

Research Article

Comparative Study of M5 Model Tree and Artificial Neural Network in Estimating Reference Evapotranspiration Using MODIS Products

Armin Alipour, Jalal Yarahmadi, and Maryam Mahdavi

Department of Irrigation and Drainage Engineering, College of Aburaihan, University of Tehran, P.O. Box 33955-159, Pakdasht, Tehran, Iran

Correspondence should be addressed to Armin Alipour; a.alipour@ut.ac.ir

Received 29 August 2014; Accepted 18 November 2014; Published 17 December 2014

Academic Editor: Ines Alvarez

Copyright © 2014 Armin Alipour et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Reference evapotranspiration (ET_o) is one of the major parameters affecting hydrological cycle. Use of satellite images can be very helpful to compensate for lack of reliable weather data. This study aimed to determine ET_o using land surface temperature (LST) data acquired from MODIS sensor. LST data were considered as inputs of two data-driven models including artificial neural network (ANN) and M5 model tree to estimate ET_o values and their results were compared with calculated ET_o by FAO-Penman-Monteith (FAO-PM) equation. Climatic data of five weather stations in Khuzestan province, which is located in the southeastern Iran, were employed in order to calculate ET_o . LST data extracted from corresponding points of MODIS images were used in training of ANN and M5 model tree. Among study stations, three stations (Amirkabir, Farabi, and Gazali) were selected for creating the models and two stations (Khazaei and Shoeybie) for testing. In Khazaei station, the coefficient of determination (R^2) values for comparison between calculated ET_o by FAO-PM and estimated ET_o by ANN and M5 tree model were 0.79 and 0.80, respectively. In a similar manner, R^2 values for Shoeybie station were 0.86 and 0.85. In general, the results showed that both models can properly estimate ET_o by means of LST data derived from MODIS sensor.

1. Introduction

Decline in availability of water for agriculture is one of the most serious challenges facing human life that has affected agricultural production in some arid and semiarid regions around the world. Determining future demands of water for agriculture section includes computation of several factors such as runoff, groundwater, precipitation, and evapotranspiration (ET) [1]. ET is identified as the combination of two different processes including evaporation from the soil surface and crop transpiration [2].

Accurate and reliable estimates of ET are necessary to determine temporal variations in irrigation requirement, improve allocation of water resources, and evaluate the effect of changes in land use and crop patterns on the water balance [3]. Considering difficulties in direct measurement of ET [4], this parameter is estimated through reference evapotranspiration (ET_o) and crop coefficient (K_c) for a specific crop [5].

Therefore, calculation of ET_c (evapotranspiration of the given plant) and subsequently crop water requirement as irrigation water depend on ET_o estimates. ET_o is defined as the rate of ET from a reference surface in such a way that the surface is assumed to be covered with a hypothetical grass with specific characteristics [2]. A large number of methods have been offered in order to model ET_o [2, 6, 7]; the majority of these methods are extremely complicated and rely on weather data such as temperature, solar radiation, wind speed, and air humidity [1, 8]. ET_o is principally calculated by physically based equations (e.g., Penman-Monteith (PM) equation) or by empirical relationships between meteorological variables.

The FAO-Penman-Monteith (FAO-PM) method is currently recommended as an accepted standard technique for computing ET_o . This method employs a variety of complex equations, which are based on weather data, including air temperature, humidity, radiation, and wind speed [9]. The requirement of various weather data, which are not mainly

available in many regions, is the major limitation of this method [10]; in addition, it seems that the reordered data are not accurate enough, especially in developing countries [11].

From the viewpoint of temporal and spatial variations of climatic data, weather stations density (the number of stations per unit of area) is not generally adequate for demonstrating spatiotemporal variations. Today, remote sensing techniques, which gather land surface information from a broader geographic scope, are known as an effective tool and methodology for extracting the ground parameters that can be used to estimate ET at regional scales [12]. To this end, in the basins that may face shortage of measured data, remote sensing data can considerably supply needed information such as albedo, leaf area index, and land surface temperature (LST) [13]. Recent advances in remote sensing techniques and enhancement of temporal, spatial, and radiometric resolution of satellite images have produced a range of useful products. These products can be widely used in various research areas, especially water and environment. Many studies have been conducted in estimating ET_O as an important element of hydrological cycle. In this regard, it can be mentioned in the studies of Guerschman et al. [14]; Maeda et al. [4]; Yang et al. [12]; Cleugh et al. [15]; Tian et al. [16]; Hankerson et al. [17]; Li et al. [18].

Recently, some data-based techniques such as artificial neural networks (ANNs), support vector machine, and M5 model tree have been broadly utilized in solving practical problems. Mentioned techniques can simulate processes in which there is not an exact solution, and it is merely possible to approach near-optimal solution [19]. In these methods, a new knowledge is not produced about the process of a problem. In fact, thanks to the proper recognition of the problem process in selecting the input and output variables, caused by the use of modern regression techniques, this helps to better fit the data [20]. Data-driven techniques can efficiently model nonlinear and complicated processes such as ET that its precise calculation depends on numerous weather data and interactions between them [21].

Recently, by the developing new machine learning methods, application of new models in the estimation of ET_O has been increased. Kişi and Öztürk [22] studied the adaptive neurofuzzy inference system (ANFIS) technique in estimation of ET_O in Los Angeles, California, and showed that it could be successfully considered in the ET_O estimation. Kişi and Çimen [23] applied SVM in modeling ET_O in central California and concluded that SVR model could be embedded as a module for estimating ET_O data in hydrological modeling studies. Tabari et al. [24] investigated the potential of SVM, ANFIS, multiple linear regression (MLR), and multiple nonlinear regression (MNL) for estimating in a semiarid highland environment in Iran. They found that the SVM and ANFIS had better performance in ET_O modeling compared to the others. Besides, several studies have shown high performance of ANNs in modeling ET against statistical methods [8, 25–29]. The substantial benefit of ANN methods, compared to conventional methods, is the ability of solving problems that are difficult to formalize [30]. Hou et al. [10] studied ANNs in simulating ET_O values and demonstrated that the models reflected a noticeable efficiency. Landaras

et al. [21] compared calculated ET_O by ANN to the ET_O values of radiation and thermal models and declared that ANN can provide more accurate results than empirical models.

Due to the simple geometric structure of model tree and providing fast computing set of IF-THEN rules, which are easily understandable [31], the model tree has been utilized in different scientific studies. Pal and Deswal [1] in a study evaluated the capability of M5 model tree to model ET_O in daily scale, and the obtained results showed that M5 model tree can be properly used in this context. Rahimikhoob et al. [11] conducted a study to assess the performance of M5 model tree in reaching ET_O values in a semiarid region of Iran by means of K_p values of pan evaporation. They asserted that this model shows more satisfactory performance than other approaches in estimating ET_O in the study region. Sattari et al. [32] compared ANN and M5 model tree in determining ET_O and concluded that both ANN model and M5 model tree can properly estimate ET_O . ANN suggested a better performance; however, M5 model tree offered simple linear regressions that can simply calculate ET_O .

