Arabian Journal of Geosciences (2021) 14: 2033
https://doi.org/10.1007/s12517-021-08150-8

ORIGINAL PAPER

Comparison of different empirical methods and data-driven models for estimating reference evapotranspiration in semi-arid Central Anatolian Region of Turkey

Ibrahim Yurtseven[1] · Yusuf Serengil[1]

Received: 6 May 2021 / Accepted: 29 July 2021 / Published online: 19 September 2021

Abstract
Evapotranspiration (ET) is a major hydrologic process to assess water budgets in terrestrial ecosystems. Since measurement of ET may involve labor intensive field technics in large areas, estimation is preferred in most cases. The FAO Penman-Monteith (PM FAO-56) is a widely recognized reference evapotranspiration (ET_o) method for potential evapotranspiration calculations. The method requires a detailed and comprehensive meteorological data set; however, some empirical methods and models have attempted to calculate ET with less data. In this study, Makkink (ET_Mak), Hargreaves–Samani (ET_Har), Thornthwaite (ET_Thor), Blaney–Criddle (ET_BC), and Penman (ET_PM) were tested against the PM FAO-56. Penman method has achieved the highest accuracy among the empirical methods. In addition, the potential of artificial neural networks (ANN), support vector machines (SVM), random forest (RF), and multiple linear regression (MLR) for estimating ET_o were investigated in a semi-arid Central Anatolian Region of Turkey. The results obtained with the ANN (based on multi-layer perceptron) and SVM models performed better than other tested data-driven models and empirical methods. These models could be used most effectively at elevation range of 850–1000 m. According to our results MLP, SVM, and Penman methods provided good performances in semi-arid regions in agricultural planning and water resources management studies. Furthermore, we concluded that integrating maximum temperature may result in improved accuracy in ET model simulations in semi-arid regions.

Keywords Reference evapotranspiration (ET_o) · Empirical methods · Artificial neural networks (ANN) · Support vector machines (SVM) · Random forest (RF) · Multiple linear regression (MLR)

Introduction

Evapotranspiration (ET) plays a key role in water resources management, agriculture, drought, climate change adaptation, and ecosystem productivity (Currie 1991). There are various methods/models to estimate potential evapotranspiration but most of them give precise outputs for specific climate zones (Lu et al. 2005). ET_p calculated under certain properties can be regarded as the reference crop ET (ET_C). ET_c is usually estimated from reference evapotranspiration (ET_o), crop, and soil coefficients. FAO and working group of the International Commission on Irrigation and Drainage recommended standardized Penman-Monteith reference evapotranspiration (ET_o) as the potential evapotranspiration for short grass or a tall reference crop (alfalfa) (Allen et al. 1998). This hypothetical evapotranspiration considers a reference surface with an assumed crop height of 0.12 m, a fixed surface resistance of 70 s/m, and an albedo of 0.23; and the reference surface closely resembling an extensive surface of green grass of uniform height, actively growing, well-watered, and completely shading the ground. The ET_o calculation is an important issue for computing crop irrigation water requirements in agriculture. The practical value of pan evaporation with empirical coefficients (relating ET_o) has been widely used for 10 days or longer periods (Allen et al. 1998). Furthermore, many empirical or physically based equations have been developed and

Responsible Editor: Zhihua Zhang

Ibrahim Yurtseven
ibrahimy@istanbul.edu.tr

Yusuf Serengil
serengil@istanbul.edu.tr

[1] Department of Watershed Management, Faculty of Forestry, Istanbul University-Cerrahpasa, 34473 Istanbul, Turkey

used to estimate ET_o under the climate regime of the country they were developed. How to choose the appropriate model to estimate ET_o among many evapotranspiration calculations is generally a major problem, and method selection under the climatic conditions of the research area is highly subjective unless certain techniques are used. Generally, empirical ET_o methods can be categorized under six groups: (1) combination (e.g., Shuttleworth); (2) radiation (e.g., Turc, Priestley and Taylor, Makking, Abtew); (3) temperature (e.g., Blaney and Criddle, Hargreaves–Samani, Thornthwaite, Hamon); (4) mass-transfer based (e.g., Penman, Dalton); (5) water budget methods (e.g. , Guitjens); and (6) pan evaporation methods (e.g., Allen et al. 1998). Many studies evaluate the reliability of these alternative empirical ET_o methods for the lack of the calculated ET_o data considering the United Nations Food and Agriculture Organization (FAO) Penman–Monteith (PM FAO-56) as the standard method (Table 1). These studies were conducted for purposes such as the effectiveness, improvement, and performance of PM FAO-56 at regional and global scales. Performances of these ET_o models have also been evaluated under different climate conditions and land cover. Assumptions and inputs are the most important causes for having different results of the methods (Maes et al. 2019).

