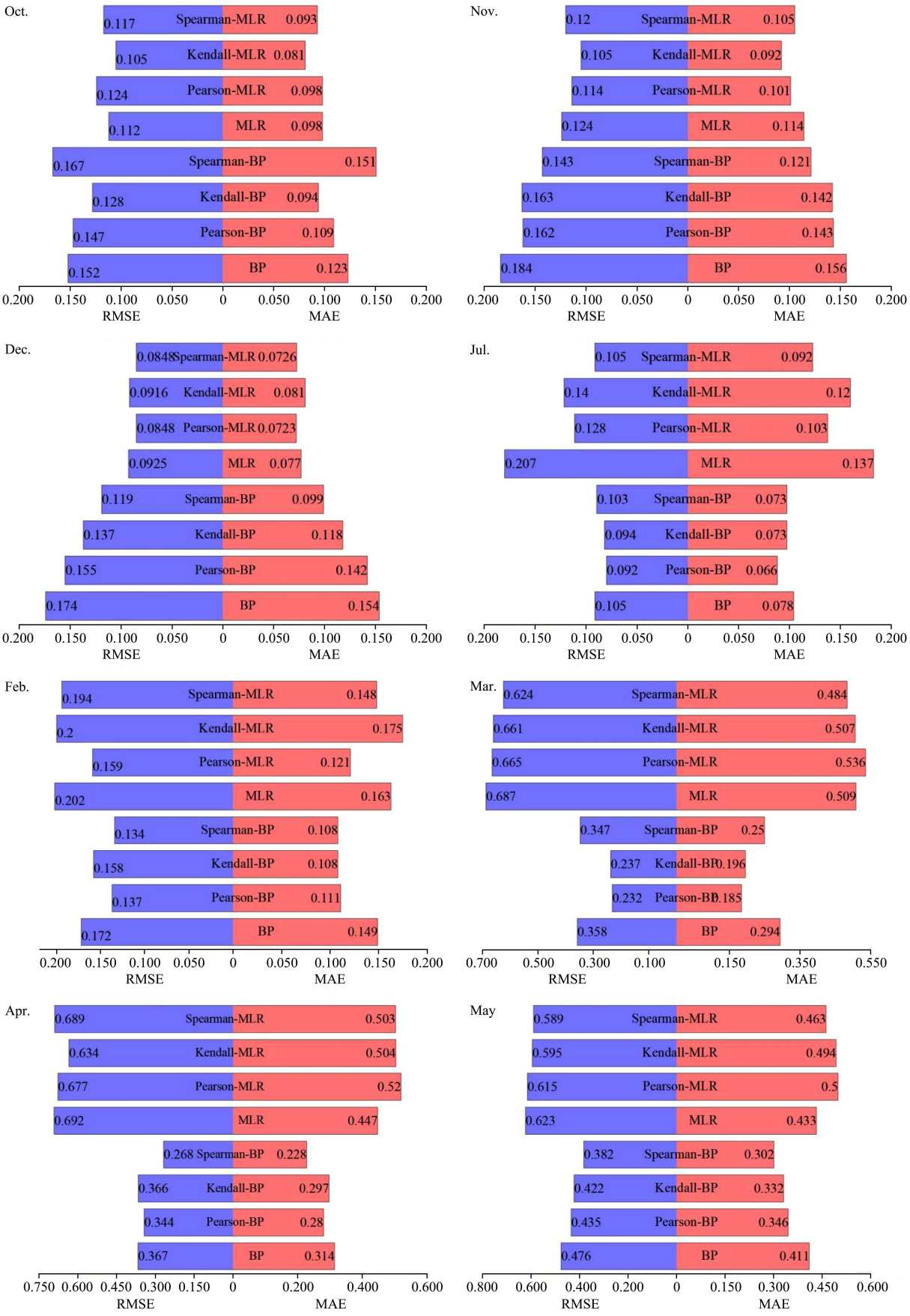

Figure 8 Modeling results under different solutions

Due to the random selection of samples for modeling and verification, the results may be haphazard, and ultimately it cannot be proved that the combination of time lag effect can effectively improve the accuracy of simulated ET_0. In order to prove the correctness of the conclusion, 500 repeated tests were carried out on MLR and BP model, and the results were presented in a box diagram. As can be seen from Figure 9, the relative errors of the multiple linear regression model and the BP neural network model

fluctuate within a range. Considering the time lag effect, the relative errors decrease compared with those without, indicating that the time lag effect plays an important role in the simulation of ET_0.

