Journal of Hydrology 444–445 (2012) 78–89

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment

Hossein Tabari [a,*], Ozgur Kisi [b], Azadeh Ezani [c], P. Hosseinzadeh Talaee [d]

[a] Department of Water Engineering, Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran
[b] Canik Basari University, Faculty of Architecture and Engineering, Civil Engineering Department, Samsun, Turkey
[c] Department of Water Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran
[d] Young Researchers Club, Hamedan Branch, Islamic Azad University, Hamedan, Iran

ARTICLE INFO

Article history:
Received 14 September 2011
Received in revised form 15 February 2012
Accepted 5 April 2012
Available online 18 April 2012
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:
Reference evapotranspiration modeling
Support vector machines
Adaptive neuro-fuzzy inference system
Climate based equations
Multiple regression

SUMMARY

The accurate estimation of reference evapotranspiration (ET_o) becomes imperative in the planning and management of irrigation practices. The Penman–Monteith FAO 56 (PMF-56) model which incorporates thermodynamic and aerodynamic aspects is recommended for estimating ET_o across the world. However, the use of the PMF-56 model is restricted by the unavailability of input climatic variables in many locations and the option is to use simple approaches with limited data requirements. In the current study, the potential of support vector machines (SVM), adaptive neuro-fuzzy inference system (ANFIS), multiple linear regression (MLR) and multiple non-linear regression (MNLR) for estimating ET_o were investigated using six input vectors of climatic data in a semi-arid highland environment in Iran. In addition, four temperature-based and eight radiation-based ET_o equations were tested against the PMF-56 model. The accuracies of the models were evaluated by using three commonly used criteria: root mean square error (RMSE), mean absolute error (MAE) and correlation coefficient (r). The results obtained with the SVM and ANFIS models for ET_o estimation were better than those achieved using the regression and climate based models and confirmed the ability of these techniques to provide useful tools in ET_o modeling in semi-arid environments. Based on the comparison of the overall performances, it was found that the SVM6 and ANFIS6 models which require mean air temperature, relative humidity, wind speed and solar radiation input variables had the best accuracy.

1. Introduction

Evapotranspiration (ET) is defined as the loss of water to the atmosphere by the combined processes of evaporation from the soil and plant surfaces and transpiration from plants (Aytek, 2008). It is necessary to quantify ET for work dealing with water resource management or environmental studies. With regard to agricultural production, its measurement is greatly relevant in arid and semi-arid regions, where it is essential for determining crop water demand and consequently for designing and managing irrigation systems. The ET quantification frequently must be preceded by the determination of reference evapotranspiration (ET_o) (Lopez-Urrea et al., 2006).

Accurate estimation of ET_o in irrigated lands is necessary for improving the planning and efficient use of water resources. The application of lysimeters is the most common method for estimating ET_o. Unfortunately, lysimeters are unsuitable for monitoring evapotranspiration as compared to direct climate based measurement at weather stations. This is not only due to their cost and complexity, but also because the limited area of a typical weather station enclosure does not provide sufficient fetch from a representative surface for these measurements to be meaningful (Sentelhas et al., 2010). In practice, ET_o can be estimated using available climatic data from a weather station (Xing et al., 2008).

Many methods based on climatic data have already been developed to estimate ET_o in various climatic and geographic conditions. Recently, the Penman–Monteith FAO-56 combination equation (PMF-56) has been recommended by the Food and Agriculture Organization of the United Nations (FAO) as the standard equation for estimating ET_o (Allen et al., 1998). The Penman–Monteith equation has two advantages over many other equations. First, it can be used globally without any local calibrations due to its physical basis. Second, it is a well documented equation that has been tested using a variety of lysimeters (Gocic and Trajkovic, 2010). The main drawback of the PMF-56 equation is the large amount of data required; including solar radiation, wind speed, humidity, and temperature. Records of these input variables are sometimes of

* Corresponding author. Tel.: +98 911 2528074; fax: +98 121 2517310.
E-mail address: tabari.ho@gmail.com (H. Tabari).

0022-1694/$ - see front matter
http://dx.doi.org/10.1016/j.jhydrol.2012.04.007

questionable quality, incomplete, or not available for a given location especially in developing countries (Martinez and Thepadia, 2009). Thus, there is an urgent need to evaluate simpler ET_o methods relative to the PMF-56 method (Tabari, 2010; Tabari et al., 2011a). Practitioners and researchers need to be provided with guidance on the choice of the most appropriate ET_o method to be adopted when weather data are insufficient to apply the PMF-56 method.

The need for a simpler method to estimate ET_o has been stated in several studies (George et al., 2002; Trajkovic, 2009; Sabziparvar et al., 2010; Tabari and Hosseinzadeh Talaee, 2011). Berengena and Gavilán (2005) tested several ET_o equations in a highly advective semi-arid environment in Spain. The locally adjusted Penman and ASCE-PM equations performed the best, followed by the PMF-56 method. Benli et al. (2010) evaluated the performance of six commonly used ET_o estimation methods with different data requirements namely Penman–Monteith FAO-56, Priestley–Taylor, Radiation-FAO24, Hargreaves, Blaney–Criddle and Class A pan versus lysimeter data in a semi-arid highland environment in Turkey. They found that the PMF-56 and Hargreaves equations were the best options to estimate ET_o in the study area. In a study in the arid and semi-arid areas of Iran, the spatially distributed maps of ET_o were prepared using the Hargreaves equation (Sabziparvar and Tabari, 2010). The results showed that the estimated total monthly ET_o revealed a significant variation during the growing season (April–September) so that the study region experienced the highest and lowest monthly ET_o values of 250 and 80 mm in July and April, respectively.