In recent years, interest has grown in testing models for non-linear relationships. Statistical tests have been proposed in many studies to help analysts check for the presence of non-linearities in an observed time series. Another alternative to ET_o estimation is the application of data-driven models. Recently, machine learning models generated simpler equations and require fewer inputs than the PM FAO-56 method. Thus, they are potentially good alternatives in ET_o calculation. As shown by numerous studies, machine-learning approaches such as Artificial Neural Networks (ANN) have been successfully applied in ET_o research (Zanetti et al. 2007; Traore et al. 2010; Käfer et al. 2020). ANN, which has a nonlinear mathematical structure, trains from the strength of correlation between input and simulated variables by checking previous trends (Yurtseven and Zengin 2013).

Sudheer et al. (2003) used radial basis function (RBF) to simulate crop evapotranspiration (ET_c) for rice crops. The simulated data was compared with the lysimetric data. The results clearly showed that RBF performed good (modeling efficiency of 98.2–99.0%) in ET_o estimation. Trajkovic et al. (2003) used a RBF type of ANN and found that the ANN gives accurate ET_0 estimates. Hashemi and Sepaskhah (2020) also reported the superiority of multi-layer perceptron with sunshine hours and wind speed and the radial basis function with sunshine hours. Zanetti et al. (2007) used the multilayer perceptron for estimating the ET_o by using only data from the maximum and minimum air temperatures in Brazil.

Machine learning approaches using support vector machine (SVM) have also been described and evaluated by many studies (Wen et al. 2015; Chia et al. 2020; Seifi and Riahi 2020). SVM, which is a useful estimator for practical applications, has the ability to provide a powerful algorithm between dependent and independent variables. This algorithm uses robust mathematical equations between dependent and independent variables to solve complex problems (Vapnik 1995). SVM has been a preferred approach as it adopts a global optimum rather than a local optimum compared to ANN method, and is less prone to overfitting than the ANN method. However, SVM models for estimating ET_o had limited applications compared to ANN models. Wen et al. (2015) developed SVM models for ET_o estimation and compared it with ANN model and three empirical models including Priestley-Taylor, Hargreaves, and Ritchie. The study showed that SVM showed relatively superior performance to ANN and empirical equations in modeling ET_o.

In recent years, the random forest (RF) model, which is an ensemble learning method for classification and regression, has become popular due to some of its advantages such as satisfactory performance, ability of preventing overfitting, and user-defined parameter selection in both classification and regression problems (Feng et al. 2017). The relative importance of variables can also be determined by this method. Wang et al. (2019) conducted a competitive analysis using different model-based approaches (random forest, gene-expression programming) on daily climatic data from the 24 meteorological stations recorded from 2010 to 2014 and concluded that random forest-based ET_o models performed slightly better than the gene expression-based models.

Multiple linear regression is a conventional model for estimating the value of one dependent variable based on two or more independent variables with linear relationship (Tabari et al. 2012). Many researchers have attempted to estimate the evaporation values from climatic variables with MLR. Yirga (2019) reported the performance of MLR in ET_o estimation. This research stated that the model is successfully employed for the estimation of the monthly reference evapotranspiration. da Silva et al. (2016) emphasized that models can be regarded as an alternative method to estimate the ET_o when the climatic variables are insufficient for other methods.

