Computers and Electronics in Agriculture 120 (2016) 10–16

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Determination of dominant weather parameters on reference evapotranspiration by path analysis theory

Xuguang Xing [a,b], Ye Liu [a,b], Wen'gang Zhao [a,b], Duan'gang Kang [a,b], Miao Yu [a,b], Xiaoyi Ma [a,*]

[a] College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Key Laboratory for Agricultural Soil and Water Engineering in Arid Area of Ministry of Education, Northwest A&F University, Yangling, Shaanxi Province 712100, China

ARTICLE INFO

Article history:
Received 23 June 2015
Received in revised form 29 September 2015
Accepted 2 November 2015
Available online 21 November 2015

Keywords:
Path analysis
Reference evapotranspiration
Penman–Monteith equation
Artificial neural network

ABSTRACT

Reference evapotranspiration (ET_0) is very important for the water cycle in farmland. Given this importance, spring ET_0 in South China was calculated by the Penman–Monteith (P–M) method based on daily meteorological data from 1969 to 2010. To reduce workload and apply in areas where climate data are not available, we researched simplification and universality of the double-factor back-propagation (BP) model for estimating ET_0. Quantitative analysis showed that in South China, based on path analysis theory, sunshine duration (N) had the strongest effect on ET_0. N was the dominant meteorological factor, with the largest correlation and path coefficients, 0.8357 and 0.5559, respectively. Minimum temperature (T_{min}), mean temperature (T_{mean}) and relative humidity (RH) were the next determining factors in different cities. Simplified BP models with only input variables N and T_{min}, and N and T_{mean} could effectively estimate ET_0, with small MAE (mean average error) both <0.2 mm, small RMSE (root mean square error) both around 0.4 mm, small MAPE (mean absolute percentage error) both <6%, and large R^2 0.9569 and 0.9479, respectively. Universality of the simplified double-factor model was further evaluated based on the selected reasonable determinants and accurate BP estimation model. Results showed that the former model for the cities of Nanning and Wuhan had great potential across South China with large R^2 of 0.9566 and 0.9609, meeting the demands of agricultural production. Further, there was a strong linear relationship between the ET_0 calculated by the P–M method and those estimated by the two universal models.

1. Introduction

Evapotranspiration, which includes evaporation from soil and vegetative surface and transpiration from plants, is considered as an important component of the hydrologic process and has an important role in field water management (Gocic et al., 2015; Petkovic et al., 2015; Shiri et al., 2014; Talebnejad and Sepaskhah, 2015). Reference evapotranspiration (ET_0) is an important basis for estimating crop evapotranspiration and quantifying crop irrigation water requirements, and has been widely used in various water resource fields, such as agricultural irrigation strategy, water balance in farmland, and land resource planning (Chang et al., 2010; Jacobsen et al., 2012; Stagnari et al., 2014).

There are many methods for ET_0 estimation. The Penman–Monteith (P–M) method is recommended as the standard by the United Nations Food and Agriculture Organization (UNFAO) and has gained worldwide acceptance and received much research interests (DeJonge et al., 2015; Gavilan et al., 2007; Lopez-Urrea et al., 2006; Petkovic et al., 2015; Shan et al., 2015). The P–M equation has been widely used in different climatic and geographic regions to assess monthly, daily or hourly ET_0 (Ladlani et al., 2012; Perera et al., 2014), and its result is more realistic and accurate than that from other methods, such as Hargreaves and Priestley–Taylor (Berti et al., 2014; Szilagyi, 2014). Although the FAO P–M method is accepted worldwide and has broad application prospects, its drawbacks are not negligible. For example, not all climate data required by the P–M equation are available in some areas (Pereira et al., 2015). Therefore, the following issues need urgent study: (1) selection of as few dominant meteorological variables as possible affecting ET_0, and (2) universal application of an established model in more regions.

In selecting key determinants from all factors, principal component analysis and factor analysis are considered effective (Shrestha et al., 2012). However, when using these analyses, there are two main requirements. The first is mutual independence among principal components, and the second is non-significant linear relations among factors. For the present study, however, there is obvious interaction among all meteorological factors in determining

* Corresponding author at: No. 23 Weihui Road, Yangling, Shaanxi Province 712100, China. Tel.: +86 29 87082860.
E-mail address: xiaoyima@vip.sina.com (X. Ma).

http://dx.doi.org/10.1016/j.compag.2015.11.001
0168-1699/

ET_0, which conflicts with the two requirements. Simple correlation analysis may not take full account of the interactions among variables, likely leading to biased results. In multiple regression analysis, although it may eliminate interaction between variables, it is difficult to compare the effects of each independent variable on the dependent variable, because the partial regression coefficient has units. Consequently, path analysis is preferred, because interaction among independent variables (i.e., climate factors) and the combined influence of all independent variables (also climate factors) on the dependent variable (i.e., ET_0) are all considered by this theory.

