Journal of Hydrology 544 (2017) 407–427

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Pan evaporation modeling using six different heuristic computing methods in different climates of China

Lunche Wang [a,*], Ozgur Kisi [b], Mohammad Zounemat-Kermani [c], Hui Li [a]

[a] Laboratory of Critical Zone Evolution, School of Earth Sciences, China University of Geosciences, Wuhan 430074, China
[b] Center for Interdisciplinary Research, International Black Sea University, Tbilisi, Georgia
[c] Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

ARTICLE INFO

Article history:
Received 3 September 2016
Received in revised form 17 October 2016
Accepted 27 November 2016
Available online 30 November 2016
This manuscript was handled by Geoff Syme, Editor-in-Chief, with the assistance of Jesús Mateo-Lázaro, Associate Editor

Keywords:
Pan evaporation
Soft computing techniques
Regression methods
Model comparison

ABSTRACT

Pan evaporation (*Ep*) plays important roles in agricultural water resources management. One of the basic challenges is modeling *Ep* using limited climatic parameters because there are a number of factors affecting the evaporation rate. This study investigated the abilities of six different soft computing methods, multi-layer perceptron (MLP), generalized regression neural network (GRNN), fuzzy genetic (FG), least square support vector machine (LSSVM), multivariate adaptive regression spline (MARS), adaptive neuro-fuzzy inference systems with grid partition (ANFIS-GP), and two regression methods, multiple linear regression (MLR) and Stephens and Stewart model (SS) in predicting monthly *Ep*. Long-term climatic data at various sites crossing a wide range of climates during 1961–2000 are used for model development and validation. The results showed that the models have different accuracies in different climates and the MLP model performed superior to the other models in predicting monthly *Ep* at most stations using local input combinations (for example, the MAE (mean absolute errors), RMSE (root mean square errors), and determination coefficient (R^2) are 0.314 mm/day, 0.405 mm/day and 0.988, respectively for HEB station), while GRNN model performed better in Tibetan Plateau (MAE, RMSE and R^2 are 0.459 mm/day, 0.592 mm/day and 0.932, respectively). The accuracies of above models ranked as: MLP, GRNN, LSSVM, FG, ANFIS-GP, MARS and MLR. The overall results indicated that the soft computing techniques generally performed better than the regression methods, but MLR and SS models can be more preferred at some climatic zones instead of complex nonlinear models, for example, the BJ (Beijing), CQ (Chongqing) and HK (Haikou) stations. Therefore, it can be concluded that *Ep* could be successfully predicted using above models in hydrological modeling studies.

1. Introduction

Evaporation is the process of conversion of liquid water to water vapor, which depends on the differences in vapor pressure between the surface and surrounding atmosphere (Kim et al., 2015). Pan evaporation (*Ep*) has been widely used as an index of lake and reservoir evaporation, potential or reference crop evapotranspiration and irrigation (Shiri et al., 2011), which plays important roles in informing water resources distribution and irrigation system design. There are many climatic factors influencing the rates of *Ep*, including solar radiation (*Rg*), air temperature (*Ta*), relative humidity (*RH*) and wind speed (*Ws*). The quantitative effects of different climatic parameters on *Ep* variations in different regions are still less well understood. Therefore, proper estimation and prediction of *Ep* is of great importance for integrating water resources management and modeling studies.

The direct measurements of *Ep* are spatially and temporally limited due to instrumental and practical issues (Shirsath and Singh, 2010; Shiri et al., 2014; Martí et al., 2015). Many researchers have tried to estimate the evaporation through indirect methods using climatic variables, for example, many empirical or semi-empirical equations have been developed to linking *Ep* to various meteorological drivers (Stephens and Stewart, 1963; Piri et al., 2009), but the applications of these techniques are often limited by data availability and completeness (Sharda et al., 2008; Kisi and Shiri, 2014; Majidi et al., 2015). Recently, advanced soft computing techniques (such as artificial neural network, ANN) have been successfully applied for modeling *Ep* due to its ability to learn complex and non-linear relationships that are difficult to model with conventional techniques (for example, Priestley-Taylor, and Stephens and Stewart models) (Kim and Kim, 2008; Goyal et al.,

* Corresponding author at: Laboratory of Critical Zone Evolution, School of Earth Sciences, China University of Geosciences, Lumo Road 388, Hongshan District, Wuhan 430074, China.

E-mail address: wang@cug.edu.cn (L. Wang).

