Geoderma 330 (2018) 52–64

Contents lists available at ScienceDirect

Geoderma

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

Non-tuned data intelligent model for soil temperature estimation: A new approach

Hadi Sanikhani[a,*], Ravinesh C. Deo[b], Zaher Mundher Yaseen[c], Okan Eray[d], Ozgur Kisi[e]

[a] University of Kurdistan, Agriculture Faculty, Water Eng. Dept., Sanandaj, Iran
[b] University of Southern Queensland, Institute of Agriculture and Environment, International Centre for Applied Climate Sciences, School of Agricultural Computational and Environmental Sciences, QLD 4300, Australia
[c] Sustainable Developments in Civil Engineering Research Group, Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam
[d] International Black Sea University, Faculty of Computer Technologies and Engineering, Tbilisi, Georgia
[e] Ilia State University, Faculty of Natural Sciences and Engineering, Tbilisi, Georgia

ARTICLE INFO

Handling Editor: Morgan Crisitne L.S.

Keywords:
Decision-support systems
Extreme learning machine
Artificial neural network
M5 model tree
Agricultural precision
Soil temperature

ABSTRACT

In knowledge-based decision-support systems, soil temperature (ST) estimation can be considered as the core modeling task used for investigating the dynamics of solar energy exchange between the land surface and the sub-surface soil layers. In addition, the impacts of meteorological processes on energy uptake into the underlying soils and the design of more resilient and sustainable agricultural systems for improved crop health and resource management. In this paper, monthly ST estimation at various soil depths (i.e., 5, 50 and 100 cm) is performed by applying data-intelligent machine learning models: extreme learning machine (ELM), artificial neural network (ANN) and M5 Model Tree (M5 Tree). The predictive models are trained using meteorological information from two stations (i.e., Mersin and Adana, Turkey). The models are constructed using monthly input variables, including air temperature (T-air), windspeed (W), relative humidity (RH), solar radiation (SR), while the objective variable is the soil temperature measurement at 5, 50 and 100 cm depths for the period 1986–2010. Multi-objective performance criteria are applied to diagnose the predictive accuracy of the data-intelligent models, entailing the correlation coefficient (R), root mean square error (RMSE), mean absolute error (MAE), Willmott's Index (WI), Nash Sutcliffe's coefficient (Nash) and the Legates & McCabe's Index (LMI). Based on the tested dataset, the ELM model is yielded the most accurate performance for Mersin station compared to the lower performance of ANN and M5 Tree models. For this case, the most accurate performance is attained for the ST estimation at a depth of 50 cm, with the highest value of $R = 0.992$, WI = 0.999, Nash = 0.981, LMI = 0.879 and the lowest value of the relative RMSE and MAE values (i.e., 4.7 and 4.0%, respectively), attained using T-air, W, RH and periodicity as the predictor variables. The performance results for the Adana station behave differently, where ELM is seen to exceed the ANN and M5 Tree models for ST estimation at the depths of 5 and 100 cm. On the other hand, the ANN model performs marginally better than the ELM and the M5 Tree model at a depth of 50 cm with a different set of input combinations. The assessment of the estimation skills reveals the efficacy of the ELM over the counterpart benchmark models. In accordance with the present results, it is concluded that ELM model can be applied as an ideal decision-support tool for soil temperature estimation at multiple depths, whilst ensuring that an appropriate combination of meteorological inputs is applied to yield an optimal model.

1. Introduction

Soil temperature is a micro-meteorological property that assumes an imperative role in environmental management and bio-meteorological processes, acting to balance the interaction of heat energy between the atmosphere and the land surface (Keshavarzi et al., 2015; Zhao et al., 2009). In addition, soil temperature has a significant role in different ecosystems, including forests and deserts, agricultural applications, and various water resources and hydrologic engineering aspects (Lai et al., 2012; Peng et al., 2009).