The main aim of this study is to evaluate the ability of M5 model tree and ANN model in estimating daily ET_O using LST values obtained from MODerate Resolution Imaging Spectroradiometer (MODIS)/Terra sensor. The accuracy of calculated ET_O by means of M5 model tree and ANN model has been investigated by comparing with the ET_O amounts of FAO-PM method.

2. Materials and Methods

2.1. Study Area and Data. Farms of Sugarcane Development Project, which are located in Khuzestan province, Iran, were considered as the study area in the present research. Khuzestan province is stretched in southeastern part of the country, with the area of approximately 63,238 km² between longitudes of 47°38' E to 50°32' E and latitudes of 29°57' N to 33°00' N. This province is located adjoining the boundary of Iraq from the west and Persian Gulf from the south. Khuzestan province according to de Martonne climate classification has a semiarid climate. The average temperature during the winter season is nearly 9.14°C, and occasionally its minimum falls down a few degrees below 0°C. The average temperature in the summers is approximately 31.2°C, whereas its maximum sometimes exceeds 50°C. Required climatic variables in order to estimate ET_O in this study were collected from weather stations, namely, Amirkabir, Farabi, Gazali, Khazaei, and Shoeybie for the years of 2006 and 2007 in the periods when satellite images existed. The meteorological data consisted of daily observations of maximum and minimum temperature, sunshine duration, relative humidity, and wind speed which have been used in estimating ET_O by FAO-PM equation. Figure 1 illustrates the location of weather stations in the study area.

2.2. MODIS Images. In the current study, data of daytime LST (LST_D) and nighttime LST (LST_N) obtained from the MODIS sensor were used as inputs to the M5 model tree and ANN models. The MODIS sensor, which is on board

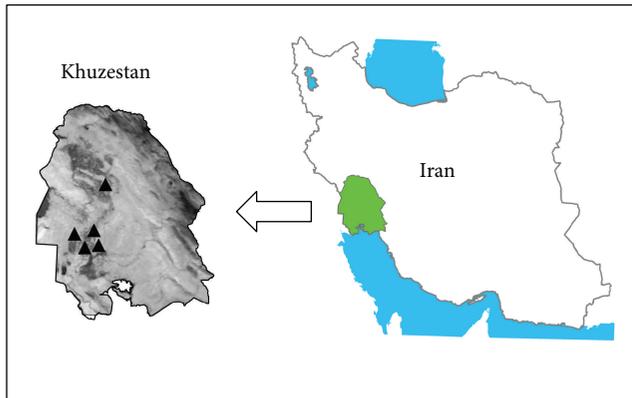


FIGURE 1: Location of weather stations in the study area.

of the Terra and Aqua satellites, is able to provide images in 36 spectral bands between 0.62 and $14.385 \mu\text{m}$. Spatial resolutions of the sensor are 250 m , 500 m , and 1 km in several bands. In the thermal-infrared (TIR), MODIS sensors have 1 km spatial resolution [19]. This sensor performs a range of daily regular observations on the sea, land, and atmosphere in the global scale; moreover, it has an extensive and continuous spectral and spatial coverage. MODIS sensor was launched by the Terra and Aqua satellites in 1999 and 2002, respectively. These satellites both have a sun-synchronous, near polar, circular orbit; Terra satellite everyday crosses over the equator from north to south, while Aqua satellite has an inverse movement from south to north. United States National Aeronautics and Space Administration (NASA) is responsible for processing and providing MODIS sensor's data. NASA supplies the data in a large variety of products applicable to land, ocean, and atmosphere uses [4]. The data and products of MODIS sensor are available on the NASA's website (<http://modis.gsfc.nasa.gov/>). The images selected for this study are the level three (L3) data of the MODIS sensor (MOD11A1), which are related to Terra satellite. This product provides daytime and nighttime LST data collected on a 1 km spatial resolution and gridded in the sinusoidal projection system [33]. The MODIS sensor LST data is acquired from two TIR channels, that is, 31 ($10.78\text{--}11.28 \mu\text{m}$) and 32 ($11.77\text{--}12.27 \mu\text{m}$) using the split-window algorithm [34].

A total of 428 images of the MODIS sensor associated with MOD11A1 production from the Terra satellite in the time periods of 2006 and 2007 covering the study area were retrieved from the Land Processes Distributed Active Archive Center (LP DAAC) (<http://lpdaac.usgs.gov>). Cloudless sky and fair weather were the reason for selecting these pictures. Changing coordinate systems from the sinusoidal to UTM (datum WGS84) was performed using MODIS Reprojection Tool (MRT). Afterward, locations of study stations were projected on the images and then LST_D and LST_N were extracted from the corresponding pixels of the images. Extracting LST values in the desired stations was done by means of Hawth's Analysis Tools (HAT) in ArcGIS 9.3 software. Unit of LST values extracted from the images was Kelvin, so before

utilizing them as model input, these values were converted to Celsius.

In the present study, whole LST values extracted from MODIS images in three stations, including Amirkabir, Farabi, and Gazali, were used to construct and train the ANN and M5 model tree, and two other stations, Khazaei and Shoeybie, were applied to test mentioned models.

2.3. Artificial Neural Network (ANN). ANNs are known as powerful machine-learning techniques, which are extensively used for numerical prediction and classification [35]. ANNs are also operative tools to model nonlinear systems and need less amount of data as inputs than customary mathematical approaches [27]. A neural network is based on processing a set of data that its analysis procedure has been inspired by biological neural systems. Neural networks consist of an interconnected group of neurons placed near to each other in layers that implement data processing through neurons and their connections. ANNs included one input layer, an output layer, and one or more hidden layers; neurons of each layer are joined to all neurons of the previous layer by means of weighted connections. The input vectors are received by neurons of the input layer and the values can be conveyed to the next layer of processing elements across connections. The number of neurons in each hidden layer is determined by the user. This process continues up to the output layer. ANN modeling is carried out in two stages including training and testing. In the training stage, after getting input data, neural network makes attempts to convert the inputs into desired outputs. Connecting weights of network neurons are determined in this stage; these weights in the testing stage are examined using various datasets.

A multilayer perceptron (MLP) ANN with one input layer, one hidden layer, and one output layer was used to estimate ET_O from LST_D , LST_N , mean LST (LST_M), and day of year (DOY) data as inputs. The transfer function in the networks was log-sigmoid for this study. The accuracy of the networks was evaluated for each epoch in the training through mean squared error (MSE). The backpropagation (BP) algorithm was employed to train MLP neural network. Levenberg-Marquardt (LM), a second-order nonlinear optimization technique, was used with an early stopping criterion to improve the network training speed and efficiency. The target error for the training of networks was equal to 10^{-4} . With the specification of this target error, the number of iteration for all ANN models was placed to 1000.

2.4. M5 Model Tree. M5 model tree was introduced by Quinlan in 1992 [36]. The model is established according to a binary decision tree in which there are linear regression functions in the terminal node (leaf), which sets a relationship between independent and dependent variables [11]. Tree-based models are made according to a divide-and-conquer method for constructing a relationship between independent and dependent variables [37]. The tree models can be also used for qualitative and quantitative data [9].