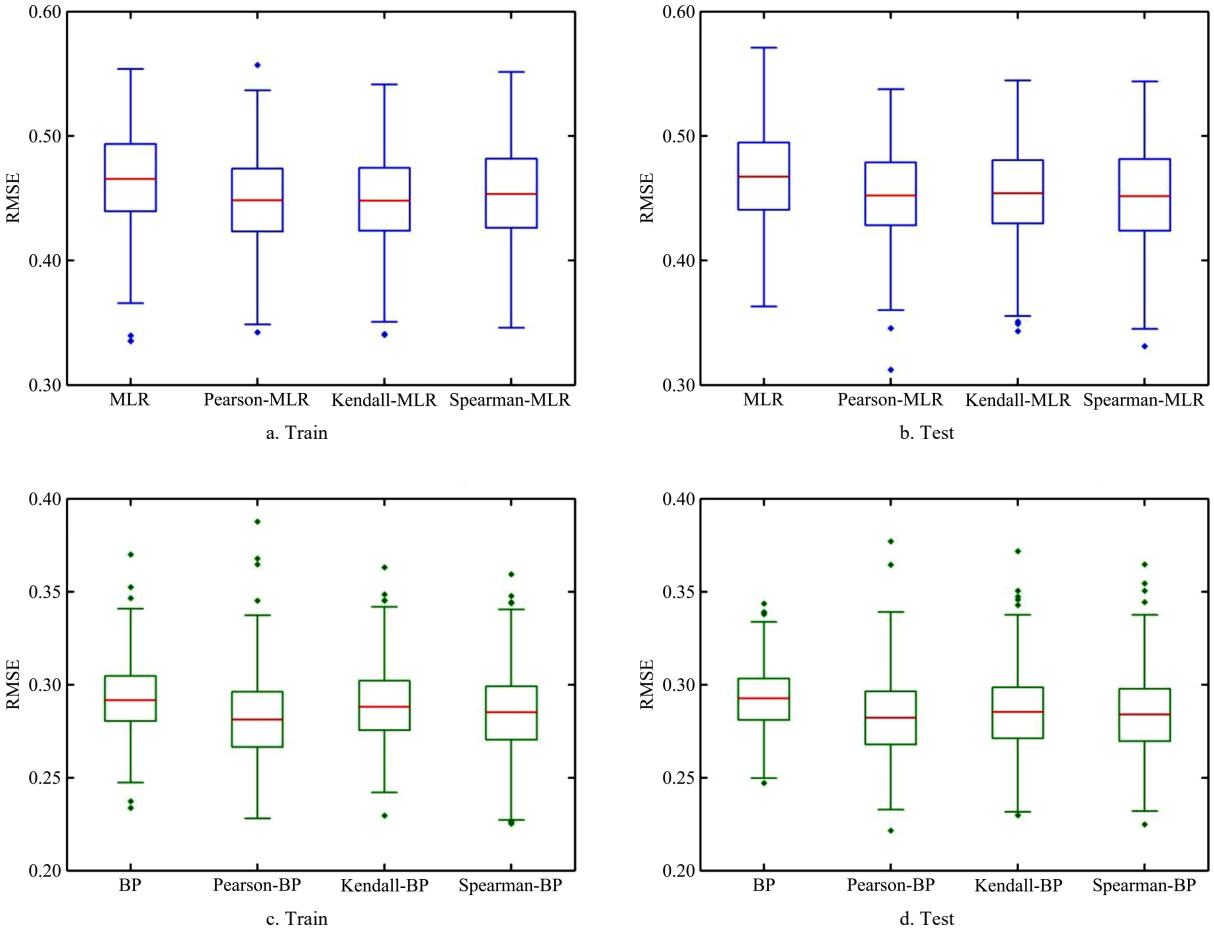

Figure 9 RMSE box diagram of 500 repetitions

4 Discussion

4.1 Time lag effect between ET_0 and meteorological factors

As mentioned earlier, three correlation coefficients used in this study determine the correlation of meteorological factors and ET_0, obviously, air temperature, solar radiation, relative humidity and wind speeds affect the effective parameters of the ET_0 process. The above parameters are generally considered as predictors for each site when estimating ET_0. This is basically consistent with the previous research conclusions. Luo analyzed the correlation between ET_0 and meteorological factor in Chongqing, and ET_0 was mainly affected by the sunshine and temperature, and its change curve was single peak type[23]. Zhang uses differential equations to attribute ET_0 in climate variables, and believes that ET_0 in the arid area is most sensitive to water, pressure, followed by solar radiation, temperature and wind speed[24]. Under the interaction of multiple climate factors, the temperature is the main control variable of ET_0 changes, while ET0 is most sensitive to RH changes[25].