ET_o is a complex non-linear phenomenon which depends on several climatic elements. Artificial neural networks (ANNs) are very appropriate for the modeling of non-linear processes, i.e. the case of ET_o. During the last decades many researchers have studied the reliability of ANN for estimating ET_o as a function of climatic variables (Trajkovic et al., 2003; Kişi, 2006, 2007; Chauhan and Shrivastava, 2008; Traore et al., 2010; Cobaner, 2011). Recently, the neuro-fuzzy and support vector machines (SVM) techniques have been successfully used for ET_o estimation. Kişi and Öztürk (2007) investigated the accuracy of the adaptive neuro-fuzzy inference system (ANFIS) technique in estimation of ET_o in Los Angeles, California. The results revealed that the ANFIS models could be employed successfully in modeling the ET_o process. Aytek (2008) examined the co-active neuro-fuzzy inference system (CANFIS) for daily ET_o modeling by using daily atmospheric parameters. They found that CANFIS can be proposed as an alternative ET_o model to the existing conventional methods. Kişi and Çimen (2009) studied the potential of SVM in modeling ET_o in central California. They found that the SVR model could be embedded as a module for estimating ET_o data in hydrological modeling studies. Shiri et al. (2012) evaluated the Gene Expression Programming (GEP), ANFIS, Priestley–Taylor and Hargreaves–Samani models for estimating ET_o in Basque Country (Northern Spain). The comparison results showed that the GEP model performed better than the others.

Due to some difficulties in working with artificial intelligence approaches, some investigators have used fast, simple and straightforward statistical methods such as regression models for expressing the dependence of a response variable on several independent (predictor) variables. In recent years, regression models have been successfully employed in modeling a wide range of hydrologic processes like soil temperature (Bilgili, 2010; Tabari et al., 2011b); snow water equivalent (Tabari et al., 2010b; Marofi et al., 2011); low and flood flows (Engeland and Hisdal, 2009; Eslamian et al., 2010) and evaporation (Tabari et al., 2010a).

The main aims of this study were (i) to examine four temperature-based (Blaney–Criddle, Hargreaves and two new versions of the Hargreaves equation) and eight radiation-based (Makkink, Turc, Jensen–Haise, McGuinness–Bordne, Priestley–Taylor, Ritchie, Abtew and Irmak) equations against the PMF-56 model as reference in a semi-arid highland environment in Iran (ii) to develop multiple linear regression (MLR) and multiple non-linear regression (MNLR) models for ET_o modeling (iii) to evaluate the ANFIS model using different membership functions with various input combinations of climatic variables (iv) to evaluate the performance of the SVM model with different kernels to predict ET_o and (v) to compare the performances of ANFIS, SVM, MLR, MNLR and climate based models.

2. Materials and methods

2.1. Study area and climate dataset

The weather data for this study were obtained from the Nozheh station (35°12′N, 48°43′E; 1680 m a.s.l.) located in Hamedan city in western Iran. Hamedan lies in the vicinity of the Alvand Mountains and has a cold, mountainous climate, with snowy winters. In fact, it is one of the coldest cities in Iran. The temperature may drop to below −30 °C on the coldest days. The mean monthly temperature in the study area varies from −3 °C in January to 25 °C in July, with an annual mean of 11 °C (Table 1). The mean annual rainfall is 330 mm. Winter precipitation is mainly snow, lasting some 6–8 months in the mountainous areas and 1–2 months on the plateau. The rest of the precipitation is provided by scarce spring and fall rains.

The weather data were collected from the Islamic Republic of Iran Meteorological Organization (IRIMO) (www.weather.ir). The data is composed of mean, maximum and minimum air temperatures, relative humidity, dew point temperature, water vapor pressure, wind speed, atmospheric pressure, precipitation, solar radiation and sunshine hours for the period 1986–2005. The first 15 years (1986–2000) data were used to train the ANN and SVM, and the remaining data were used for validation. It should be noted that the mean monthly values of the weather data were used for the analysis.

2.2. Climate based methods

In this study, those ET_o equations that require only air temperature as an input variable are considered as temperature-based equations. The temperature-based equations evaluated here are Blaney–Criddle (Blaney and Criddle, 1950; Doorenbos and Pruitt, 1977), Hargreaves (Hargreaves and Samani, 1985; Jensen et al., 1997) and two new versions of the Hargreaves equation (Allen, 1993; Droogers and Allen, 2002). Allen (1993) modified the Hargreaves equation by fitting coefficients based on monthly calculations of ET_o by the PMF-56 method using the FAO Climwat data set (Smith, 1993) consisting of 3200 stations and using lysimeter measurements of ET_o from Davis, California. A second attempt was made by Droogers and Allen (2002) to improve the agreement of the Hargreaves equation with the PMF-56 method using the IWMI (International Water Management Institute) Climate Atlas data grids. Comparisons around the globe using the grid were used to adjust two parameters in the original Hargreaves equation.

Eight radiation-based equations were used in this study: Makkink (Makkink, 1957), Turc (Turc, 1961), Jensen–Haise (Jensen and Haise, 1963), McGuinness–Bordne (McGuinness and Bordne, 1972), Priestley–Taylor (Priestley and Taylor, 1972), Ritchie (Jones and Ritchie, 1990), Abtew (Abtew, 1996), and Irmak's solar radiation-based equation (Irmak et al., 2003). Readers interested in the details and formulations of the temperature and radiation based equations are referred to Tabari (2010) and Tabari et al., 2011a.