Building the M5 model tree involves two separate steps [36, 38]. The first one consists of dividing data into subsets

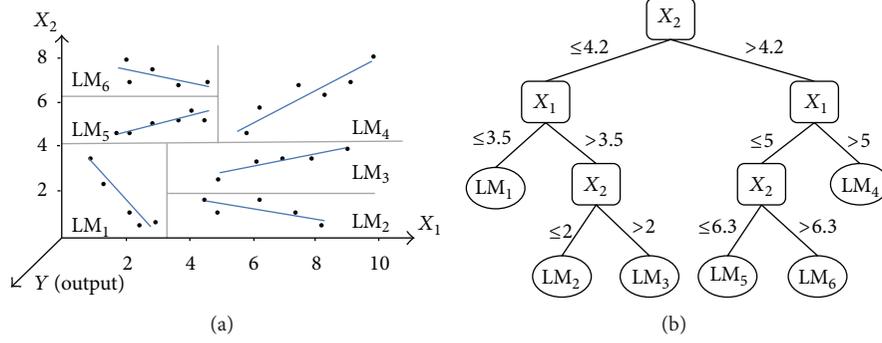


FIGURE 2: (a) Example of M5 model tree, a splitting the input space $x_1 \times x_2$ by M5 model tree algorithm and (b) diagram of model tree with six linear regression models at the leaves [11].

to construct the model tree. The dividing criterion is based on the standard deviation of the subset values that reach a node as a measure of error value in that node and additionally calculating the expected reduction in the error as a result of testing each attribute at that node. Equation of standard deviation reduction (SDR) is given as follows [1]:

$$\text{SDR} = \text{SD}(T) - \sum \frac{T_i}{T} \times \text{SD}(T_i), \quad (1)$$

where T is defined as a set of examples that reaches the node; T_i demonstrates the subset of examples that have the i th output of the potential set, and SD is the standard deviation [39]. Due to branching process, data in child nodes (subtree or smaller nodes) have fewer SD than parent nodes (greater nodes). After examining all possible structures, a structure would be picked out that has the maximum expected error reduction. This dividing often creates a great tree-like structure that leads to overfit structure. In order to avoid overfitting, in the second step the overgrown tree is pruned, and pruned subtrees are replaced with linear regression functions. Figure 2(a) illustrates splitting the input space $x_1 \times x_2$ (independent variables) into six linear regression functions at the leaves (labeled LM1 through LM6) by M5 model tree algorithm. General form of the model is $y = a_0 + a_1x_1 + a_2x_2$, in which a_0 , a_1 , and a_2 are linear regression constants. Figure 2(b) shows branches relations in the form of a tree diagram [19].

2.5. FAO-PM Equation. FAO-PM equation, a physically based method, is widely used in the calculation of ET_O and is a reliable method in the estimation of ET_O in various climate conditions [2, 40, 41]. Moreover, FAO proposed this equation as a standard method for estimating ET_O . Although the most precise method for estimating ET_O is to employ lysimeter data, considering the absence of such data in the study region, assessment of ANN and M5 model tree performance was made by FAO-PM equation. The equation for estimating daily ET_O has been presented as follows [2]:

$$\text{ET}_O = \frac{0.484\Delta(R_n - G) + \gamma(900/(T_a + 273))U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)}, \quad (2)$$

where ET_O is the reference evapotranspiration (mm d^{-1}), Δ is the slope of the saturation vapor pressure curve ($\text{kPa } ^\circ\text{C}^{-1}$) at the daily mean air temperature ($^\circ\text{C}$), R_n and G are the net solar radiation and soil heat flux density ($\text{MJ m}^{-2} \text{d}^{-1}$), γ is the psychrometric constant ($\text{kPa } ^\circ\text{C}^{-1}$), T is the daily mean temperature ($^\circ\text{C}$), U_2 is the mean wind speed (ms^{-1}), e_s is the saturation vapor pressure (kPa), and e_a is the actual vapor pressure (kPa) [2].

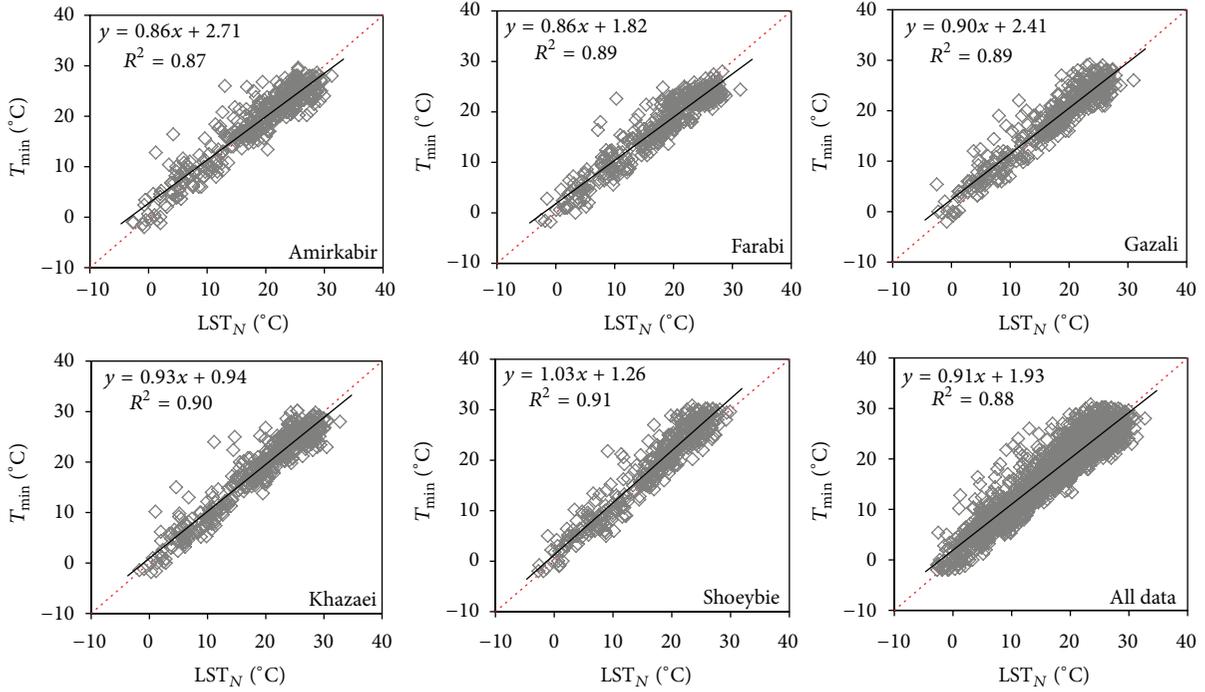
In current study, the daily values of Δ , R_n , e_s , and e_a were calculated using the equation given by Allen et al. [2]. For calculation of R_n , it is needed to calculate daily solar radiation (R_s). Allen et al. [2] proposed R_s to be calculated using the Angstrom formula, which relates solar radiation to extraterrestrial radiation (R_a) and relative sunshine duration. The formula has the following form:

$$R_s = \left(a + b \frac{n}{N}\right) R_a, \quad (3)$$

where n and N are, respectively, the actual daily sunshine duration and daily maximum possible sunshine duration, R_s and R_a are, respectively, the daily global solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$) and the daily extraterrestrial solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$) on a horizontal surface, and a and b are empirical coefficients which depend on location. The values of a and b coefficients are considered as default values, 0.25 and 0.5, respectively, proposed by Allen et al. [2].