In this study, we tested some of the recent approaches in ET_o estimation. Our study area was the middle Anatolian region that is the driest region in Turkey. Drought has become an important and prominent phenomenon in Turkey, and especially semi-humid (semi-dry) drought classes have shifted to semi-dry (dry) conditions in Central Anatolia regions. Besides, the Central Anatolian Region has strong and big potential for marketing and growing of cereal production (wheat, barley, oat, etc.). In recent years, less rain and more ET have led to crop failure and economic losses in the region. Spatial

Table 1 Summary of various methods or model adopted by different authors for ET_o estimation

No	Author(s)	Origin	Using method*	Reference output	Length of record	Conclusion
1	Douglas et al. (2009)	USA	Comparing (TU, PT, PM)	ET (micrometeorological techniques)	1999–2006	PT method was found to be the best performing models for estimating ET_o
2	Li et al. (2016)	China	Comparing (BC, HA, PT, DA, PE and SW)	ET (eddy covariance)	2008–2012	PE, SW, and PT models were more useful for estimating crop ET_o and ET_a in arid regions
3	Sentelhas et al. (2010)	Canada	Comparing (PT, HA, TH, THref)	ETo (PM FAO-56)	2001–2007	PT method was reliable and accurate option for estimating ETo when vapor pressure was deficit and wind speed data were missing, mainly when calibrated locally. When only temperature data were available, adjusted HA and THref methods gave best estimated results to estimate ETo.
4	Tukimat et al. (2012)	Malaysia	Comparing (PT, HA, MK, HA, TH, and BC)	ETo (PM FAO-56)	1972–2001	The radiation-based methods gave more accurate performance compared to temperature-based methods in estimation of ET in the study area of Malaysia.
5	Rácz et al. (2013)	Hungaria	Comparing (PT, FAO-56, SW, SZ, MK, WMO-1966, MA, BC, and PE)	ET (Pan evaporation)	2005–2010	The performance of PT, FAO-56,SW, SZ, and MK methods appears to be superior to the other methods for ET_o calculation
6	Xu and Singh (2002)	Switzerland	Comparing (HA, BC, MK, PT, and RO)	ETo (PM FAO-56)	1990–1994	Method performance gives the following rank of accuracy as compared with the FAO-56 estimates: PT,MK, HA, BC, and RO
7	Lu et al. (2005)	USA	Comparing (TH, HM, HA, TU, MK, and PT)	ET (meteorological techniques)	1978–1990	PT, TU, and HM methods are recommended for regional applications of alternative reference evapotranspiration in the southeastern USA.
8	Lang et al. (2017)	China	Comparing (MK, AB, PT, HA, TH, HM, LI, and BC)	ETo (PM FAO-56)	1962–2013	Radiation-based methods for ET_O estimation was shown the better performing models than temperature-based methods among the selected methods in the study area. Among the radiation-based methods, MK method performed the best results, while HS showed the best performance among the temperature-based methods.
9	Alexandris et al. (2008)	Belgrade	Comparing (PT, TU, MK, HA, and CO)	ETo (PM FAO-56)	2005–2006	PT and CO methods were produced best performing results for ET_o calculation
10	Efthimiou et al. (2013)	Greece	Comparing (FAO-24, MK, TU, PE, PT, LI, KP, HA, and CO)	ETo (PM FAO-56)	1961–2010	The Priestley-Taylor method had the best correlation to the FAO-56-PM method at Krania station, while at Kozane station, the Turc method gave the best estimated values.
11	Tellen (2017)	Cameroon	Comparing (TU, PA, SSmod, BC, BCmod)	ETo (PM FAO-56)	1967–1982	In general, the Stephens and Stewart (1963) modified by Jansen and Haise method produced best statistics result
12	Kisi (2014)	Turkey	Comparing (CO, TU,HA, HS, RI, IR)	ETo (Valiantzas)	1972–2002	CO equation gave the best results out of the nine methods. The worst estimates were shown in the Turc method.
13	Rahimikhoob et al. (2012)	Iran	Comparing (MK, TU, PT, HG)	ETo (PM FAO-56)	1996–2005	PT and HG equations are more applicable in an intermediate humidity region due to higher values of coefficient of determination with PM56 method
14	Fisher and Pringle III (2013)	USA	Comparing (ReS,TU, HA)	ETo (PM FAO-56)	1997–2012	TU method was found to provide better estimates of FAO-56 than the other methods.
15	Issaka et al. (2017)	Qatar	Comparing (BC, HS, JH, LI, TU)	ETo (PM FAO-56)	1985–1998	TU method is easily measurable over an arid area and it is suitable for arid areas such as Doha in Qatar.
16	Hadria et al. (2021)	Morocco	Comparing (HS, RZ, BE, TR, DO)	ETo (PM FAO-56)	2011–2019	Temperature-based models can simulate reliably and accurately reference evapotranspiration. Hargreaves method is not accurate in mountainous zones.

