Expert Systems with Applications 41 (2014) 5267–5276

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS

Manish Kumar Goyal [a], Birendra Bharti [b], John Quilty [c], Jan Adamowski [c,*], Ashish Pandey [b]

[a] Dept. of Civil Engineering, Indian Institute of Technology, Guwahati 781039, India
[b] Dept. of Water Resources Development Management, Indian Institute of Technology, Roorkee 247667, India
[c] Dept. of Bioresource Engineering, McGill University, 21 111 Lakeshore Road, Ste. Anne de Bellevue, QC H9X 3V9, Canada

ARTICLE INFO

Keywords:
Evaporation
Modeling
Gamma Test
ANN
LS-SVR
Fuzzy Logic
ANFIS

ABSTRACT

This paper investigates the abilities of Artificial Neural Networks (ANN), Least Squares – Support Vector Regression (LS-SVR), Fuzzy Logic, and Adaptive Neuro-Fuzzy Inference System (ANFIS) techniques to improve the accuracy of daily pan evaporation estimation in sub-tropical climates. Meteorological data from the Karso watershed in India (consisting of 3801 daily records from the year 2000 to 2010) were used to develop and test the models for daily pan evaporation estimation. The measured meteorological variables include daily observations of rainfall, minimum and maximum air temperatures, minimum and maximum humidity, and sunshine hours. Prior to model development, the Gamma Test (GT) was used to derive estimates of the noise variance for each input–output set in order to identify the most useful predictors for use in the machine learning approaches used in this study. The ANN models consisted of feed forward backpropagation (FFBP) models with Bayesian Regularization (BR), along with the Levenberg–Marquardt (LM) algorithm. A comparison was made between the estimates provided by the ANN, LS-SVR, Fuzzy Logic, and ANFIS models. The empirical Hargreaves and Samani method (HGS), as well as the Stephens–Stewart (SS) method, were also considered for comparison with the newer machine learning methods. The Root Mean Square Error (RMSE) and Correlation Coefficient (CORR) were the statistical performance indices that were used to evaluate the accuracy of the various models. Based on the comparison, it was found that the Fuzzy Logic and LS-SVR approaches can be employed successfully in modeling the daily evaporation process from the available climatic data. In addition, results showed that the machine learning models outperform the traditional HGS and SS empirical methods.

1. Introduction

Evaporation, as a major component of the hydrologic cycle, is important in water resources development and management since it affects the yield of river basins, the capacity of reservoirs, the consumptive use of water by crops and the yield of underground supplies. In many parts of the world, where availability of water resources is scarce, the estimation of this evaporation loss is very important in the planning and management of irrigation practices, and these losses should be considered in the design of various water resources and irrigation systems (Tabari, Marofi, & Sabziparvar, 2009). In areas with little rainfall, evaporation losses can represent a significant part of the water budget for a lake or reservoir, and may contribute significantly to the lowering of the water surface elevation (McCuen, 1998). Despite this significance, evaporation is one of the least understood components of the hydrologic cycle (Brutsaert, 1982; Jackson, 1985). Empirical and semi-empirical models reported in the literature are based on relationships between evapotranspiration and a limited number of meteorological variables. A number of researchers have attempted to estimate evaporation values from various climatic variables (Burman, 1976; Coulomb, Legesse, Gasse, Travi, & Chernet, 2001; Gavin & Agnew, 2004; Linarce, 1967; Reis & Dias, 1998; Stephens & Stewart, 1963), and most of these methods require data that are not easily available. Furthermore, some of these methods are valid only under specific climatic and agronomic conditions, and they cannot be applied under conditions which are different from those they were originally developed for. Simple methods that have been reported (e.g., Stephens & Stewart, 1963) try to fit a linear relationship zbetween the explanatory variables. However, the process of evaporation is highly non-linear in nature, as evidenced by many of the estimation procedures (Sudheer, Gosain, Rangan, & Saheb, 2002).

* Corresponding author. Tel.: +1 5143987786.
E-mail addresses: mkgoyal@iitg.ernet.in (M.K. Goyal), jan.adamowski@mcgill.ca (J. Adamowski).

http://dx.doi.org/10.1016/j.eswa.2014.02.047

Several studies have emphasized the need for accurate estimates of evaporation in hydrologic modeling studies (e.g., Sudheer et al., 2002; Szilagyi & Jozsa, 2009). This requirement could be addressed through better models that address the inherent non-linearity in the process. Recently, machine learning approaches such as Artificial Neural Networks (ANN), Least-Squares – Support Vector Regression (LS-SVR), Fuzzy Logic, and Adaptive Neuro-Fuzzy Inference System (ANFIS) have been successfully applied in a number of diverse fields, including water resources. Most of the applications reported in the literature are related to estimation, prediction, and classification problems. In the hydrological context, recent studies have reported that ANN, LS-SVR, Fuzzy Logic, and ANFIS approaches may offer a promising alternative to traditional hydrological forecasting approaches, including evaporation (Cancelliere, Giusiano, Ancarani, & Rossi, 2002; Chang, Chang, Kao, & Wu, 2010; Cigizoglu & Kisi, 2005, 2006; Ismail, Shabri, & Samsudin, 2012; Kalra, Li, & Ahmad, 2013; Kisi, 2004a, 2004b, 2005a, 2005b; Kisi & Yildirim, 2005a, 2005b; Kumar, Raju, & Sathish, 2004; Minnes & Hall, 1996; Sudheer et al., 2002; Supharatid, 2003; Tayfur, 2002; Twarakavi, Misra, & Bandopadhyay, 2006).