2.6. Statistical Analyses. In this study, ET_O results obtained from FAO-PM equation were utilized as reference values and the results achieved by ANN and M5 model tree were taken into account as estimated values. The comparison of ANN model and M5 model tree values with FAO-PM equation was made using statistical indices such as coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), in addition to plotting graphs. These indices are defined as follows:

$$R^2 = \left(\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \right)^2$$

FIGURE 3: Scatter plot of T_{\min} against LST_N .

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i|,$$

(4)

where x_i is ET_O computed by FAO-PM, \bar{x} denotes the average amount of computed ET_O computed by FAO-PM, y_i is ET_O computed by ANN/M5, and \bar{y} is the average computed ET_O amounts by last models.

3. Results and Discussion

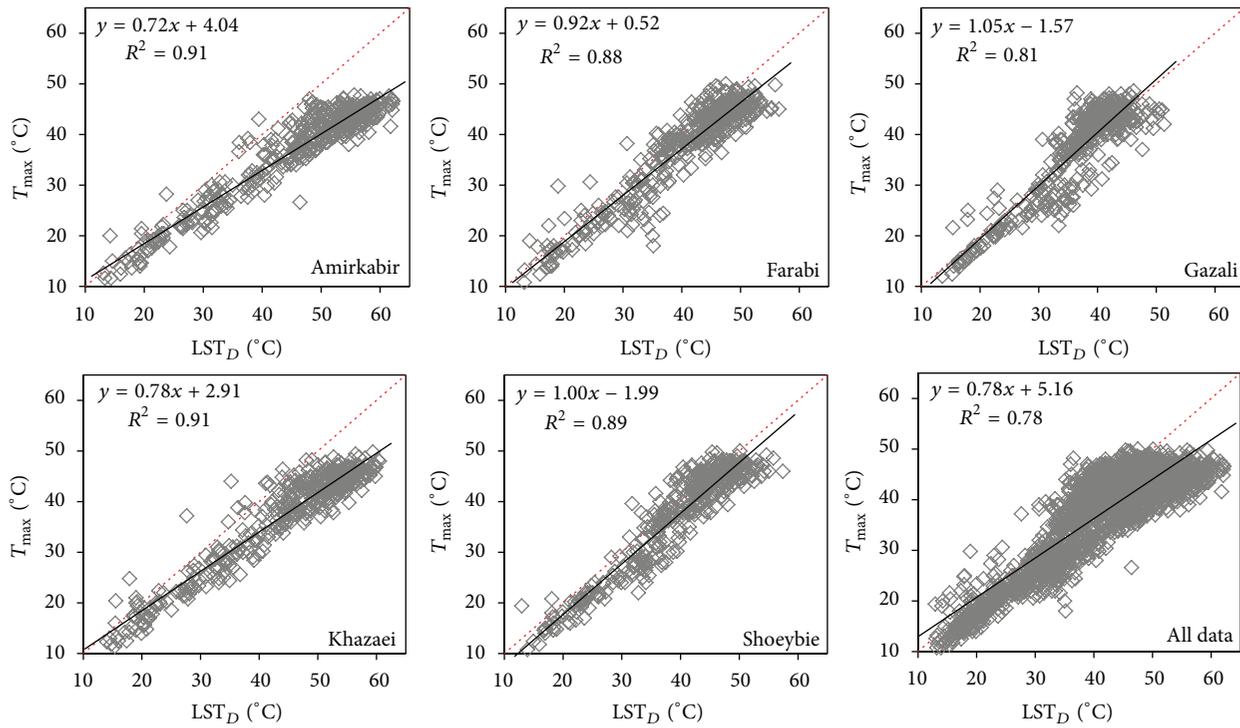
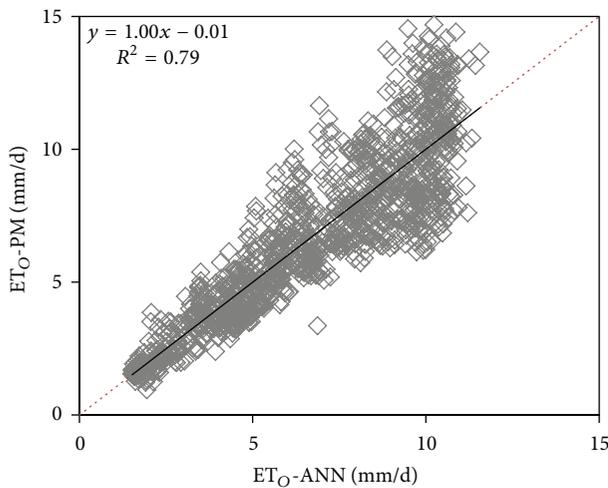
In the first stage of the study, before calculating the ET_O , the correlation between LST obtained from MODIS sensor and air temperature was determined. Accordingly, scatter plots of LST values provided by MODIS images versus air temperature are presented for different stations in Figures 3 and 4. As it can be perceived in Figure 3, there is a high correlation between the LST_N and minimum air temperature (T_{\min}). The greatest R^2 was observed between LST_N and T_{\min} in Shoeybie station, which was equal to 0.91; the minimum R^2 was seen in the Amirkabir station with the value of 0.87. Furthermore, for all stations, R^2 of two aforementioned parameters was entirely assessed at 0.88. In Figure 3, symmetrical and appropriate distribution of points around the line of best fit (line 1:1) indicates a strong correlation between the two parameters. In Figure 4, relation between the LST_D and maximum air temperature (T_{\max}) displays that the greatest R^2 between LST_D and T_{\max} is related to Amirkabir and Khazaei stations

($R^2 = 0.91$); the minimum amount of this coefficient is associated with Gazali station ($R^2 = 0.81$). For all study stations, R^2 was generally equal to 0.78. Overall, it can be concluded that there is a strong relationship between these two parameters. The scatter of points around the line of best fit in the Figure 4 depicts an overestimation of LST_D over T_{\max} . This overestimation is more pronounced for the temperatures over 30°C .

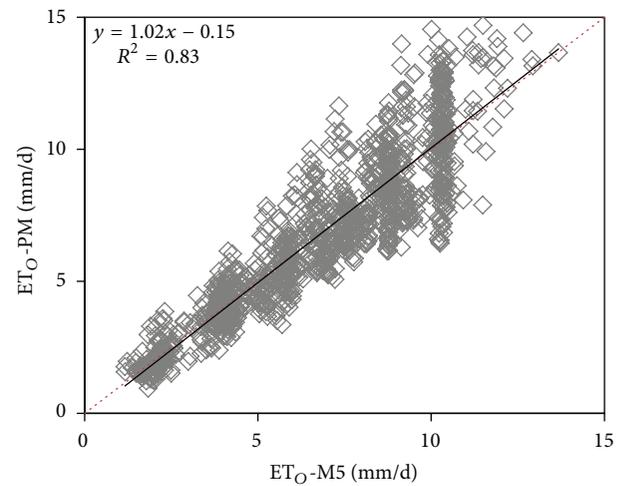
For all examined datasets, the correlation between LST_N and T_{\min} is higher compared to the correlation between LST_D and T_{\max} for thermal products of MODIS sensor. Moreover, previous studies concerning estimation of air temperature have similarly confirmed that correlation between LST_N and T_{\min} is greater than correlation between LST_D and T_{\max} [12, 19].