BC, Blaney-Criddle; *MK*, Makking; *TU*, Turc; *HA*, Hargreaves; *HS*, Hargreaves-Samani; *PT*, Priestley-Taylor; *DA*, Dalton; *PE*, Penman; *TH*, Thornthwaite; *THTef*, Thornthwaite with effective temperature; *FAO-56*, Penman-Monteith-FAO-56; *SZ*, Szász; *MA*, Mahringer; *PE*, Pereira; *RO*, Rohwer; *HM*, Hamon; *AB*, 0Abtew; *LI*, Linacre; *CO*, Copais; *PE*, Penman; *KP*, Kimberly-Penman; *PA*, Papadakis; *SS*, Stephen-Steward; *RI*, Ritchie; *IR*, Irmak; *ReS*, reduced set; *JH*, Jensen-Haise; *RZ*, Ravazzani; *BE*, Berti; *TR*, Trajkovic; *DO*, Dorji; *SW*, Shuttleworth variability of precipitation regimes influenced by the topography has been studied before (Türkeş and Tatlı 2011; Schemmel et al. 2013). The complex biotic and abiotic environment in different elevation zones makes it difficult to measure and estimate ET directly or indirectly. With respect to the climatic composition at different elevations, the variation of the ET_o in different elevation is quite complex. The elevation causes a manifold effect in ET_o in different locations since the dynamics of climatic parameters at different altitudes are also different. For example, relative humidity is one of the most relevant meteorological factors in ET_o measurement, and it is affected by elevation with a reverse relationship. Moisture availability affected by relative humidity and absolute vapor pressure decreases with elevation (Duane et al. 2008). Therefore, elevation dominates climatic parameters that affect ET_o at elevation gradients. Furthermore, the spatial variation in ET_o is also affected with R_S received by the surface (Vicente-Serrano et al. 2007). Ma et al. (2019) reported that available energy (shortwave radiation and air temperature) increased with elevation is a more influential factor than water vapor. Sun et al. (2020) found that net R_S leaf area index and air temperature have strong relationship with ET_o in mountainous regions. Wang et al. (2020) emphasized that the FAO-Penman Monteith (PM) and Hargreaves-Samani (HS) perform well as appropriate ET_o estimation methods in high elevation zones. Understanding the topographic characteristics, especially elevation controlling the ET_o in Central Anatolia and its variability, is one of the scientific gaps of climate research of Turkey. Furthermore, the elevation and ET_o interaction in dry regions of Turkey are poorly characterized despite obvious practical importance.

One of the main objectives of this study is to determine possible variations in ET_o at different elevations and the performance of selected methods/models that can be used in estimation of ET_o.

Other objectives were the following:

- To calculate ET_o using six different empirical methods (FAO-56 Penman-Monteith method-ET_o, Hargreaves-ET_Har, Penman-ET_PM, Makking-ET_Mak, Thornthwaite-ET_Thor, and FAO-Blaney-Criddle-ET_BC) and make comparison between these five ET_o methods and the FAO-56 Penman-Monteith method (ET_o) in regional average values of 45 meteorological stations (represent the average of Central Anatolian Region) and four different elevation groups (650–850 m-G1, 850–1100 m-G2, 1100–1350 m-G3, and 1350–1600 m-G4).
- To investigate the accuracy of data-driven modeling such as two different artificial neural network (ANN) techniques, namely the multi-layer perceptrons (MLPs), radial basis neural networks (RBNNs), support vector machine (SVM), random forest (RF), and multi linear regression (MLR) in estimating long-term monthly ET_o by using data from the same 45 stations in Central Anatolian Region in Turkey.
- Statistical evaluation of the outputs of all ET_o approaches and climatic parameters used in the assessment.

