Computers and Electronics in Agriculture 144 (2018) 114–125

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Assessing the potential of data-driven models for estimation of long-term monthly temperatures

Saeid Mehdizadeh

Department of Water Engineering, Urmia University, Urmia, Iran

ARTICLE INFO

Keywords:
Long-term monthly air temperatures
Data-driven models
Estimation

ABSTRACT

Having information on air temperature components consisting minimum (T_{min}), maximum (T_{max}) and mean (T) temperatures plays a crucial role in various aspects of agriculture such as agricultural meteorology, soil science, agronomy, etc. The present study explores the performance of four data-driven models including artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM) and multivariate adaptive regression splines (MARS) for estimation of long-term monthly T_{min}, T_{max} and T. For this purpose, the long-term monthly temperatures of 50 stations all over Iran were used. The data of 35 and 15 stations were utilized to train and test the models, respectively. To feed the models, the geographical information (latitude, longitude, altitude) and periodicity component (the number of months) were employed as input parameters. The obtained results demonstrated that the long-term monthly temperatures of the studied regions can be estimated as a function of geographical information and periodicity component. Comparing the overall performance of the models at training stage revealed that the ANN outperformed the other models for estimating the long-term monthly T_{min}, T_{max} and T. That's while the SVM, ANN and ANFIS had superiority over the others at testing stage for estimation of the long-term monthly T_{min}, T_{max} and T, respectively. Furthermore, the MARS model presented the weakest performance for estimating the long-term monthly temperatures at both training and testing stages.

1. Introduction

Air temperature is one of the most important parameters required for different fields of agricultural sciences such as horticulture, agronomy, soil science, agricultural meteorology and so on. It is an important parameter which is used to determine the appropriateness of crops with the desired area. Moreover, optimum growth of plants depends on the favorable soil temperature so that it affects the planting date of seeds. Soil temperature is influenced by various factors such as air temperature, solar radiation, precipitation and other effective parameters. Meanwhile, air temperature can be considered as one of the most important factors affecting soil temperature. Given the issues raised, knowledge of the air temperature is essential to achieve maximum agricultural products.

In the recent years, the use of data-driven models has received much attention in various engineering sciences for modeling of the meteorological, hydrological and climatological phenomena. It can be pointed out some researches such as prediction of solar radiation (Wang et al., 2016, 2017a, 2017b; Mehdizadeh et al., 2016), pan evaporation modeling (Wang et al., 2017c, 2017d, 2017e), estimating the precipitation/rainfall (Mehdizadeh et al., 2017a, 2017b). Some of the data-driven models are artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM), multivariate adaptive regression splines (MARS) and so on. Recently, the data-driven models have widely been utilized to estimate air temperatures. Herein, some of these studies are briefly presented:

Jang et al. (2004) investigated the accuracy of multilayer feed-forward neural networks (FFNN) for estimating air temperatures in Southern Québec (Canada) using Advanced Very High Resolution Radiometer (AVHRR) images. They stated that the accuracy of air temperature estimates was improved by the inclusion of Julian day. Ustaoglu et al. (2008) used three artificial neural networks models namely feed-forward back propagation (FFBP), radial basis function (RBF) and generalized regression neural networks (GRNN) for estimation of daily minimum, maximum and mean temperatures at Geyve and Sakarya, Turkey. FFBP and RBF models had the best performances in predicting daily minimum and maximum temperatures. However, FFBP performed better than others for estimation of daily mean temperature. Dombayci and Golcu (2009) estimated mean daily temperatures by the ANN in Denizli, South-Western Turkey. The obtained results showed that the ANN can be successfully used to predict mean daily temperatures. Smith et al. (2009) predicted year-round temperature using the

E-mail address: saied.mehdizadeh@gmail.com.

https://doi.org/10.1016/j.compag.2017.11.038
Received 18 August 2017; Received in revised form 10 November 2017; Accepted 26 November 2017