In the second stage, in order to calculate ET_O using LST values extracted from the MODIS sensor (LST_D and LST_N), ANN models and M5 model tree were employed. In addition to LST_D and LST_N , two parameters including mean LST (LST_M) and day of year (DOY) were considered as inputs for generating model in order to increase accuracy of ET_O estimates. Thus, in producing ANN model and M5 model tree, four parameters including LST_D , LST_N , DOY, and LST_M were taken into consideration as input and ET_O calculated from weather data using FAO-PM equation as output.

For generating ANN model, training data (whole data of Amirkabir, Farabi, and Gazali stations) were applied, and the number of nodes in the hidden layer was obtained using trial and error method. The number of nodes in the hidden layer was considered varying from 1 to 12; based on error statistical parameters, a network with 3 nodes in the hidden layer indicated the lowest error in comparison of ET_O values

FIGURE 4: Scatter plot of T_{\max} against LST_D .FIGURE 5: Scatter plot of estimated ET_O using ANN against calculated ET_O by FAO-PM equation for training data.

calculated by FAO-PM method. Table 1 presents the values of error parameters for the neural network with the structure of three nodes in the hidden layer. In Figure 5, scatter diagram of calculated ET_O values through ANN for all training data has been plotted against ET_O values obtained by FAO-PM equation. As shown in Figure 3, $R^2 = 0.79$ demonstrates that the ANN has provided an acceptable determination coefficient in approximating ET_O . In this figure, the points density increases around the line 1 : 1 in large amounts of ET_O ; that is, it represents the overestimation or underestimation of the data that are situated in this area.

FIGURE 6: Scatter plot of estimated ET_O using M5 model tree against calculated ET_O by FAO-PM equation for training data.TABLE 1: Error statistics in estimating ET_O for model tree and ANN methods using training data.

Model	RMSE (mm/d)	MAE (mm/d)
ANN	1.40	1.05
M5	1.27	0.96

For generating M5 model tree, WEKA 3.7 software [42] was used considering the same data applied to the training ANN model. Algorithm 1 shows M5 model tree produced

LSTn <= 19.43:	LM num: 1
LSTms <= 15.55: LM1	Pm = 0.0147 * Tn + 0.1048 * Tms - 0.0007 * DOY + 0.8963
LSTms > 15.55:	LM num: 2
DOY <= 282.5:	Pm = 0.0349 * Tn + 0.0312 * Tms + 0.028 * DOY + 1.3107
DOY <= 81.5: LM2	LM num: 3
DOY > 81.5:	Pm = 0.046 * Tn + 0.081 * Tms - 0.0134 * DOY + 3.9221
DOY <= 106:	LM num: 4
LSTd <= 31.2: LM3	Pm = 0.0719 * Tn + 0.0642 * Tms - 0.0088 * DOY + 3.8864
LSTd > 31.2: LM4	LM num: 5
DOY > 106:	Pm = 0.0705 * Tn + 0.0073 * Td + 0.0301 * Tms + 0.0003 * DOY + 4.6615
DOY <= 207:	LM num: 6
LSTn <= 17.72:	Pm = 0.201 * Tn + 0.0147 * Td + 0.0301 * Tms + 0.0003 * DOY + 2.5252
DOY <= 129: LM5	LM num: 7
DOY > 129: LM6	Pm = 0.118 * Tn + 0.0301 * Tms + 0.0026 * DOY + 4.7479
LSTn > 17.72: LM7	LM num: 8
DOY > 207:	Pm = 0.0529 * Tn + 0.0301 * Tms - 0.0087 * DOY + 6.254
DOY <= 272.5: LM8	LM num: 9
DOY > 272.5: LM9	Pm = 0.0529 * Tn - 0.0348 * Tms - 0.0116 * DOY + 8.4773
DOY > 282.5:	LM num: 10
DOY <= 311.5: LM10	Pm = 0.0328 * Tn + 0.0201 * Tms - 0.0012 * DOY + 3.367
DOY > 311.5: LM11	LM num: 11
LSTn > 19.43:	Pm = 0.1437 * Tn + 0.0201 * Tms - 0.0012 * DOY + 1.7414
DOY <= 227.5:	LM num: 12
DOY <= 159.5:	Pm = -0.1076 * Tn + 0.0018 * Tms - 0.0393 * DOY + 15.9787
DOY <= 147.5: LM12	LM num: 13
DOY > 147.5: LM13	Pm = 0.0297 * Tn + 0.0018 * Tms + 0.0032 * DOY + 8.0036
DOY > 159.5:	LM num: 14
DOY <= 198.5:	Pm = 0.3445 * Tn + 0.0769 * Td + 0.0018 * Tms - 0.1664 * DOY + 26.3097
DOY <= 167.5: LM14	LM num: 15
DOY > 167.5: LM15	Pm = 0.0316 * Tn + 0.0018 * Tms - 0.006 * DOY + 10.5743
DOY > 198.5:	LM num: 16
DOY <= 216.5:	Pm = -0.0175 * Tn + 0.0018 * Tms + 0.0225 * DOY + 4.5799
DOY <= 206.5: LM16	LM num: 17
DOY > 206.5: LM17	Pm = -0.206 * Tn + 0.0018 * Tms + 0.0185 * DOY + 11.1645
DOY > 216.5: LM18	LM num: 18
DOY > 227.5: LM19	Pm = 0.0015 * Tn + 0.0018 * Tms - 0.0046 * DOY + 9.6864
	LM num: 19
	Pm = 0.022 * Tn + 0.0018 * Tms - 0.0387 * DOY + 16.0569

ALGORITHM 1: Tree-structured regression model generated by M5 algorithm using training data.

by the software and established linear regression relations. Afterward, a scatter plot of estimated ET_O by M5 model tree versus calculated ET_O by FAO-PM equation has been presented for training data (Figure 6); R^2 value equal to 0.83 indicates proper agreement. The distribution of points around the best fit line represents higher accuracy of this method in the estimation of low ET_O values; in the great values of ET_O the error amount increases as well.

The comparison of error statistics concerning developed models by ANN model and M5 model tree in estimating ET_O using training data can be drawn in Table 1. As seen, M5 model tree, which has shown RMSE and MAE values, respectively, equal to 1.27 and 0.96 $mm\ d^{-1}$, has resulted in higher accuracy than ANN in estimating ET_O .