Time lag between meteorological factors and ET_0 sequences is identified by the dislocation correlation method. As can be seen from Table 1, ET_0 lags behind solar radiation by about 10 minutes, and is about 120, 130 and 0~30 minutes ahead of air temperature, relative humidity and wind speed. After the time lag effect is considered, the correlation between meteorological factors and ET_0 increases, especially atmospheric temperature and relative humidity. This may be because air, as a continuous and stable system, needs a long time to react to external influences. There have been many reports of plants on time lag effect of evapotranspiration transpiration and meteorological factors. Di analyzed the Populus flow velocity and the corresponding temperature, air humidity, net radiation and water vapor pressure difference. The SAP flow velocity was 30 to 90 minutes ahead of air temperature, 30 to 90 minutes ahead of air humidity, 50 to 120 minutes ahead of air temperature, and 10 to 40 minutes behind net radiation [26]. Oguntunde analyzed the time lag between evapotranspiration and vapor pressure difference, and the results showed delays of 43, 46 and 75 minutes in November, December and January, respectively[27]. In addition, many papers have reported the existence of periodic and seasonal delays between evapotranspiration and meteorological factors. Bo studied the time-lag characteristics of seed-maize liquid flow and believed that the time-lag of maize liquid flow was significantly different in different growth stages[28]. Guo believes that the crop evapotranspiration in different growth stages shows different lag models[15]. In this study, the time lag between meteorological factors and ET_0 in different months is stable and regular, without seasonal or periodicity. This may be because ET_0 itself assumes a state of transpiration that is growing well.

4.2 Influence of time lag effect on evapotranspiration model

The BP model is equipped with a higher equivalent of the traditional MLR model, which can be used for prediction and simulation of ET_0. The BP model gives the improved consistency between the measured values and the analog values, mainly because they can find any complex and non-line input-output relationships and simulate from the data, without the need for any

restrictions on the entire process[15]. Liu reported the superior performance of the machine learning model in simulated transpiration[29]. Kisi used an artificial neural network model to estimate ET_0 and the results show that it is better than the traditional empirical method to estimate ET_0[30]. Granata established M5P Regression Tree, Bagging, Random Forest and Support Vector regression model to evaluate ET_0, and the results show that the machine learning model is better than the traditional model in predicting evapotranspiration[31].

Some researchers also take the time lag effect into account when establishing various evapotranspiration models. Regression analysis showed that the coefficient of determination of microclimate factors and poplar SAP flow velocity were 0.855 and 0.903 respectively without and with time lag effect, and the coefficient of determination was increased by 2.04% when time lag effect was included[26]. Ferhrman predicted the future ET_0 trend based on deep learning, and believed that the accuracy of ET_0 prediction could be improved by considering the lag effect[32]. In this study, the MLR model and BP model significantly improved the fitting accuracy of ET_0 due to the consideration of time lag effect, especially in the month with large ET_0, indicating that the transpiration process of water vapor absorbing energy from the evapotranspiration surface to the air takes a certain time, rather than instantaneous.

However, there are some deficiencies in this paper. Since the time lag between meteorological factors and ET_0 is several hours or even tens of minutes, it may be necessary to monitor meteorological data every few minutes, while general meteorological stations use hourly meteorological data. In addition, there may be a certain collinearity between meteorological data, that is, the correlation between meteorological factors. The increase in solar radiation will cause the increase in atmospheric temperature, and also affect the decrease of relative humidity.

Transpiration is an important component of the water balance of any forest ecosystem. Accurate estimation of evapotranspiration is a prerequisite for understanding its dynamics and its relationship to various environmental variables, which has important management implications for water supply and a variety of other ecological functions or services. Our study clearly shows that BP models can improve estimation of transpiration, especially when time lag effects are taken into account. These findings contribute to the accurate prediction of transportation processes, thus promoting the development of ecological science. In addition, our results may help to better assess the impact of climate change on ecological processes, as the variables we identified in this study may be affected by future climate change scenarios.

