Computers and Electronics in Agriculture 140 (2017) 203–213

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Pan evaporation modeling using four different heuristic approaches

Lunche Wang [a,*], Zigeng Niu [a], Ozgur Kisi [b], Chang'an Li [a], Deqing Yu [c]

[a] Laboratory of Critical Zone Evolution, School of Earth Sciences, China University of Geosciences, Wuhan 430074, China
[b] School of Natural Sciences and Engineering, Ilia State University, Tbilisi, Georgia
[c] Remote Sensing Centre of Hunan Province, Changsha 410007, China

ARTICLE INFO

Article history:
Received 11 January 2017
Received in revised form 27 May 2017
Accepted 29 May 2017
Available online 15 June 2017

Keywords:
Fuzzy genetic
Least square support vector regression
M5 model tree
Multivariate adaptive regression spline
Dongting Lake
Evaporation estimation

ABSTRACT

Evaporation plays important roles in regional water resources management, climate change and agricultural production. This study investigates the abilities of fuzzy genetic (FG), least square support vector regression (LSSVR), multivariate adaptive regression spline (MARS), M5 model tree (M5Tree) and multiple linear regression (MLR) in estimating daily pan evaporation (*Ep*). Daily climatic data, air temperature (*Ta*), surface temperature (*Ts*), wind speed (*Ws*), relative humidity (*RH*) and sunshine hours (*Hs*) at eight stations in the Dongting Lake Basin, China are used for model development and validation. The first part of this study focuses on testing the model accuracies at each station using local input and output data. The results show that LSSVR and FG models with more input variables perform better than the MARS, M5Tree and MLR models in predicting daily *Ep* at most stations with respect to mean absolute errors (MAE), root mean square errors (RMSE) and determination coefficient (R^2). In the second part of this study, the models are tested using cross-validation method in two different applications. The daily *Ep* of Yueyang station is estimated using the input and output data of Jingzhou and Changsha, respectively. Comparisons of the models indicate that the FG, LSSVR and MARS models outperform the M5Tree model, *Ts*, *Hs* and *Ta* are major influencing factors and adding *Ws* or *RH* into model inputs significantly improve the model performances. The overall results indicate that above models can be successfully used for estimating daily *Ep* using local input and output data while the FG and LSSVR generally perform better than the other models without local input and outputs.

1. Introduction

Pan evaporation (*Ep*) is an important indicator for atmospheric evaporative demand, which is extensively used in designing regional water resources and irrigation systems (Roderick and Farquhar, 2002; Azorin-Molina et al., 2015). It is an important hydrological component in understanding of the balance of water and energy cycles in the context of global climate change (Wang et al., 2007; Yang and Yang, 2012; Li et al., 2013b). Many studies have analyzed the trends and causes of *Ep* in some regions of the world (Liu et al., 2004, 2011; Zhang et al., 2015), however, it is widely recognized that evaporation is still one of the less understood components of hydrologic cycle, especially in areas where water resources are rare (Goyal et al., 2014; Valipour and Eslamian, 2014; Majidi et al., 2015). Accurate observation, estimation and prediction of *Ep* are of great importance to integrated water resources management, irrigation control and agricultural production at both farm and regional scales (Rahimikhoob et al., 2013; Shiri et al., 2014).

There are generally two direct and indirect approaches for observing and calculating evaporation. One of the most widely used instruments for measuring evaporation is *Ep*, however, this direct method is affected by instrumental limits and practical issues such as measurement errors and maintenance, which may reduce the accuracies of evaporation measurements (Piri et al., 2009; Shirsath and Singh, 2010). Thus, lots of methods have been proposed to predict evaporation using observed climatic parameters, for example, several researchers fit a linear relationship between *Ep* and meteorological data (air temperature (*Ta*), air pressure (*Pa*), solar radiation (*Rg*), wind speed (*Ws*), relative humidity (*RH*) and sunshine hours (*Hs*), etc.) (Majidi et al., 2015). However, the empirical methods should be recalibrated when applied to other sites. In addition, the evaporation process is incidental, nonlinear, complex and unsteady (Kim et al., 2012; Kisi, 2015), it is difficult to derive an accurate relationship between *Ep* and climatic factors to represent all the physical processes involved (Kisi et al.,

* Corresponding author at: Laboratory of Critical Zone Evolution, School of Earth Sciences, China University of Geosciences, Lumo Road 388, Hongshan, Wuhan 430074, China.
E-mail addresses: wang@cug.edu.cn, lunchewang@whu.edu.cn (L. Wang).

http://dx.doi.org/10.1016/j.compag.2017.05.036

2012; Kim et al., 2015). Therefore, many researchers have emphasized the necessity for precise modeling of *Ep* in hydrologic modeling studies using nonlinear data driven methods (Shirsath and Singh, 2010; Majidi et al., 2015).