For testing the methods of ANN and M5 model tree, data of Khazaei and Shoeybie stations have been used. In Figure 7, the scatter plot of calculated ET_O through two models of

ANN and M5 model tree versus FAO-PM values for the mentioned stations have been shown. R^2 values associated with calculated ET_O by ANN against FAO-PM equation ET_O values for Khazaei and Shoeybie stations were 0.79 and 0.86, respectively. Additionally, the same R^2 values related to M5 model tree for aforementioned stations were 0.80 and 0.85, respectively. In both methods, ET_O values were properly distributed around line 1:1. For smaller amounts of ET_O , there was been a better compatibility between the calculated values by two models and FAO-PM values compared to greater amounts of ET_O . In other words, underestimation or overestimation of ET_O is related to warm seasons further than cold seasons. The evolutions of calculated ET_O values over time have been illustrated for two test stations in Figure 8. As perceived in this graph, the maximum amounts of ET_O by these two models were not entirely compatible with FAO-PM values. This could be due to the lack of consideration of some

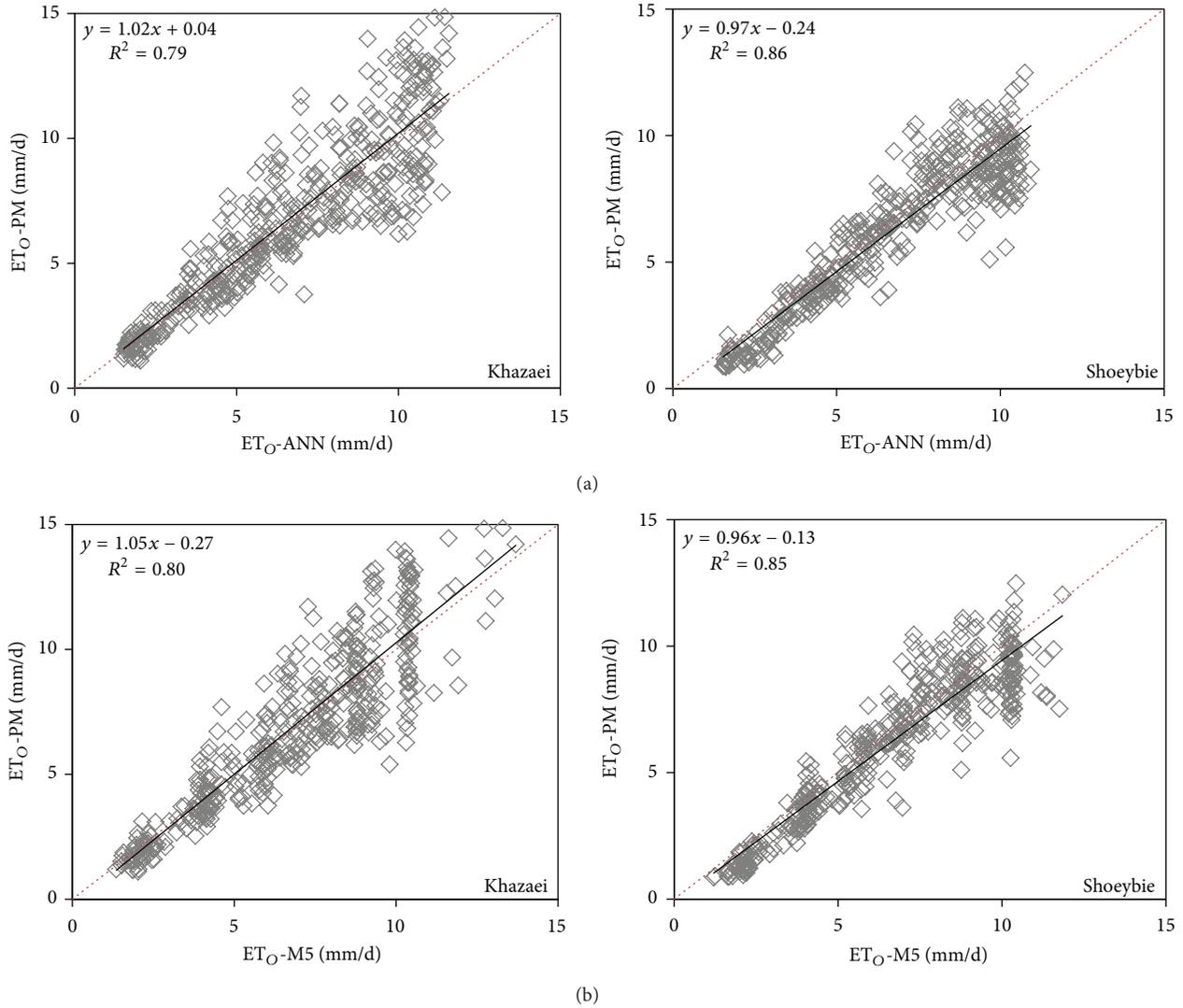


FIGURE 7: Scatter plot of estimated ET_O using ANN and M5 model tree against calculated ET_O by FAO-PM method, (a) ANN method, and (b) M5 method.

weather parameters affecting ET_O rate (e.g., wind speed, solar radiation, and relative humidity) as inputs of ANN model and M5 model tree.

In Table 2, error statistics of ET_O estimation by ANN and M5 model tree have been presented for two test stations. According to the statistics of RMSE and MAE, both methods performed equally well. Similar results have been reported about close performance of ANN and M5 models in rainfall-runoff modeling [43] and river flow forecasting [44]. The main advantage of M5 model tree is that it provides the results in a simple and comprehensible form of regression equations [1], which can be easily used in the calculation of ET_O .

It seems generally that LST_N and LST_D values obtained from MODIS sensor, making the use of both ANN and M5 models, can be properly utilized in the estimation of ET_O .

4. Conclusion

In this study, attempts were made to calculate ET_O using LST values extracted from images of MODIS sensor. For this

TABLE 2: Error statistics in estimating ET_O by ANN and M5 model tree using testing data.

Station	Model	R^2	RMSE (mm/d)	MAE (mm/d)
Khazaei	ANN	0.79	1.54	1.14
	M5	0.80	1.49	1.12
Shoeybie	ANN	0.86	1.15	0.90
	M5	0.85	1.17	0.90

purpose, two methods of ANN and M5 model tree were selected in order to estimate ET_O ; four parameters including LST_D , LST_N , LST_M , and DOY were considered as model inputs, and ET_O was calculated from FAO-PM equation as model outputs. A comparison of estimated ET_O using each mentioned model with the values obtained from the FAO-PM method revealed high accuracy of both methods. The average amounts of RSME in two test stations for ANN and M5 model tree were, respectively, 1.35 and 1.34 mm/day. According to the error values, it can be observed that both

