Computers and Electronics in Agriculture 122 (2016) 112–117

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Daily pan evaporation modeling using chi-squared automatic interaction detector, neural networks, classification and regression tree

Ozgur Kisi [a,*], Onur Genc [b], Semih Dinc [c], Mohammad Zounemat-Kermani [d]

[a] Department of Civil Engineering, Canik Basari University, Samsun, Turkey
[b] Department of Civil Engineering, Meliksah University, Kayseri, Turkey
[c] Department of Computer Science, University of Alabama, Hunstville, AL, USA
[d] Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

ARTICLE INFO

Article history:
Received 5 September 2015
Received in revised form 20 January 2016
Accepted 24 January 2016
Available online 5 February 2016

Keywords:
ANN
CHAID
C&RT
Machine learning algorithms
Pan evaporation

ABSTRACT

Accurate prediction of daily pan evaporations plays a crucial role in water resources management since it has direct effect on water reservoirs, therefore drinking water supply systems. In this study, three machine learning (ML) methods such as classification and regression tree (C&RT), chi-squared automatic interaction detector (CHAID) and artificial neural networks (ANN) are applied to predict the daily pan evaporations in Ankara and Polatli stations in Turkey, both of which have dry climatic conditions. First part of the study focuses investigating the ability of C&RT, CHAID and ANN methods on predicting the daily pan evaporations in Ankara and Polatli stations have been investigated, separately. The estimations are made by using the inputs from the associated stations itselves. However, in the second part of the study, daily pan evaporations in Polatli station have been estimated by using the inputs obtained from Ankara station. The inputs for both parts consist of daily climatic data of maximum and minimum temperature, solar radiation, relative humidity and wind speed. Therefore, this study not only aims to compare the ability of ML models on two different stations, but it also investigates the ability of these models on prediction of daily pan evaporations in a station by using the inputs that were obtained from a nearby station. By comparing these models, it has been revealed that, ANN model has performed slightly better than the other models in both applications. Therefore, it can be concluded that daily pan evaporations could be successfully predicted by employing ANN model in both type of applications.

1. Introduction

Evaporation, which is considered as a major component of hydrological cycle, plays an important role in many applications in water resource management projects and environmental studies. Therefore, proper estimation and prediction of evaporation especially in the arid and semiarid areas is of great importance to integrated water resources management and modeling studies (Ribot et al., 2005; Nourani and Fard, 2012; Malik and Kumar, 2015). Generally two main approaches of direct and indirect methods can be applied for calculating and estimation of evaporation. One of the direct methods for evaporation measurement is the pan evaporation (Ep) (Eslamian et al., 2008; Goyal and Ojha, 2011). Although using pans for measuring evaporation seems to be the best and realistic technique, but due to the shortage of pan evaporation data as well as high expenses of the installing and maintaining evaporation pans and their measurement devices, applying indirect methods for the estimation of EP seems necessary. In other words, using the indirect methods, the evaporation can be estimated from the other meteorological parameters (e.g. temperature, relative humidity, solar radiation rainfall, wind speed).

In this respect, data driven methods (such as artificial neural networks) have been successfully applied for pan evaporation modeling and proved their high capacity in simulating meteorological phenomenon (Bruton et al., 2000; Sudheer et al., 2002; Kisi, 2006; Dogan et al., 2007; Kim and Kim, 2008; Piri et al., 2009; Moghaddamnia et al., 2010; Samui and Dixon, 2012; Lin et al., 2013).