5 Conclusion

Monthly evapotranspiration is estimated for the period October 2020 to May 2021 in this study. Therefore, based on meteorological data (Rs, Ta, RH and u2) and PM-ET_0, Pearson, Kendall and Spearman pretreatment techniques were used to analyze the correlation between meteorological data and ET_0. And the time lag between them was calculated by the dislocation correlation method. The MLR model and BP model are used as independent models respectively by considering two different types of scenarios with and without time lag effect. The results of this study can be summarized as follows: In general, ET_0 lags behind solar radiation, but precedes atmospheric temperature and relative humidity. Furthermore, the correlation between ET_0 and atmospheric temperature and relative humidity is significantly improved when time lag effect is considered. The BP model is superior to the MLR model in predicting ET_0 without considering the time lag effect, because it can identify the complex and nonlinear processes between evapotranspiration and meteorological factors. In addition, for the MLR model and BP model, considering the time lag effect can effectively reduce the error and improve the accuracy of ET_0 simulation. More importantly, we find that the estimation of ET_0 can be greatly improved by taking into account the time lag effect in spring and summer. We hope that through more accurate projections of transpiration, our results may have useful implications for advancing the science of ecohydrology and improving the assessment of climate change impacts.

Author Contributions

Jialiang Huang and Yuhong Guo conceived and designed the experiments, analyzed the data, prepared figures and tables, authored or reviewed drafts of the paper, and approved the final draft.

Junying Chen conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Yifei Yao, Youzhen Xiang and Zhitao Zhang proofread and revised the drafts of the paper, and approved the final version.

Xiyu Zuo and Xin Wang conceived and designed the experiments, performed the experiments, reviewed drafts of the paper, and approved the final draft.

Funding

This study was financially supported by the National Natural Science Foundation of China (51979232), the National Natural Science Foundation of China (52179044) and Natural Science Foundation in Shaanxi Province of China (2019JM-066).

[References]

[1] McCabe M F, Wood E F. Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors. Remote Sensing of Environment, 2006, 105(4): 271–285. DOI: 10.1016/j.rse.2006.07.006.

[2] Wang K, Dickinson R E. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. Reviews of Geophysics, 2012, 50(2). DOI: 10.1016/j.rse.2006.07.006.

[3] Stöckle C O, Kjelgaard J, Bellocchi G. Evaluation of estimated weather data for calculating Penman-Monteith reference crop evapotranspiration. Irrigation science, 2004, 23(1): 39–46. DOI: 10.1007/s00271-004-0091-0.

[4] Tabari H. Evaluation of reference crop evapotranspiration equations in various climates. Water resources management, 2010, 24(10): 2311–2337. DOI: 10.1007/s11269-009-9553-8.

[5] Masoner J R, Stannard D I, Christenson S C. Differences in evaporation between a floating pan and class a pan on land 1. JAWRA Journal of the American Water Resources Association, 2008, 44(3): 552–561.

[6] Yang C, Zhu Z, Tan L, et al. Analysis on evapotranspiration of maize field measured by Lysimeters in Huailai. Plateau Meteorol, 2015, 34(4): 1095–1106.

[7] Drexler J Z, Snyder R L, Spano D, et al. A review of models and micrometeorological methods used to estimate wetland evapotranspiration. Hydrological processes, 2004, 18(11): 2071–2101. DOI: 10.1002/hyp.1462.

[8] Allen R G, Pereira L S, Raes D, et al. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome, 1998, 300(9): D05109.

[9] Jabloun M, Sahli A. Evaluation of FAO-56 methodology for estimating reference evapotranspiration using limited climatic data: Application to Tunisia. Agricultural water management, 2008, 95(6): 707–715. DOI: 10.1016/j.agwat.2010.09.005.

[10] Valiantzas J D. Simple ET_0 forms of Penman's equation without wind

and/or humidity data. II: Comparisons with reduced set-FAO and other methodologies. Journal of Irrigation and Drainage Engineering, 2013, 139(1): 9–19.

[11] Droogers P, Allen R G. Estimating reference evapotranspiration under inaccurate data conditions. Irrigation and drainage systems, 2002, 16(1): 33–45. DOI: 10.1023/A:1015508322413.