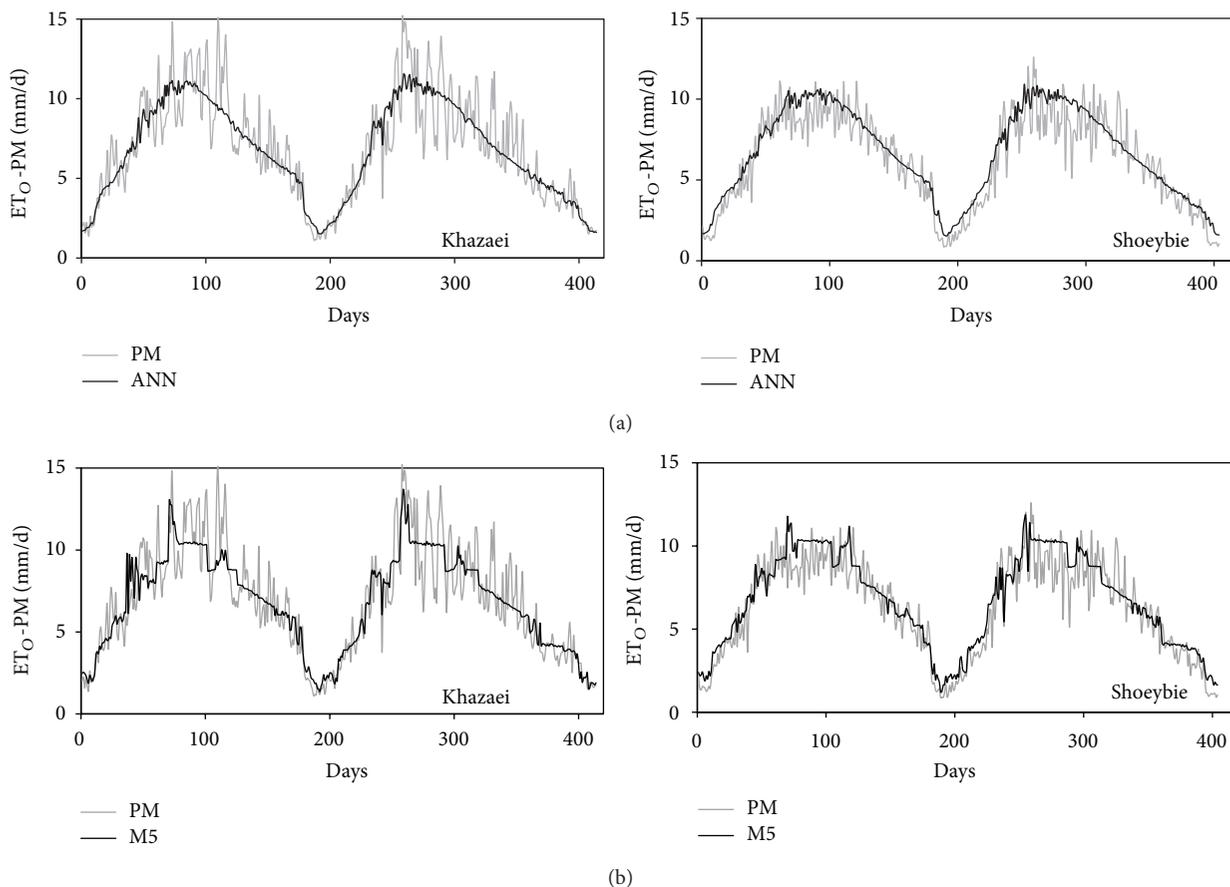


FIGURE 8: Time variations of estimated ET_O using ANN and M5 model tree for test stations, (a) ANN method, and (b) M5 method.

models have shown close performance in determining ET_O . It seems that considering other significant variables affecting the ET process such as relative humidity and wind speed (as model inputs) can lead to higher accuracy.

Overall, it can be concluded that, unlike complex ANN method, M5 model tree due to its simple structure and providing simple regression equations, is a more appropriate method in determining ET_O using LST data derived from MODIS sensor.

In this paper A M5 tree model with some simple linear regressions was developed to estimate ET_O using MODIS LST data. Considering simple linear equations and minor data required in the developed M5 tree model, it is expected that this method could be an alternative to calculate ET_O in the study area. However, this model provided acceptable results in the current area but it seems that its applicability in the other region of the world or the region with similar climate condition needs more investigation.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] M. Pal and S. Deswal, "M5 model tree based modelling of reference evapotranspiration," *Hydrological Processes*, vol. 23, no. 10, pp. 1437–1443, 2009.
- [2] R. G. Allen, L. S. Pereira, D. Raes, and M. Smith, "Crop evapotranspiration—guidelines for computing crop water requirements," FAO Irrigation and Drainage Paper 56, FAO, Rome, Italy, 1998.
- [3] S. Ortega-Farias, S. Irmak, and R. H. Cuenca, "Special issue on evapotranspiration measurement and modeling," *Irrigation Science*, vol. 28, no. 1, pp. 1–3, 2009.
- [4] E. E. Maeda, D. A. Wiberg, and P. K. E. Pellikka, "Estimating reference evapotranspiration using remote sensing and empirical models in a region with limited ground data availability in Kenya," *Applied Geography*, vol. 31, no. 1, pp. 251–258, 2011.
- [5] A. A. Sabziparvar, H. Tabari, A. Aeini, and M. Ghafouri, "Evaluation of class a pan coefficient models for estimation of reference crop evapotranspiration in cold semi-arid and warm arid climates," *Water Resources Management*, vol. 24, no. 5, pp. 909–920, 2010.
- [6] M. E. Jensen, R. D. Burman, and R. G. Allen, "Evapotranspiration and irrigation water requirements," ASCEM Annuals and Reports on Engineering Practice no. 70, ASCE, New York, NY, USA, 1990.