Material and method

Study work-flow

The study work-flow (Fig. 1) presents the research steps represented by subdivision method. The methodology essentially seeks the possibility of different PET methods, ANN (MLP and RBF), SVM, RF, and MLR model as an alternative to the respective FAO-PM (ET_o). The flowchart illustrates the primary structure of the model involving three main parts, i.e., calculate ET_o with five simple empirical ET_o methods and PM FAO-56 method and generate alternative ET_o using data-driven models (ANN, SVM, RF, and MLR). The study was carried out in two steps/stages. In the first step, in which denotes regional average, a data formation was prepared by taking the average of the climate data of 45 meteorological stations used in all analysis. At this stage, 45 meteorological stations were evaluated as one station to represent the entire Central Anatolian Region of Turkey. In the second step, the data were grouped according to elevation of meteorological station as main data formations in the paper; this step is termed the "elevation group." Thus, a large dataset was grouped along four different elevation gradients (650–850 m, 850–1100 m, 1100–1350 m, 1350–1600 m) using the elevation of 45 meteorological stations. All analyses were evaluated separately for both data formations. The conceptual background of the study consists of two main parts. First, ET_o was calculated with the equations of different researchers using the unnormalized climate data to compare ET_o (PM FAO–56). Second, the climate data were used in alternative data-driven model based on ET_o calculations. Performance evaluation was used for determining appropriate method or model to estimate ET_o in regional average and grouped data. Therefore, the coefficient of determination (R^2), mean absolute deviation (MAD), Nash–Sutcliffe efficiency (NSE), the index of agreement (d), and percent bias (PBIAS) were used to identify the best method among the empirical methods and data-driven ET_o models.

Study area and data acquisition

According to multiple-year local assessments, Turkey is classified under seven geographical and 8 ecological regions (ecozones) (Serengil 2018). This research was conducted in

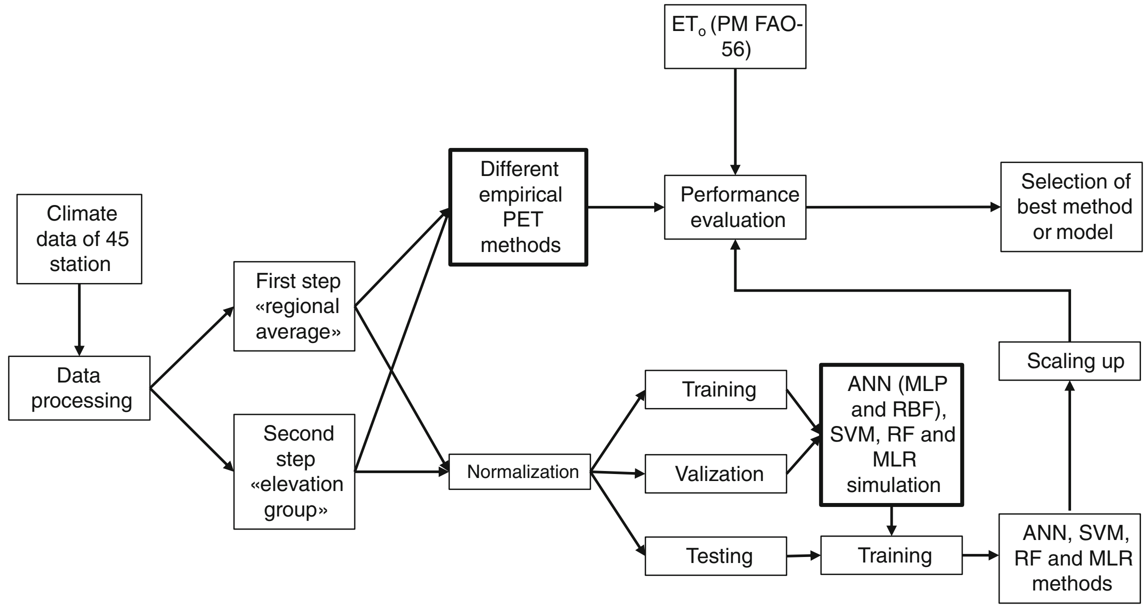

Fig. 1 Study work-flow

the Central Anatolia geographical region, and Central Anatolia Steppe ecozone. Climatological data from 47 synoptic stations located in Central Anatolian Region of Turkey were obtained from the Climate Forecast System Reanalysis (CFSR) global meteorological dataset. The CFRS dataset consist of hourly weather forecast generated by National Weather Service's NCEP Global Forecast Syetems. Studies showed that the CFSR data used in hydrological models provide satisfactory results (Fuka et al. 2014; Dile and Srinivasan 2014). All stations, with 35 years of monthly meteorological data, were selected for analysis. The data covered the time period between January 1979 and December 2013. The locations of the 47 stations are given in Fig. 2, and Table 2 shows some characteristics of these stations.