http://dx.doi.org/10.1016/j.jhydrol.2016.11.059
0022-1694/

2014; Shiri et al., 2015). For example, Kisi (2009) investigated the abilities of three different ANN techniques and it was found that the MLP and radial basis neural network (RBNN) computing techniques could be employed successfully to model the evaporation process using the available climatic data. Piri et al. (2009) improved the ANN model by incorporating an autoregressive external input (ARX) component and evaluated the models for *Ep* estimation at a hot and dry site of Southeast Iran. The results showed that NNARX was better than the ANN and Marciano method, the models with inputs of wind and vapor pressures performed much better than the ones with temperature and dew point. Chang et al. (2010) proposed a self-organizing map neural network (SOMN) to assess the variability of daily evaporation based on meteorological variables, the results demonstrated that the topological structures of SOMN could give a meaningful map to present the clusters of meteorological variables and the networks could well estimate the daily evaporation (Kim et al., 2015). Kim et al. (2012) applied multilayer perceptron-neural networks (MLP), generalized regression neural networks (GRNN) and support vector machine-neural networks (SVM) to estimate *Ep* in temperate and arid climatic zones and the results indicated that these ANN models performed better than the empirical Linacre model and MLR model. Goyal et al. (2014) investigated the abilities of ANN, Least Squares Support Vector Machine (LSSVM), Fuzzy Logic (FG), Adaptive Neuro-Fuzzy Inference System (ANFIS) techniques, Hargreaves and Samani method (HGS), as well as the Stephens-Stewart (SS) method to improve the accuracy of daily *Ep* estimation in sub-tropical climates of India. Results showed that the above soft computing models outperformed the HGS and SS methods, and the LSSVM and FG models produced the highest accuracies. Kisi (2015) investigated the accuracy of LSSVM, multivariate adaptive regression splines (MARS) and M5 Model Tree (M5Tree) in modeling *Ep* at Mersin and Antalya stations in Mediterranean region of Turkey, which indicated that the LSSVM model could be successfully used for estimating *Ep* using local input and output data while the MARS model performed better than the LSSVM model using data from other stations. Several studies have also been performed in order to compare and assess *Ep* models with limited data around the world (Majidi et al., 2015). In contrast, only a few studies have been conducted to find the most appropriate methods to estimate *Ep*, and most of these studies focused on comparing only two or three models. Therefore, there is no clear consensus on which methods are better to employ when lacking important long term measured data such as radiation and heat fluxes. Meanwhile, the *Ep* models are only tested at few stations and they have not been tested in different climates. For example, Keskin et al. (2004) only compared the FG model with empirical Penman method at Lake Eğirdir in Turkey; Sanikhani et al. (2012) compared two different ANFIS models including grid partitioning (GP) and subtractive clustering (SC), in modeling *Ep* at San Francisco and San Diego in California, however, there are almost no studies using large number of stations (>3) for obtaining more generalized conclusions. In addition, there are not any studies in literature that compare different methods in estimating *Ep* at different climates (for example, the arid continental climate, desert climate, semi humid monsoon climate, plateau climate and the tropical maritime monsoon climate), which impede for the present investigation for revealing a more robust and applicable *Ep* estimation model.

The aim of this study is to investigate capability and usability of six different soft computing methods, ANFIS-GP, FG, GRNN, LSSVM, MARS and MLP, and two regression methods, MLR and SS, in *Ep* modeling with different combinations of climate inputs. Data from eight stations in different climatic zones are used for training and testing above models. The model performances will be compared and discussed through: (i) estimating *Ep* of each station using different local input combinations; (ii) estimating *Ep* of eight stations using eight different models. To the knowledge of the authors, no similar studies have been reported using above mentioned methods for modeling *Ep*, this will be the first study to compare the accuracy of multiple soft computing models for *Ep* estimation in different climates.

2. Methods and materials

2.1. Modeling strategies

2.1.1. Multi-layer perceptron neural network

The MLP is a well-known and efficient neural network widely used for a variety of problems such as classification, time series modeling and regression. MLPs are organized as hierarchical networks with several layers including an input layer, hidden layer (s) and an output layer (Wang et al., 2016). There are one or more hidden layers between the input and output layers which are connected by neurons (including synaptic weights, biases and activation or transfer functions). Each neuron receives input value(s) from the input vector (or the antecedent hidden layer's output) and then calculates a weighted sum of input values passing through the transfer function, which generates the output of the neuron (Fig. 1a). MLPs are feed-forward networks, using the error back-propagation (BP) algorithm for network training. In the BP algorithm, an iterative process changes the weights and biases of the network to optimize the solution by reducing the overall error between the output and target (generally the observed parameters) values. More details about the MLP model can be found in Zounemat-Kermani (2012). In current study, Levenberg-Marquardt algorithm was used for MLP training. This algorithm is similar to the Gauss-Newton method and converges faster and more accurately towards an error minimum (Hagan and Menhaj,

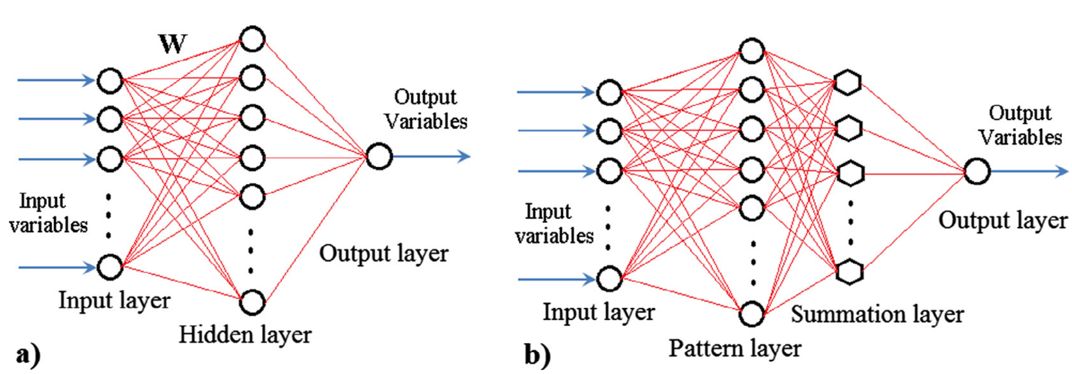

Fig. 1. Schematic architecture of: (a) MLP neural network; (b) GRNN.