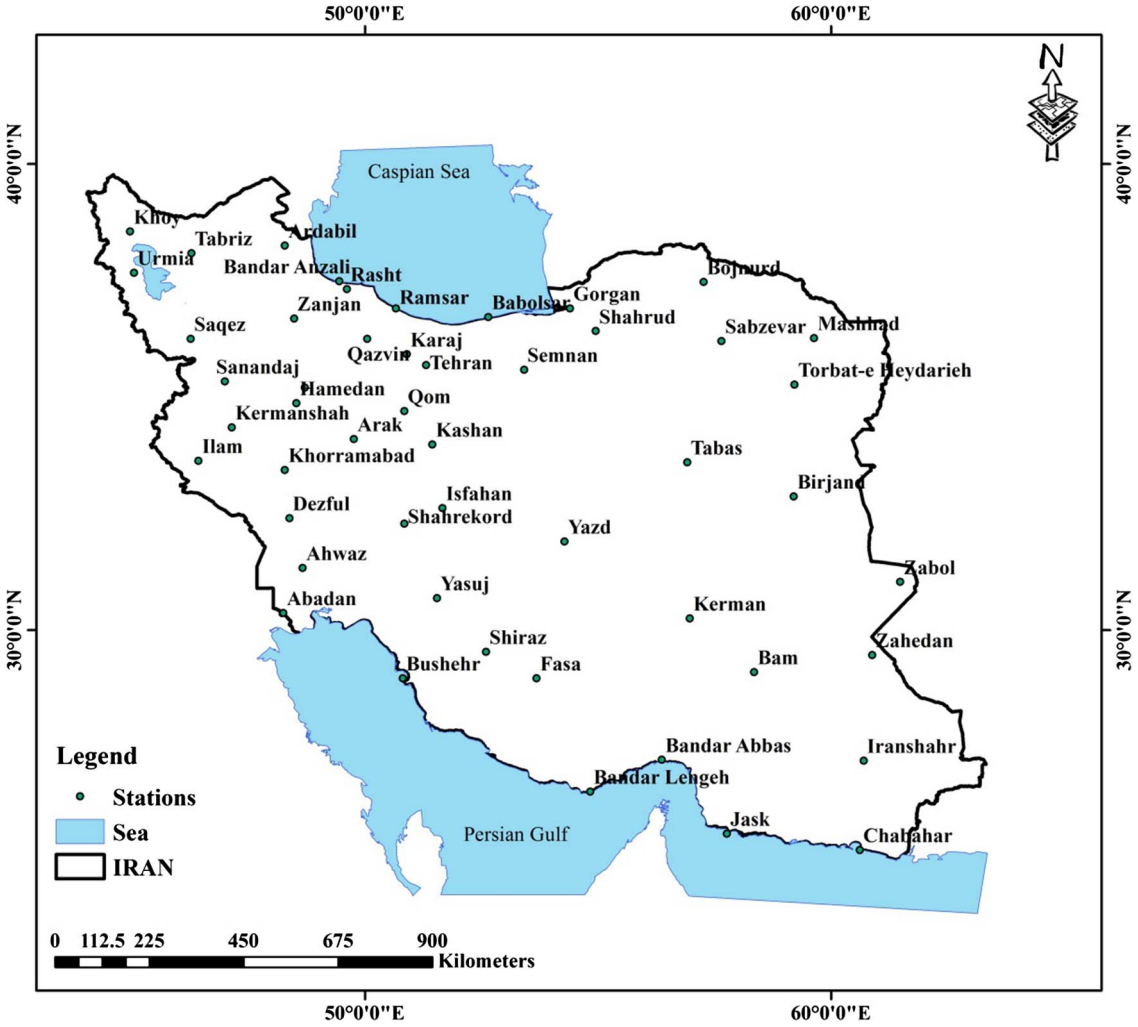

Fig. 1. Spatial distribution of the studied stations in Iran.

dataset of Georgia Automated Environmental Monitoring Network. It was found that the ANN approach provides the possibility of year-round temperature prediction. Also, the accuracy of ANN was increased by considering the rainfall as input. Bilgili and Sahin (2010) estimated monthly mean temperatures and precipitation by the ANN at 76 stations in Turkey. The obtained results demonstrated that the geographical information and the number of months can be employed to estimate long-term monthly rainfall and temperature in the studied regions. Kisi et al. (2011) investigated the ability of proposed wavelet-gene expression programming (W-GEP) and conventional GEP models to estimate daily and monthly air temperatures at Urmia and Mahabad, Iran. It was concluded that the hybrid W-GEP had better accuracy than single GEP. Paniagua-Tineo et al. (2011) compared the performance of SVM, multi-layer perceptron (MLP) and extreme learning machine (ELM) for predicting the daily maximum temperature at 10 stations considered in Germany, Holland, Spain, Norway and Russia. The SVM showed better results compared to the MLP and ELM models. Sahin (2012) modeled long-term monthly air temperatures at 20 stations in Turkey using the ANN and remote sensing approaches. He stated that these techniques can be used for estimating the air temperatures as a function of geographical information, the number of months and monthly mean land surface temperatures. Cobaner et al. (2014) evaluated the accuracy of ANN, ANFIS and multiple linear regression (MLR) to estimate monthly mean temperatures at 275 stations in Turkey. They employed four parameters as inputs consisting latitude, longitude, altitude and the number of months. The results indicated that the ANFIS outperformed the ANN and MLR methods. Salcedo-Sanz et al. (2016) predicted monthly air temperature in Australia and New Zealand by the SVM, MLP, MLR, and baseline persistence and stationarity approaches. Their findings revealed that the SVM presented the best accuracy.

Literature review shows that the geographical information and periodicity component have been used as predictors in some studies such as estimation of long-term monthly precipitation (Kisi and Sanikhani, 2015), soil temperature (Mehdizadeh et al., 2017c) and reference evapotranspiration (Kisi et al., 2015). In this study, the applicability of four data-driven models including ANN, ANFIS, SVM and MARS is assessed for estimating the long-term monthly minimum (T_{min}), maximum (T_{max}) and mean (T) air temperatures at 50 selected stations in Iran. For this purpose, the geographical information (latitude, longitude, altitude) and the periodicity component (the number of months) are used to feed the models.

2. Materials and methods

2.1. Study area and data used

Study region of this research is Iran which is situated in the Southwest of Asia. Iran is located between the latitudes of 40° and 25° N and the longitudes of 44° and 64° E. Fig. 1 shows the spatial distribution of 50 considered stations all over the country. Moreover, the summary of geographical information (latitude, longitude, altitude), as well as the long-term annual temperatures of the studied stations are given in Table 1.