The Central Anatolian Region of Turkey is a generally semiarid area based on United Nations Environment Program (UNEP) aridity index (Middleton and Thomas 1997) with a size of about 151,000 km^2, representing 21% of the country. The region is located between 31° 21′ to 38° 07′ E longitude and 36° 59′ to 40° 55′ N latitude. In this region, average altitude of 1000 m and low precipitation plateaus are located and it is limited by Bolu-Köroğlu Mountain to the north, Sündiken and Uludağ Mountains to the west, Toros Mountain to the south, and Tecer Mountains of Turkey to the east. As the region is surrounded by high mountains, the humid mild sea air cannot easily penetrate into the region. Therefore, the region has a continental climate with hot and dry summers and cold and snowy winters. In the region, the terrestrial effect increases due to the increase in altitude, and winter temperatures reach extremely low values towards the east. The annual average temperature of the region is 10–11 °C (Table 1). Annual precipitation averages about 418 mm, and the actual amount is determined by elevation. Low precipitation amount in some areas of the region is not sufficient to satisfy the water need of the crops during especially summer months. In a dry period, it would thus be necessary to irrigate the crops, while in average wet seasons, irrigation is not needed in agricultural areas. Low precipitation generally causes low productivity in agriculture. Drought necessitates fallow practice in grain agriculture. The natural vegetation is mostly composed of steppes since drought prevents forest growth.

In this study, data processing follows the raw data download and converts into usable or readable form. The monthly values of maximum temperature (T_{max}), minimum temperature (T_{min}), average temperature (T_{avg}), precipitation (P), average wind speed (U), average (RH_{avg}), maximum (RH_{max}), minimum relative humidity (RH_{min}), and average solar radiation (R_S) were obtained for 45 stations located in Central Anatolian Region. There are two steps in the study. (1) The regional average of the Central Anatolian Region for each parameter was calculated by taking the average of the values obtained from 45 stations. Therefore, each station has not been evaluated separately in this first step. (2) In the second step, the 45 different climate stations were divided into four different elevation groups as follows: 650–850 m considered as "low elevation group-G1," 850–1100 m considered as "moderate elevation group-G2," 1100–1350 m considered as "high elevation group-G3," and 1350–1600 m considered as "very high elevation group-G4." The results of five different literature-based equations (ET_Har, ET_PM, ET_Mak, ET_Thor, and ET_BC), ANN (MLP and RBF), SVR, RF,