In recent decades, the soft computing techniques (e.g., artificial neural networks, fuzzy and neuro-fuzzy systems) have been successfully applied for modeling ecohydrological process around the world (Lin et al., 2013; Kisi et al., 2013; Kisi and Shiri, 2014; Kisi and Zounemat-Kermani, 2014; Bewoor et al., 2016), for example, Abghari et al. (2012) predicted the daily *Ep* using wavelet neural networks. Several studies have been performed in order to compare the model performances in limited data conditions, however, only a few studies have been conducted to find the most appropriate methods to estimate evaporation in the conditions that long term data are not available. Meanwhile, there is also no much information about the appropriate models to apply when lacking some important parameters (Goyal et al., 2014), especially the solar radiation and heat fluxes data in areas of lakes and reservoirs of China (Wang et al., 2016, 2017). The precise estimation of *Ep* using advanced techniques is still much needed for understanding the water and energy cycles.

Dongting Lake is the second largest freshwater lake in China, the water cycle process varied very significantly during last decades. The impoundment of the Three Gorges Dam (TGD) changed the hydrological regime downstream and the patterns of the lake wetlands (Yuan et al., 2015). Many attentions have been paid to the hydrological process in this region (Wang et al., 2007; Huang et al., 2014; Zhang et al., 2014), for example, Yuan et al. (2016) investigated the contributions of climate variability and human activity to stream flow alteration in this region. However, there is very few knowledge about the evaporation characteristics in Dongting Lake Basin due to the lack of observations, which promotes the *Ep* estimation using available climatic data in this region.

Therefore, there are critical requirements for investigating and evaluating the potential of different soft computing techniques (FG, LSSVR, M5Tree and MARS) in predicting *Ep* in Dongting Lake Basin using limited climatic variables. The cross validation method is employed and the model performances will be compared and analyzed through: (i) estimating *Ep* of each station using different local input combinations; (ii) estimating *Ep* without local input or output data. The most appropriate model for *Ep* estimation will be chosen, which constituted the first study of evaporation process in Dongting Lake Basin in the hydrological context.

2. Methods and materials

2.1. Modeling strategies

2.1.1. Fuzzy genetic approach

Fuzzy logic is a developed method for computing based on the "degrees of truth" rather than the usual "true or false" (1 or 0) logic. Many specifications of fuzzy logic methods have to be determined while fuzzy systems are studied. Some of them include the number of membership functions of input/output variables, types and parameters of membership functions, types of inference engines, types of fuzzy operators, difuzzification methods. There are many different types of membership functions (triangular, Gaussian, etc.), but Gaussian functions are generally preferred (Russel and Campbell, 1996) and widely used in various studies.

The fuzzy inference system (FIS) including sets of input and output data is presented in Fig. 1. The FIS consists of three main components: (1) a rule base comprising fuzzy IF-THEN rules; (2) a database composed of membership functions used in fuzzy rules; (3) an inference mechanism that combines these rules to relate a set of outputs to a set of inputs to derive a reasonable output. In the fuzzification, the input or output data are partitioned into subsets determined by linguistic terms (e.g., small, medium, big) and membership degrees. The main concepts of the fuzzy approaches can be explained as below:

The input and output parameters are first divided into a number of subsets with Gaussian membership functions. There are c^n fuzzy rules where c and n indicate the numbers of subsets and input parameters, respectively. If the number of subset increases, the higher accuracy may be obtained. In this case, however, the rule base gets bigger (Sen, 1998). Assuming that we have two inputs with two fuzzy subsets or membership functions labeled as "small" and "big" and one output with three fuzzy subsets labeled as "small", "medium" and "big", then there will be three rules:

R1: IF x is small THEN y_1
R2: IF x is medium THEN y_2
R3: IF x is big THEN y_3

Thus, the weighted average of the outputs from three rules results a single weighted output, y, as:

$$y = \frac{\sum_{n=1}^{3} w_n \cdot y_n}{\sum_{n=1}^{3} w_n} \tag{1}$$

where w_n indicates membership degree, for x is figured to be appointed to the corresponding output y_n for each triggered rule.

In this manner, the output values, y, can be obtained from Eq. (1) for any combinations of input fuzzy subsets after setting up the rule base. More detailed information about fuzzy concepts and operational calculations are also given in previous studies (Sen, 1998; Ross, 1995).

The optimal parameters (e.g. membership functions) of a fuzzy model can be determined using genetic algorithms (GA). Genetic algorithms work as a machine learning algorithm in genetic fuzzy systems. The principle idea of GA is to simulate the process of the natural evolution and natural selection mechanisms of chromosomes, involving reproduction, crossover, and mutation. Selection is utilized to choose chromosomes (refer to string structures) from the present population for reproduction according to their fitness values. Crossover is utilized to create new parameter sets by changing sets of strings. Mutation is utilized to change in one of the strings locations. These processes are repeated and stop only

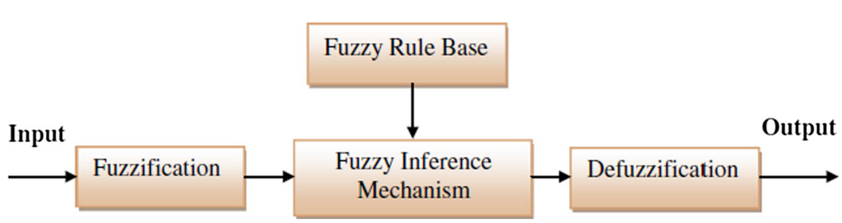

Fig. 1. Schematic diagram of a fuzzy inference system.