- [7] M. Smith, *Report on the Expert Consultation on Revision of FAO Methodologies for Crop Water Requirements*, FAO, Rome, Italy, 1992.
- [8] M. Kumar, N. S. Raghuwanshi, R. Singh, W. W. Wallender, and W. O. Pruitt, "Estimating evapotranspiration using artificial neural network," *Journal of Irrigation and Drainage Engineering*, vol. 128, no. 4, pp. 224–233, 2002.
- [9] P. Dittthakit and C. Chinnarasri, "Estimation of pan coefficient using M5 model tree," *American Journal of Environmental Sciences*, vol. 8, no. 2, pp. 95–103, 2012.
- [10] Z. Huo, S. Feng, S. Kang, and X. Dai, "Artificial neural network models for reference evapotranspiration in an arid area of northwest China," *Journal of Arid Environments*, vol. 82, pp. 81–90, 2012.
- [11] A. Rahimikhoob, M. Asadi, and M. Mashal, "A comparison between conventional and M5 model tree methods for converting pan evaporation to reference evapotranspiration for semi-arid region," *Water Resources Management*, vol. 27, no. 14, pp. 4815–4826, 2013.
- [12] Y. Yang, S. Shang, and L. Jiang, "Remote sensing temporal and spatial patterns of evapotranspiration and the responses to water management in a large irrigation district of North China," *Agricultural and Forest Meteorology*, vol. 164, pp. 112–122, 2012.
- [13] Z. Wan, "New refinements and validation of the MODIS land-surface temperature/emissivity products," *Remote Sensing of Environment*, vol. 112, no. 1, pp. 59–74, 2008.
- [14] J. P. Guerschman, A. I. J. M. van Dijk, G. Mattersdorf et al., "Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia," *Journal of Hydrology*, vol. 369, no. 1-2, pp. 107–119, 2009.
- [15] H. A. Cleugh, R. Leuning, Q. Mu, and S. W. Running, "Regional evaporation estimates from flux tower and MODIS satellite data," *Remote Sensing of Environment*, vol. 106, no. 3, pp. 285–304, 2007.
- [16] H. Tian, J. Wen, C.-H. Wang, R. Liu, and D.-R. Lu, "Effect of pixel scale on evapotranspiration estimation by remote sensing over oasis areas in north-western China," *Environmental Earth Sciences*, vol. 67, no. 8, pp. 2301–2313, 2012.
- [17] B. Hankerson, J. Kjaersgaard, and C. Hay, "Estimation of evapotranspiration from fields with and without cover crops using remote sensing and in situ methods," *Remote Sensing*, vol. 4, no. 12, pp. 3796–3812, 2012.
- [18] X. Li, L. Lua, W. Yangc, and G. Chenga, "Estimation of evapotranspiration in an arid region by remote sensing—a case study in the middle reaches of the Heihe River Basin," *International Journal of Applied Earth Observation and Geoinformation*, vol. 17, no. 1, pp. 85–93, 2012.
- [19] S. Emamifar, A. Rahimikhoob, and A. A. Noroozi, "Daily mean air temperature estimation from MODIS land surface temperature products based on M5 model tree," *International Journal of Climatology*, vol. 33, no. 15, pp. 3174–3181, 2013.
- [20] B. Bhattacharya and D. P. Solomatine, "Neural networks and M5 model trees in modelling water level-discharge relationship," *Neurocomputing*, vol. 63, pp. 381–396, 2005.
- [21] G. Landeras, A. Ortiz-Barredo, and J. J. López, "Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain)," *Agricultural Water Management*, vol. 95, no. 5, pp. 553–565, 2008.
- [22] Ö. Kişi and Ö. Öztürk, "Adaptive neurofuzzy computing technique for evapotranspiration estimation," *Journal of Irrigation and Drainage Engineering*, vol. 133, no. 4, pp. 368–379, 2007.
- [23] O. Kişi and M. Çimen, "Evapotranspiration modelling using support vector machines," *Hydrological Sciences Journal*, vol. 54, no. 5, pp. 918–928, 2009.
- [24] H. Tabari, O. Kisi, A. Ezani, and P. H. Talae, "SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment," *Journal of Hydrology*, vol. 444–445, pp. 78–89, 2012.
- [25] P. B. Shirsath and A. K. Singh, "A comparative study of daily pan evaporation estimation using ANN, regression and climate based models," *Water Resources Management*, vol. 24, no. 8, pp. 1571–1581, 2010.
- [26] S. S. Zanetti, E. F. Sousa, V. P. S. Oliveira, F. T. Almeida, and S. Bernardo, "Estimating evapotranspiration using artificial neural network and minimum climatological data," *Journal of Irrigation and Drainage Engineering*, vol. 133, no. 2, pp. 83–89, 2007.
- [27] A. Rahimikhoob, "Estimating sunshine duration from other climatic data by artificial neural network for ET_0 estimation in an arid environment," *Theoretical and Applied Climatology*, vol. 118, no. 1-2, pp. 1–8, 2014.
- [28] I. El-Baroudy, A. Elshorbagy, S. K. Carey, O. Giustolisi, and D. Savic, "Comparison of three data-driven techniques in modelling the evapotranspiration process," *Journal of Hydroinformatics*, vol. 12, no. 4, pp. 365–379, 2010.
- [29] M.-B. Aghajanloo, A.-A. Sabziparvar, and P. Hosseinzadeh Talae, "Artificial neural network-genetic algorithm for estimation of crop evapotranspiration in a semi-arid region of Iran," *Neural Computing and Applications*, vol. 23, no. 5, pp. 1387–1393, 2013.
- [30] K. P. Sudheer, A. K. Gosain, and K. S. Ramasastri, "Estimating actual evapotranspiration from limited climatic data using neural computing technique," *Journal of Irrigation and Drainage Engineering*, vol. 129, no. 3, pp. 214–218, 2003.
- [31] V. Jothiprakash and A. S. Kote, "Effect of pruning and smoothing while using m5 model tree technique for reservoir inflow prediction," *Journal of Hydrologic Engineering*, vol. 16, no. 7, pp. 563–574, 2011.
- [32] M. T. Sattari, M. Pal, K. Yürekli, and A. Ünlükara, "M5 model trees and neural network based modelling of ET_0 in Ankara, Turkey," *Turkish Journal of Engineering & Environmental Sciences*, vol. 37, no. 2, pp. 211–219, 2013.
- [33] K. Wang, Z. Wan, P. Wang et al., "Estimation of surface long wave radiation and broadband emissivity using moderate resolution imaging spectroradiometer (MODIS) land surface temperature/emissivity products," *Journal of Geophysical Research D: Atmospheres*, vol. 110, no. 11, pp. 1–12, 2005.
- [34] C. Vancutsem, P. Ceccato, T. Dinku, and S. J. Connor, "Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa," *Remote Sensing of Environment*, vol. 114, no. 2, pp. 449–465, 2010.
- [35] K. K. Singh, M. Pal, and V. P. Singh, "Estimation of mean annual flood in indian catchments using backpropagation neural network and M5 model tree," *Water Resources Management*, vol. 24, no. 10, pp. 2007–2019, 2010.
- [36] J. R. Quinlan, "Learning with continuous classes," in *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence (AI '92)*, pp. 343–348, World Scientific, Singapore, 1992.

- [37] M. Pal, "M5 model tree for land cover classification," *International Journal of Remote Sensing*, vol. 27, no. 4, pp. 825–831, 2006.
- [38] D. P. Solomatine and Y. Xue, "M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China," *Journal of Hydrologic Engineering*, vol. 9, no. 6, pp. 491–501, 2004.
- [39] Y. Wang and I. H. Witten, "Induction of model trees for predicting continuous classes," in *Proceedings of the Poster Papers of the European Conference on Machine Learning*, University of Economics, Faculty of Informatics and Statistics, Prague, Czech Republic, 1997.
- [40] D. Itenfisu, R. L. Elliott, R. G. Allen, and I. A. Walter, "Comparison of reference evapotranspiration calculations as part of the ASCE standardization effort," *Journal of Irrigation and Drainage Engineering*, vol. 129, no. 6, pp. 440–448, 2003.
- [41] P. Gavilán, I. J. Lorite, S. Tornero, and J. Berengena, "Regional calibration of Hargreaves equation for estimating reference et in a semiarid environment," *Agricultural Water Management*, vol. 81, no. 3, pp. 257–281, 2006.
- [42] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, San Francisco, Calif, USA, 2005.
- [43] D. P. Solomatine and K. N. Dulal, "Model trees as an alternative to neural networks in rainfall—runoff modelling," *Hydrological Sciences Journal*, vol. 48, no. 3, pp. 399–411, 2003.
- [44] S. Londhe and S. Charhate, "Comparison of data-driven modelling techniques for river flow forecasting," *Hydrological Sciences Journal*, vol. 55, no. 7, pp. 1163–1174, 2010.