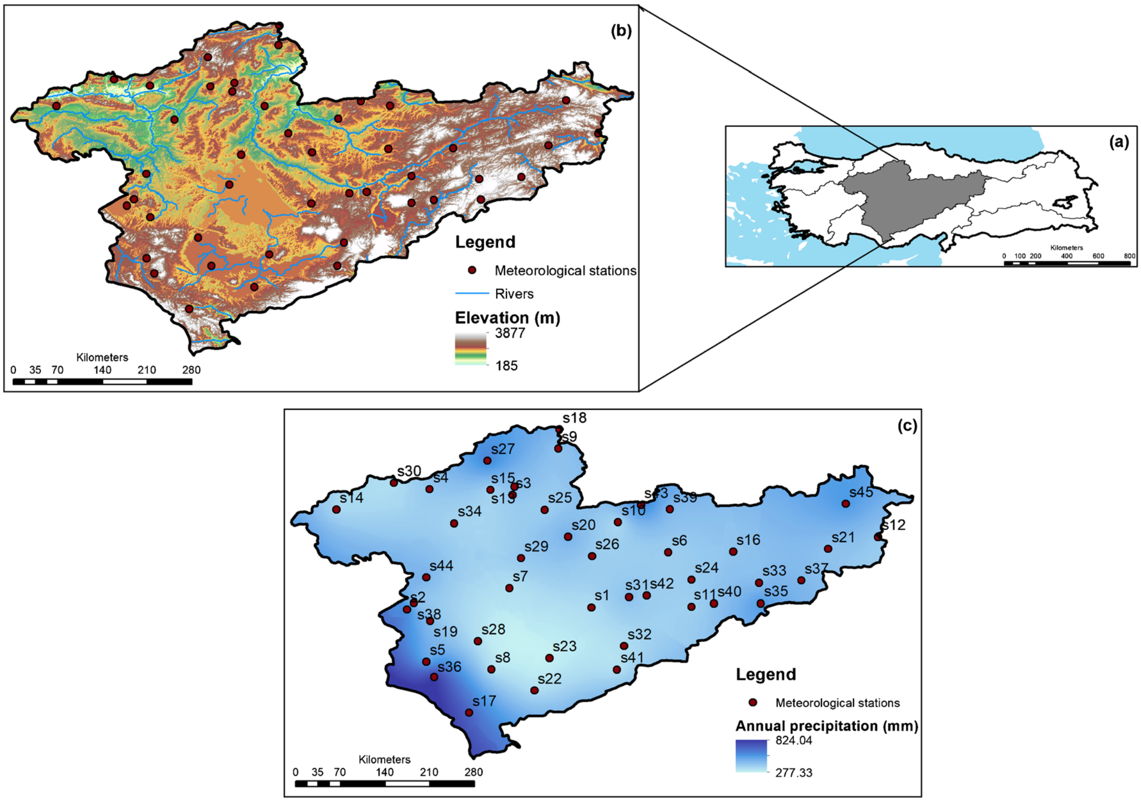

Fig. 2 Elevation (**b**) and precipitation zones (**c**) with meteorological stations in Central Anatolian Region of Turkey (**a**)

and MLR models for 4 different elevation groups were subjected to performance evaluations with target output ET_o. The objective was to compare models for different elevation groups located in different local climatic conditions.

Empirical ETo methods

The following ET methods have been chosen for the assessment:

(a) FAO-56 Penman–Monteith method (ET_o): This method is considered the most precise method to estimate ET_o. The FAO Penman-Monteith method for calculating reference (potential) evapotranspiration ET_o can be expressed as (Allen et al. 1998) follows:

$$\mathrm{ETo} = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T_a + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{1}$$

where ET_o= reference evapotranspiration (mm day^{-1}); Δ is the slope of the saturated vapor pressure curve (kPa ^{8}C^{-1}); R_n is the net radiation (MJ m^{-2} day^{-1}); G is the soil heat flux density (MJ m^{-2} day^{-1}), considered as null for daily estimates; T is the daily mean air temperature (°C) at 2 m, based on the average of maximum and minimum temperatures; U_2 is the average wind speed at 2 m height (m s^{-1}); e_s is the saturation vapor pressure (kPa); e_a is the actual vapor pressure (kPa); ($e_s - e_a$) is the saturation vapor pressure deficit (Δe, kPa) at temperature T; and γ is the psychrometric constant (0.0677 kPa °C^{-1}).

The following equations were recommended by Allen et al. (1998) to estimate Rn:

$$R_n = R_{ns} - R_{nl} \tag{2}$$

$$R_{ns} = 0.77SR \tag{3}$$

$$R_{nl} = \left[\sigma\left(\frac{T\max_K{}^4 + T\min_K{}^4}{2}\right)\left(0.34 - 0.14\sqrt{e_a}\right)\left(1.35\frac{Rs}{Rso} - 0.35\right)\right] \tag{4}$$

$$Rso = 0.75Ra \tag{5}$$

where R_{ns} is the net shortwave radiation (MJ m^{-2} day^{-1}); R_{nl} is the net longwave radiation (MJ m^{-2} day^{-1}); R_s is the incoming solar radiation (MJ m^{-2} day^{-1}); σ is the Stefan–Boltzmann constant (4.903×10^{-9} MJ K^{-4} m^{-2} day^{-1}); $T\max_K$ is the

Table 2 Properties of 45 meteorological stations with long-term average climatic conditions. Elevation of stations, elevation group, annual average temperature (T_{avg}), annual total precipitation (P), annual total evapotranspiration (ET_O), UNEP aridity index (Middleton and Thomas 1997), and UNEP aridity index zone (Middleton and Thomas 1997)