The estimation of soil temperature is very important for agriculture, and this property is greatly affected by the different environmental factors that include meteorological parameters, such as solar radiation, air temperature, physical soil parameters (surface albedo, water

* Corresponding author.
E-mail addresses: h.sanikhani@uok.ac.ir, hsanikhani12@gmail.com (H. Sanikhani).

https://doi.org/10.1016/j.geoderma.2018.05.030
Received 12 September 2017; Received in revised form 9 April 2018; Accepted 25 May 2018

content, soil texture and capacity of infiltration; topographical variables) and the other land surface characteristics (Kang et al., 2000; Liang et al., 2014).

Recently, the implementation of soft computing models has shown a remarkable progress on earth science researches (Chen et al., 2015; Diop et al., 2018; Elzwayie et al., 2016; Gholami et al., 2015; Hameed et al., 2016; Taormina et al., 2015; Wang et al., 2015). Within the topic of soil temperature simulations, there are several previous studies that have accomplished an acceptable degree of success in this regard. Different statistical interpolation methods and various modeling techniques have been used, including the multiple linear regression (MLR), gene expression programming (GEP), adaptive neuro fuzzy inference system (ANFIS), stepwise regression (SR) and the genetic programming (GP)-based models (Jebamalar et al., 2012; Kim and Singh, 2014; Kisi et al., 2016; Mehdizadeh et al., 2017).

One of the earliest research conducted on soil temperature estimation conducted by (Yang et al., 1997), they developed a model to estimate ST at different depths (i.e., 10, 50 and 150 cm) using the application of ANN technique. In that study, five years of climatic-based input data, measured at a weather station at the Central Experimental Farm in Ottawa, Ontario, Canada were used. George (2001) applied an ANN model to estimate weekly scale of ST, assuming that the ST data was affected by the relative humidity, wind speed and air temperature. A single layer neural network was considered, which incorporated the McCulloch-Pitts type neuronal system with the Widdrow-Hoff algorithm applied to train the network. A estimation of daily ST data at six depths (i.e., 5, 10, 20, 30, 50 and 100 cm) at the four stations of Iran conducted by (Tabari et al., 2011). In this study, the artificial neural network model was developed by using a Multi-Layer Perceptron (MLP) model and tested in an arid region where only four climatic datasets, comprised of: air temperature, solar radiation, relative humidity and precipitation were used as the model's inputs. Bilgili et al. (2013) estimated monthly mean ST data by using a nonlinear regression, linear regression and MLR model for different soil depths (i.e., 5, 10, 20, 50 and 100 cm) where the greatest accuracy was found at a depth of 5 cm. Wu et al. (2013) estimated the monthly mean ST data at the depth of 10 cm by using the ANN method.

The study of Napagoda and Tilakaratne (2012) used an MLP model to estimate ST at different depths of 5 cm and 10 cm in the Bathalagoda area. Ozturk et al. (2011) developed a set of data-intelligent models for the monthly estimation of ST data using the classical MLP at different depths with standard geographical and meteorological data including altitude, latitude, longitude, month, year, monthly solar radiation, monthly sunshine duration and monthly mean air temperature. Bilgili (2011) developed a set of models to estimate monthly mean ST with an ANN algorithm based on SR at Izmir in Turkey where a total of 7 neighboring measuring stations data were adopted in the model development and the model evaluation phases. All the studies mentioned above used ANN method which has some disadvantages (e.g., slower in learning, higher computation time and less accuracy) compared to ELM.