[12] Cai J, Liu Y, Lei T, et al. Estimating reference evapotranspiration with the FAO Penman–Monteith equation using daily weather forecast messages. Agricultural and Forest Meteorology, 2007, 145(1-2): 22–35. DOI: 10.1016/j.agrformet.2007.04.012.

[13] Allen R G, Pereira L S, Howell T A, et al. Evapotranspiration information reporting: I. Factors governing measurement accuracy. Agricultural Water Management, 2011, 98(6): 899–920. DOI: 10.1016/j.agwat.2010.12.015.

[14] Zhang Q, Manzoni S, Katul G, et al. The hysteretic evapotranspiration—Vapor pressure deficit relation. Journal of Geophysical Research: Biogeosciences, 2014, 119(2): 125–140. DOI: 10.1002/2013jg002484.

[15] Guo X, Shi J. Progress and trend of evapotranspiration research at home and abroad. Geological Review, 2019, 65(06): 1473–1486. DOI: 10.16509/j.georeview.2019.06.014. (in Chinese)

[16] Doorenbos J. Guidelines for predicting crop water requirements. Food and Agriculture organization. Rome, Irrig. Drainage pap., 1975, 24.

[17] Sabziparvar A A, Mirmasoudi S H, Tabari H, et al. ENSO teleconnection impacts on reference evapotranspiration variability in some warm climates of Iran. International Journal of Climatology, 2011, 31(11): 1710–1723. DOI: 10.1002/joc.2187.

[18] Pearson K. Notes on the history of correlation. Biometrika, 1920, 13(1): 25–45.

[19] Kendall M G. A new measure of rank correlation. Biometrika, 1938, 30(1/2): 81–93.

[20] Spearman C. The proof and measurement of association between two things, 1961.

[21] Draper N R, Smith H. Applied regression analysis. John Wiley & Sons, 1998.

[22] Cui K, Jing X. Research on prediction model of geotechnical parameters based on BP neural network. Neural Computing and Applications, 2019, 31(12): 8205–8215. DOI: 10.1007/s00521-018-3902-6.

[23] Luo Z, Yang Y, Yang Shi. Spatial-temporal Characteristics of Reference Crop Evapotranspiration and Climatic Influence Factors in Chongqing. Water Saving Irrigation, 2012(10): 5–9. (in Chinese)

[24] Zhang D, Liu X, Hong H. Assessing the effect of climate change on reference evapotranspiration in China. Stochastic Environmental Research and Risk Assessment, 2013, 27(8): 1871–1881. DOI: 10.1007/s00477-013-0723-0.

[25] Zhao J, Xia H, Yue Q, et al. Spatiotemporal variation in reference evapotranspiration and its contributing climatic factors in China under future scenarios. International Journal of Climatology, 2020, 40(8): 3813–3831. DOI: 10.1002/joc.6429.

[26] Di S, De-xin G, Feng-hui Y, et al. Time lag effect between poplar's sap flow velocity and microclimate factors in agroforestry system in west Liaoning Province. Yingyong Shengtai Xuebao, 2010, 21(11). DOI: 10.13287/j.1001-9332.2010.0426. (in Chinese)

[27] Oguntunde P G, Van de Giesen N C, Vlek P L G, et al. Water flux in a cashew orchard during a wet-to-dry transition period: analysis of sap flow and eddy correlation measurements Earth Interactions, 2004, 8(15): 1–17. DOI: 10.1175/1087-3562(2004)8.

[28] Bo X, Du T, Ding R, et al. Time lag characteristics of sap flow in seed-maize and their implications for modeling transpiration in an arid region of Northwest China. Journal of Arid Land, 2017, 9(4): 515–529. DOI: 10.1007/s40333-017-0024-4.

[29] Liu X, Kang S, Li F. Simulation of artificial neural network model for trunk sap flow of Pyrus pyrifolia and its comparison with multiple-linear regression. Agricultural Water Management, 2009, 96(6): 939–945. DOI: 10.1016/j.agwat.2009.01.003.

[30] Kisi O. The potential of different ANN techniques in evapotranspiration modeling. Hydrological Processes: An International Journal, 2008, 22(14): 2449–2460. DOI: 10.1002/hyp.6837.

[31] Granata F. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. Agricultural Water Management, 2019, 217: 303–315. DOI: 10.1016/j.agwat.2019.03.015.

[32] Fehrman B, Gess B, Jentzen A. Convergence rates for the stochastic gradient descent method for non-convex objective functions. Journal of Machine Learning Research, 2020, 21: 136.