Journal of Hydrology 528 (2015) 312–320

Contents lists available at ScienceDirect

Journal of Hydrology

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

Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree

Ozgur Kisi

Canik Basari University, Architecture and Engineering Faculty, Civil Eng. Dept., 55080 Samsun, Turkey

ARTICLE INFO

Article history:
Received 1 June 2015
Received in revised form 20 June 2015
Accepted 22 June 2015
Available online 27 June 2015
This manuscript was handled by Geoff Syme, Editor-in-Chief

Keywords:
Pan evaporation
Estimation
Cross-station application
Least square support vector machine
Multivariate adaptive regression splines
M5 model tree

SUMMARY

Pan evaporation (Ep) modeling is an important issue in reservoir management, regional water resources planning and evaluation of drinking-water supplies. The main purpose of this study is to investigate the accuracy of least square support vector machine (LSSVM), multivariate adaptive regression splines (MARS) and M5 Model Tree (M5Tree) in modeling Ep. The first part of the study focused on testing the ability of the LSSVM, MARS and M5Tree models in estimating the Ep data of Mersin and Antalya stations located in Mediterranean Region of Turkey by using cross-validation method. The LSSVM models outperformed the MARS and M5Tree models in estimating Ep of Mersin and Antalya stations with local input and output data. The average root mean square error (RMSE) of the M5Tree and MARS models was decreased by 24–32.1% and 10.8–18.9% using LSSVM models for the Mersin and Antalya stations, respectively. The ability of three different methods was examined in estimation of Ep using input air temperature, solar radiation, relative humidity and wind speed data from nearby station in the second part of the study (cross-station application without local input data). The results showed that the MARS models provided better accuracy than the LSSVM and M5Tree models with respect to RMSE, mean absolute error (MAE) and determination coefficient (R^2) criteria. The average RMSE accuracy of the LSSVM and M5Tree was increased by 3.7% and 16.5% using MARS. In the case of without local input data, the average RMSE accuracy of the LSSVM and M5Tree was respectively increased by 11.4% and 18.4% using MARS. In the third part of the study, the ability of the applied models was examined in Ep estimation using input and output data of nearby station. The results reported that the MARS models performed better than the other models with respect to RMSE, MAE and R^2 criteria. The average RMSE of the LSSVM and M5Tree was respectively decreased by 54% and 3.4% using MARS. The overall results indicated that the LSSVM could be successfully used in estimating Ep by using local input and output data while the MARS model performed better than the LSSVM in the case of without local input and outputs.

1. Introduction

Accurately estimation of evaporation is very important for regional water resources planning and reservoir controlling; allocation of water supplies for diverse sectors, for instance domestic, agriculture, industry and energy; and drought management (Abghari et al., 2012). Evaporation losses should be well-thought-out in the plan of various water resources and irrigation systems. In the areas receiving little rainfall, evaporation losses can characterize a significant amount of the water budget for a lake or reservoir, and contribute significantly to the dropping of water surface level (McCuen, 1998). For that reason, precise estimation of evaporation loss from the water body has primary importance for monitoring and allocating water resources, at farm scales as well as at regional scales (Piri et al., 2009b).

The artificial intelligence methods has been successfully applied for modeling pan evaporation (Ep) in the last decades (Bruton et al., 2000; Dogan et al., 2007; Goyal et al., 2014; Karimi-Googhari, 2010; Kim and Kim, 2008; Kim et al., 2014; Kisi, 2005, 2006, 2009a,b, 2013; Lin et al., 2013; Malik and Kumar, 2015; Moghaddamnia et al., 2010; Nourani and Sayyah Fard, 2012; Piri et al., 2009a; Samui and Dixon, 2012; Sanikhani et al., 2012; Sudheer et al., 2002; Terzi and Keskin, 2008; Yang, 2013). Bruton et al. (2000) developed artificial neural network (ANN) models for estimating daily Ep. They used weather data from Pome, Plains, and Watkinsville, Georgia, consisting of 2044 daily records from 1992 to 1996 to develop the Ep models. Sudheer et al. (2002) investigated the prediction of Ep from minimum climatic data using ANN technique. Their study showed that Ep values

E-mail address: okisi@basari.edu.tr

http://dx.doi.org/10.1016/j.jhydrol.2015.06.052
0022-1694/

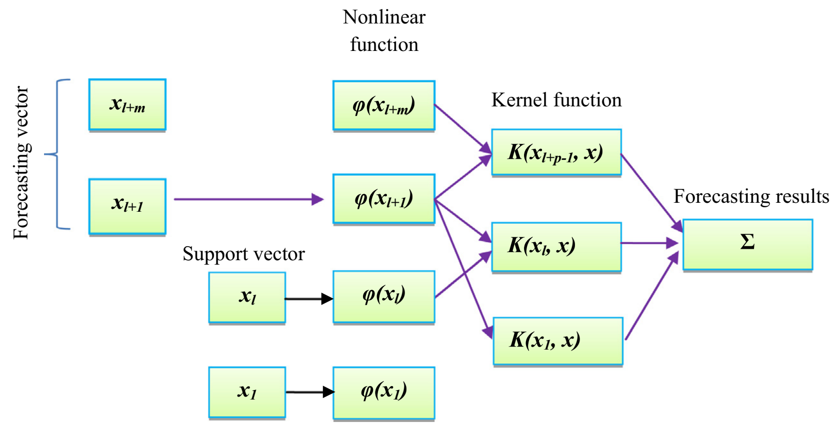

Fig. 1. LSSVM model for Ep modeling.