In another study, Mehdizadeh et al. (2017) compared an ANN, ANFIS and GEP methods to estimate monthly mean soil temperature, where the best accuracy was obtained from the ANFIS model. Hosseinzadeh Talaee (2014) estimated daily ST by using an ANFIS model using meteorological data at different depths in the region of Isfahan, Iran. In this study, the best ST estimations were obtained using ANFIS model, which included the Takagi–Sugeno–Kang (TSK) model that used Gaussian membership function. Abyaneh et al. (2016) applied an ANN model and a co-active neuro fuzzy inference system (CANFIS) for an estimation of ST using air temperature time series in different (i.e., dry and humid) climate in Iran. The developed ANN model resulted in a good accuracy for the dry climate. Kisi et al. (2016) used an ANN, ANFIS, and the GP model to estimate the monthly ST at different depths using climatic data of air temperature, wind speed, solar radiation, and relative humidity in Adana and Mersin in Turkey and showed that the GP model mostly performed better than the ANN and ANFIS in monthly ST. However, the ANN models outperformed the ANFIS and GP model for estimating the monthly ST of Mersin Station without local climatic inputs. Notably, this was the first study that estimated the monthly ST with a GP model. It was evident from the related literature; the previous studies have generally used the ANN, ANFIS and GEP methods in the estimation of ST. The GEP and ANFIS or CANFIS models are associated with several disadvantages compared to the ELM model. The ANFIS (CANFIS) methods with large rule bases and high number of parameters (or weights) in its structure have high computational time and sometimes overtraining problem (Tahmasebi and Hezarkhani, 2010). GEP method uses evolutionary computation (genetic algorithm) and therefore it has slower approximation compared to ANN method. On the other hand, in application of intelligent models to hydrological or hydrogeological data, there is no algorithm that is superior to all problems as echoed by (Bárdossy and Singh, 2008). This necessitates the comparative studies including application of various models on ST estimation. Although several types of data intelligent models (including ANN, ANFIS, GP, GEP, etc.) were used, the extreme learning machine (ELM) abd M5 Tree models utilized in this study as a new data-intelligent modeling strategies, which have not been explored previously in the estimation of soil temperature.

The premise of embracing the ELM model as a data-intelligent tool is certainly a desirable task by virtue of its fast and efficient learning ability, universal approximation capability, user-friendly and enigmatic-less design and its powerful non-linear estimation skill. Such advantageous features suffices the ELM model to be potentially used as a real-time decision-support tool, and the construction of a real-time, expert system in agriculture, water management and the estimation of micro-meteorological events (Akusok et al., 2017). Several recent studies have demonstrated the ELM model to be an efficient connectionist algorithm that were applied in data-intelligent frameworks for solving regression problems and practical decision system problems (e.g., (Avci, 2013; Chen and Ou, 2011; Deo et al., 2017; Scardapane et al., 2014)). Originally, the ELM model was proposed by Huang and Siew (2004) as a pertinent learning tool that is able to learn much faster and more efficiently than its conventional counterparts (i.e., the ANN) model. Basically, the ELM model is based on a simple, three-layer feed forward neural network (Wang et al., 2012). Due to the randomized selection of hidden neurons and biases, the ELM model is able to reduce the required computational time for feature extraction and prediction process compared to several other learning algorithms (e.g., the ANN, support vector machines, evolutionary computing, ANFIS, etc.).

In recent years, the ELM technique has successfully been applied in many different research fields, such as the function approximation (Huang et al., 2006a, 2006b), regression-based modeling (Yu et al., 2014) and classification problems (Akusok et al., 2015). Furthermore, the ELM model has been able to incredibly enhance the neural training speed and overall prediction accuracy (Huang et al. 2011), which was evident in a number of recent studies (Deo et al., 2016). In spite of the widespread embracement of the ELM model as an efficient data intelligent tool, its application in the area of soil temperature estimation (i.e., as a potential decision-support tool) is yet to be fully explored.

The primary contribution of this research paper is to design and evaluate a non-tuned predictive model based on the ELM algorithm applied of long-term (i.e., monthly) soil temperature (ST) estimation. The predictive ability of the proposed ELM model is also validated with respect to the conventional ANN and M5 Tree-based data-intelligent models. Upon designing the ELM model, the predictive performance is also assessed in respect to the three different soil temperature elevations: at a depth of 5, 50 and 100 cm. Through a statistical assessment of the observed and estimated ST data and a diagnostic evaluation of the model is performed at a real agricultural district (namely the Adana and Mersin) located in Turkey.