In most recent studies, Sattari et al. (2013) investigated the potential of back propagation neural network and M5 model tree based regression approaches to model monthly reference evapotranspiration using climatic data. They claimed that the ANN model performed better that the M5 model tree. Chang et al. (2013) proposed a hybrid model (BD) that combines back-propagation neural networks (BPNN) and dynamic factor analysis

* Corresponding author.

http://dx.doi.org/10.1016/j.compag.2016.01.026
0168-1699/

(DFA) to simultaneously estimate pan evaporation at multiple meteorological stations in northern Taiwan. The results demonstrated that the proposed BD model has good reliability and applicability in simultaneously estimating pan evaporation for multiple meteorological stations. Goyal et al. (2014) investigated the abilities of 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. Based on the comparison, it was found that the Fuzzy Logic and LS-SVR approaches can be successfully employed in modeling the daily evaporation process from the available climatic data. Kim et al. (2014) evaluated a multilayer perceptron neural networks model (MLP-NNM) and a cascade correlation neural networks model (CCNNM) for estimating daily pan evaporation for inland and coastal stations in Republic of Korea. They reported that in general statistical results of CCNNM are better than those of MLP-NNM during the test period for homogeneous and nonhomogeneous weather stations. Kim et al. (2015) develops and applied some soft computing models including the MLP-NNM, Kohonen self-organizing feature maps-neural networks model (KSOFM-NNM), and gene expression programming (GEP) to predict daily PE in a dry climate region of south-western Iran. Comparison of soft computing models and multiple linear regression (MLR) demonstrated the superiority of MLP-NNM, KSOFM-NNM, and GEP over MLR. Kisi (2015) investigated the accuracy of least square support vector machine (LSSVM), multivariate adaptive regression splines (MARS) and M5 Model Tree in modeling Ep. The overall results showed that the LSSVM could be successfully applied 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. Tezel and Buyukyildiz (2015) studied the usability of multilayer perceptron (MLP), radial basis function network (RBFN)) and ε-support vector regression (SVR) to estimate monthly pan evaporation. They claimed that the ANN and ε-SVR had similar results, however, the ANNs and ε-SVR methods performed better than the empirical Romanenko and Meyer methods. Malik and Kumar (2015) applied artificial neural network, co-active neuro-fuzzy inference system (CANFIS) and MLR to simulate daily pan evaporation (Ep) at Pantnagar in India. The results showed that the accuracy of ANN model was superior to the CANFIS and MLR models; however, the performance of CANFIS models was better than MLR models.

The objective of this study is to investigate capability and usability of chi-squared automatic interaction detector (CHAID), artificial neural networks (ANNs), classification and regression tree (C-RT) methods for modeling daily pan evaporation using meteorological data at two automated weather stations in Turkey (Ankara Station and Polatli Station). To the knowledge of authors, no similar study has been reported using the above mentioned methods for modeling Ep. Hence, the main contribution of this study is referred to applying CHIAD and C&RT methods as comparative approaches to the frequent ANN model.

2. Materials and methods

Turkey, which has an irregular topography, is situated in Mediterranean geographical and climatically location. Therefore, there are a lot of distinct geographic weather and climatic regions in Turkey. The dataset utilized in this paper were collected in Ankara and Polatli stations that are located in the central Anatolia region, which has a steppe climate, in Turkey (Fig. 1). In these stations, temperatures vary between the degrees −25 °C to 40 °C, with rainfall of only 382 mm per year. Central Anatolia is evaluated as under the drought risk region by the Turkish State Meteorological Service (DMI) (URL 1).

The daily meteorological variables from two automated weather stations, Ankara Station, (39°97′25″N, 32°86′39″E), and Polatli Station (39°58′34″N, 32°16′24″E), operated by DMI are used in this paper. The pan evaporation observations are carried out after the frost season every year. Class A pans are used to obtain the open water evaporation in these stations by the DMI. These stations are an automatic meteorological observation station, automatic meteorological observation station, AMOS, and synoptic weather station. The Geographical features of the Ankara and Polatli stations have good conditions for measurement of evaporation. They both have 9 × 9 square meter areas. They have been partially grassed through their around. The sunshine readily reaches to Class A pan in these stations. There are no building which is not barrier to sunshine around the Class A pans. All meteorological parameters are measured every 3 h in these stations. The altitudes of the Ankara and Polatli stations are 887 and 886 m above MSL (mean sea level), respectively. The data set consists of thirty years (1985–2005) of daily average climatic data records of the maximum and minimum temperature, solar radiation, relative

Fig. 1. The location of the Ankara and Polatlı stations.