Eslamiam, Gohari, and Malekian (2008) estimated monthly pan evaporation using ANNs and support vector machines with climatic variables such as air temperature, solar radiation, wind speed, relative humidity and precipitation as input data for the models. The results showed that both ANN and support vector machines provided accurate pan evaporation estimates. Kim and Kim (2008) applied ANN and genetic algorithm models for modeling pan evaporation and evapotranspiration. The study further confirmed the capabilities of ANN and genetic algorithm models as effective tools for estimation of pan evaporation and evapotranspiration. Tabari et al. (2009) estimated daily pan evaporation using artificial neural network (ANN) and multivariate non-linear regression (MNLR) methods in a semi-arid region of Iran. The results indicated that the ANN method provided the best estimates of daily pan evaporation in comparison to MNLR. Recently, Guo, Sun, and Ma (2011) studied the applicability of LS-SVR models for real-time prediction of daily reference crop evapotranspiration. The authors considered public weather forecast variables such as: minimum and maximum air temperature, average relative humidity, wind scale, and weather conditions. The authors compared the forecast of their LS-SVR model against Penman–Monteith estimated daily crop reference evapotranspiration. It was found that the LS-SVR model provided a 90% model efficiency score; it was concluded that the LS-SVR approach was a successful tool that can be used to measure daily crop reference evapotranspiration using public weather forecasts. The authors also recommended that future research be carried out in other catchments to further explore the applicability of the LS-SVR method as a crop reference evapotranspiration estimation tool. Similarly to Guo et al. (2011), Kisi (2013) investigated the applicability of LS-SVR for daily reference crop evapotranspiration estimation at two sites in Southern California. This study included empirical methods (Priestley-Taylor, Hargreaves, and Ritchie methods) and also considered feed forward ANNs as comparison methods to the LS-SVR model. The following variables were considered in developing the models: daily weather data, solar radiation, air temperature, relative humidity, and wind speed. The LS-SVR method outperformed all empirical methods as well as the ANN models, with coefficients of determination in excess of 0.96 for the best LS-SVR models at both sites. The most useful sets of predictors were solar radiation, air temperature, relative humidity, and wind speed.

Keskin, Özlem, and Dilek (2004) applied the Fuzzy Logic method to estimate daily pan evaporation based on meteorological data for Lake Eğirdiris and compared this with the Penman method. It was concluded that the Fuzzy Logic approach can be used to estimate daily pan evaporations effectively. Atiaa Alaa and Abdul-Qadir Amal (2012) used Fuzzy Logic for estimating monthly pan evaporation from meteorological data from Emara meteorological station in southern Iraq, and found that the method was useful. Kisi (2006) investigated the abilities of the ANFIS technique against ANN and SS methods for daily evaporation estimation and found that the ANFIS computing technique could be employed successfully in modeling evaporation processes from the available climatic data as it outperformed both techniques. Kumar, Kumar Jaipaul, and Tiwari (2012) developed ANN and ANFIS models to forecast monthly potential evaporation in Pantagar, India and found that the ANFIS model outperformed the ANN model. To the best knowledge of the authors, no studies have compared the ANN, LS-SVR, Fuzzy Logic, and ANFIS machine learning methods, and no studies have compared these newer machine learning methods with the traditional HGS and SS empirical methods for evaporation modeling/forecasting.

In this study, the potential of four new machine learning techniques (i.e., ANN, LS-SVR, Fuzzy Logic, and ANFIS) for the estimation of pan evaporation using climatic variables was investigated. Multiple input variable combinations were assessed using the Gamma Test (GT) before model development to provide insight to regarding the most useful sets of predictors to model pan evaporation. The performance of these models was compared with that of the traditional Hargreaves and Samani (HGS) and Stephens–Stewart (SS) empirical methods. The Hargreaves-Samani method is a well-known method to estimate daily incoming solar radiation values by using available data. The SS method was chosen following the suggestion of Al-Shalan and Salih (1987), who evaluated 23 well-known climatic methods of evaporation estimation and concluded that the Stephens–Stewart model performed the best of all the methods tested.

1.1. Study area

The study area, the Karso watershed in India, is a part of the Damodar Barakar catchment, and is situated between 85° 23′ 30″E to 85° 28′E longitude and 24° 12′N to 24° 18′N latitude with an elevation ranging from 390–650 m above MSL (Mean Sea Level) (Fig. 1). The geographical area of the watershed is approximately 27.93 km^2. The watershed receives an average annual rainfall of 1300 mm, and 75 percent of the rainfall occurs during the monsoon season (June to October). The minimum and maximum temperature varies in the range of 3–42 °C. The mean relative humidity varies from a minimum of 40 percent in April to a maximum of 85 percent in the month of July. The overall climate of the area can be classified as sub humid tropical. The soil is mainly sandy loam type and the soil depth ranges from 0 to 45 cm. The average land slope of the watershed varies from 0 to 8%, and the maximum slope of some hilly sections of the watershed is close to 22%. The dominant crop in the study area is paddy. Erosion problems are prevalent in the study area due to the rolling topography and improper agricultural management practices. Daily rainfall data of the watershed was collected from the automatic rain gauge station located at the outlet of the study watershed by the Soil Conservation Department, Damodar Valley Corporation (DVC), Hazaribag, Jharkhand, India. The datasets used in this study are discussed in the Input Variable Determination section and descriptive statistics of this dataset can be found in Table 1.

2. Methodology

This section provides a short introduction to the Gamma Test for identifying the most useful input variable set combinations, as well as descriptions of the ANN, LS-SVR, Fuzzy Logic, and ANFIS methods.