could be reasonably estimated using temperature data only through the ANN technique. Kisi (2006) compared the accuracy of neuro-fuzzy (NF) and ANN techniques for estimating daily Ep using various combinations of daily air temperature, solar radiation wind speed, pressure and humidity. He found that the NF computing technique performed better than the ANN in modeling Ep process from the available climatic data. Dogan et al. (2007) compared the accuracy of feed forward neural network (FFNN) and radial basis neural network (RBNN) models in estimating daily Ep of Lake Sapanca and they found that the FFNN model yielded better results than the RBNN. Piri et al. (2009a) modeled Ep using ANN model in a hot and dry region. They found that ANN works very well at the studied region and, further, an integrated ANN and autoregressive with exogenous inputs can have an improved performance over the traditional ANN. Kisi (2009b) applied multi-layer perceptrons (MLP), RBNN and generalized regression neural network for estimating daily Ep. Based on the comparisons, it was found that the MLP and RBNN methods could be successfully employed to model Ep process using the available climatic data. Kim et al. (2014) evaluated the accuracy of MLP and cascade correlation neural networks (CCNN) in estimating daily Ep for inland and coastal stations in Republic of Korea and they indicated that the CCNN model was better than the MLP during the test period for homogeneous and nonhomogeneous weather stations. Lin et al. (2013) compared support vector machine (SVM) and MLP models in Ep estimation and found that the SVM-based model was more appropriate than the MLP model because of its higher accuracy, robustness and efficiency. Malik and Kumar (2015) compared ANN, co-active adaptive neuro-fuzzy inference system (CANFIS) and multi-linear regression (MLR) models in estimating daily Ep at Pantnagar, located at the foothills of Himalayas in the Uttarakhand State of India. The results indicated that the performance of ANN model was generally superior to the CANFIS and MLR models. To the knowledge of the author, there is not any published work in the literature related to application of least square support vector machine (LSSVM), multivariate adaptive regression splines (MARS) and M5 Model Tree (M5Tree) models for estimating Ep.

The study aims to investigate the ability of LSSVM, MARS and M5Tree methods in (i) locally modeling of monthly Ep of Mersin and Antalya stations, (ii) estimating Ep of Antalya Station using input data of Mersin Station and (iii) estimating Ep of Antalya Station using the data of Mersin Station without local input and output data.

2. Methods

2.1. Least square support vector machine

The LSSVM evolved from the SVM is a powerful method for solving non-linear problems, classification and function estimation (Kumar and Kar, 2009). The procedure of the LSSVM, first proposed by Suykens and Vandewalle (1999) is illustrated in Fig. 1. By considering inputs x_i (climatic data) and output y_i (Ep values) time series, the nonlinear function LSSVM function can be expressed as

$$f(X) = w^T \varphi(X) + b \tag{1}$$

where w, φ and b are the m-dimensional weight vector, mapping function and bias term, respectively (Shu-gang et al., 2008).

Using the function estimation error, the regression problem can be expressed regarding structural minimization principle as

$$min\ J(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{m} e_i^2 \tag{2}$$

which subject to following constraints

$$y_i = w^T \varphi(x_i) + b + e_i \quad (i = 1, 2, \ldots, m) \tag{3}$$

where γ refers the regularization constant and e_i is the training error for x_i.

To find the solutions of w and e, the Lagrange multiplier optimal programming method is employed to solve Eq. (2). The objective function can be determined by altering the constraint problem into an unconstraint problem. The Lagrange function L can be expressed as

$$L(w, b, e, \propto) = J(w, e) - \sum_{i=1}^{m} \propto_i \{ w^T \varphi(x_i) + b + e_i - y_i \} \tag{4}$$

where $\propto_i$ is the Lagrange multipliers.

Taking into account the Karush–Kuhn–Tucker (Flecher, 1987), the optimal conditions can be obtained by taking the partial derivatives of Eq. (4) with respect to w, b, e and $\propto$, respectively as

$$\begin{cases} w = \sum_{i=1}^{m} \propto_i \varphi(x_i) \\ \sum_{i=1}^{m} \propto_i = 0 \\ \propto_i = \gamma e_i \\ w^T \varphi(x_i) + b + e_i - y_i = 0 \end{cases} \tag{5}$$