When establishing ET_0 estimation models via selected dominant climate factors, another alternative is the application of intelligent computational models, such as the artificial neural network (ANN) (Huo et al., 2012), and the ANN method has been effectively applied with high accuracy in ET_0 estimation (Gocic et al., 2015; Shiri et al., 2014; Trajkovic, 2010; Traore et al., 2010). However, in many researches on ET_0 estimation via ANN models, some traditionally crucial climate factors (e.g., maximum and minimum temperature, relative humidity, radiation and wind speed) and other factors required by the P–M equation have been considered as input variables (Falamarzi et al., 2014; Huo et al., 2012; Landeras et al., 2008; Traore et al., 2010). This often results in a waste of time and energy. Therefore, a three-layer BP network with one input layer, one hidden/intermediate layer and one output layer is selected for the present study. More importantly, not many but only one or two dominant climate factors determined by path analysis are used as input variables to ANN models.

In this study on spring ET_0 estimation, the original climate data required were obtained from agro-meteorological observation stations. First, path analysis was used for determining as few dominant meteorological factors as possible, and these factors were viewed as independent (input) variables of the ANN model. Second, for model establishment, a three-layer BP network was selected because of its wide application and high accuracy. For verifying the established ANN model, its accuracy was determined by the difference between ET_0 values computed by the established model and the FAO P–M equation. The primary objectives were as follows: (1) Propose a more reasonable method for selecting as few determinants as possible among many climate factors for ET_0 estimation; (2) Simplify ANN models using only those few determinants in computing ET_0, evaluate the universal performance of the optimal ANN model, and finally generalize the model to all China and more regions with a similar climate; and (3) Put the established model into practice to predict ET_0 in local areas where lacks meteorological data.

2. Materials and methods

2.1. Study area and weather data source

China is divided into north and south by a Qinling Mountain–Huaihe River line, each with significantly different climate features (Fig. 1). Generally, cultivated area in the south is smaller, and demands intensive cultivation and greater crop productivity. Consequently, the south (around 18–35°N and 98–122°E) was selected as the study area to enhance the universality of the simplified model for ET_0 calculation, thereby promoting crop yields. Hong Kong and Macao special administrative regions and Taiwan were omitted, leaving 14 provincial capital cities in the south for analysis (Fig. 1).

All original spring (March–May) meteorological data from 1969 to 2010 were obtained from the China Meteorological Data Sharing Service System website. Data from 1969 to 2010 were used to determine dominant meteorological factors and establish BP estimation models based on those factors. Data from 1969 to 2003 were used to calibrate the BP models, including training models and determining inner parameters. Data from 2004 to 2010 were used to validate the established BP models. The original data were of high accuracy and reliability, owing to continuous observation by the automatic agro-meteorological observation stations in each Chinese province.

2.2. Reference evapotranspiration

The FAO P–M equation is written as Eq. (1), which is considered the standard method for calculation of crop ET_0. The calculated ET_0 data were thus viewed as standard values, which were used to train the ANNs and verify fitted values from the simplified model.

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

where ET_0 is reference evapotranspiration (mm d^{-1}), R_n is net radiation (MJ m^{-2} d^{-1}), G is soil heat flux density (MJ m^{-2} d^{-1}), γ is the psychrometric constant (kPa °C^{-1}), Δ is the slope of the saturation vapor pressure curve (kPa °C^{-1}), u_2 is wind speed at 2 m height (m s^{-1}), e_s is saturation vapor pressure (kPa), e_a is actual vapor pressure (kPa), and T is mean air temperature (°C). And the meteorological factors in Eq. (1) could be obtained through maximum temperature (T_{max}, °C), minimum temperature (T_{min}, °C), mean temperature (T_{mean}, °C), mean wind speed at 2 m height (u_2, m s^{-1}), mean relative humidity (RH, %) and sunshine duration (N, h) (Gocic and Trajkovic, 2010). Meanwhile, in consideration of easy access to meteorological parameters, the main goal of this study is to determine how such six parameters affect ET_0.

2.3. Path analysis theory

Path analysis is a type of multivariate statistical analysis for studying relationships among variables, and it can reveal the strength of effect of independent variables on a dependent variable (Cai et al., 2008). Path analysis can determine direct and indirect effects of independent variables on the dependent variable, multi-collinear independent variables resulting from their own strong correlations, and optimal regression equations without unnecessary independent variables. The path coefficient is a type of standard partial regression coefficient without units that expresses causalities among related variables, and is also a directional correlation coefficient between independent and dependent variables (Cai et al., 2008).

For a correlative system with one dependent variable y and independent variables x_i (i = 1, 2, …, n), the linear equation is:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n \quad (2)$$

A regular matrix equation based on Eq. (2) can be established as:

$$\begin{bmatrix} 1 & r_{x_nx_1} & \cdots & r_{x_1x_n} \\ r_{x_2x_1} & 1 & \cdots & r_{x_1x_n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{x_nx_1} & r_{x_nx_2} & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} P_{yx_1} \\ P_{yx_2} \\ \vdots \\ P_{yx_n} \end{bmatrix} = \begin{bmatrix} r_{x_1y} \\ r_{x_2y} \\ \vdots \\ r_{x_ny} \end{bmatrix} \quad (3)$$

where $r_{x_ix_j}$ is the simple correlation coefficient of x_i and x_j, P_{yx_i} is the path coefficient.

$$P_{yx_i} = b_i\frac{\sigma_{x_i}}{\sigma_y}, \quad (i = 1, 2, \cdots, n) \quad (4)$$

where b_i is the partial regression coefficient of y to x_i, σ_{x_i} and σ_y are the standard deviations of x_i and y. P_{yx_i} is the direct path coefficient