Station code	Near city name	Elevation (m)	Elevation group	Annual T_{avg} (°C)	Annual P (mm)	Annual ET_O (mm)	P/ET_O (UNEP aridity index)	UNEP aridity index zone
s1	Aksaray	960.77	G2	17.89	337.46	1389.16	0.24	Semi-arid
s2	Akşehir	1002	G2	16.13	577.58	1573.51	0.37	Semi-arid
s3	Ankara	890.52	G2	17.10	406.01	1404.85	0.29	Semi-arid
s4	Beypazarı	682	G1	17.83	398.36	1480.10	0.27	Semi-arid
s5	Beyşehir	1148	G3	15.52	489.08	1347.45	0.36	Semi-arid
s6	Boğazlıyan	1067	G2	16.27	372.57	1548.98	0.24	Semi-arid
s7	Cihanbeyli	968.73	G2	18.39	323.16	1588.45	0.20	Semi-arid
s8	Cumra	1013	G2	18.78	322.01	1518.72	0.21	Semi-arid
s9	Çankırı	751	G1	15.10	407.62	1281.01	0.32	Semi-arid
s10	Çiçekdağı	900	G2	16.10	347.60	1464.67	0.24	Semi-arid
s11	Develi	1180	G3	14.80	369.47	1474.53	0.25	Semi-arid
s12	Divriği	1120	G3	13.11	390.16	1361.68	0.29	Semi-arid
s13	Esenboğa	959.33	G2	4.32	410.83	1682.88	0.24	Semi-arid
s14	Eskişehir	787	G1	4.77	348.09	1678.95	0.21	Semi-arid
s15	Etimesgut	806.15	G1	4.55	360.11	1766.77	0.20	Semi-arid
s16	Gemerek	1171	G3	2.94	403.90	1758.17	0.23	Semi-arid
s17	Hadim	1552	G4	3.25	659.43	1544.51	0.43	Semi-arid
s18	Ilgaz	885	G2	2.08	462.81	1311.70	0.35	Semi-arid
s19	Ilgın	1034	G2	4.07	424.16	1775.59	0.24	Semi-arid
s20	Kaman	1075	G2	5.08	455.35	1874.10	0.24	Semi-arid
s21	Kangal	1541	G4	1.75	409.00	1611.47	0.25	Semi-arid
s22	Karaman	1023.05	G2	5.40	330.46	2249.01	0.15	Arid
s23	Karapınar	1004	G2	5.92	286.73	1804.24	0.16	Arid
s24	Kayseri	1092	G2	15.94	390.17	1534.96	0.25	Semi-arid
s25	Kırıkkale	750.88	G1	5.15	374.63	1834.01	0.20	Semi-arid
s26	Kırşehir	1007.17	G2	5.19	382.13	1901.69	0.20	Semi-arid
s27	Kızılcahamam	1033	G2	14.24	581.15	1224.40	0.47	Semi-arid
s28	Konya	1030.61	G2	17.23	323.09	1401.44	0.23	Semi-arid
s29	Kulu	1010	G2	17.30	391.24	1542.12	0.25	Semi-arid
s30	Nallıhan	650	G1	16.88	331.94	1435.93	0.23	Semi-arid
s31	Nevşehir	1259.54	G3	16.37	414.33	1483.35	0.28	Semi-arid
s32	Niğde	1210.5	G3	15.18	332.55	1423.71	0.23	Semi-arid
s33	Pınarbaşı	1500	G4	13.08	423.69	1377.75	0.31	Semi-arid
s34	Polatlı	886	G2	17.92	356.86	1399.78	0.25	Semi-arid
s35	Sarız	1500	G4	13.08	521.54	1371.26	0.38	Semi-arid
s36	Seydişehir	1131	G3	15.36	772.11	1369.61	0.56	Dry subhumid
s37	Sivas	1285	G3	13.43	438.95	1391.10	0.32	Semi-arid
s38	Sivrihisar	1070	G2	17.26	402.14	1418.86	0.28	Semi-arid
s39	Sorgun	1110	G3	15.65	431.66	1417.85	0.30	Semi-arid
s40	Tomarza	1397	G4	14.29	403.84	1459.90	0.28	Semi-arid
s41	Ulukışla	1453	G4	15.32	320.08	1407.04	0.23	Semi-arid
s42	Ürgüp	1060	G2	16.38	380.47	1491.67	0.26	Semi-arid
s43	Yozgat	1298.43	G3	5.19	587.72	1423.86	0.41	Semi-arid
s44	Yunak	1140	G3	18.34	450.17	1464.30	0.31	Semi-arid
s45	Zara	1347	G3	13.00	531.80	1320.66	0.40	Semi-arid

maximum temperature (K); $T\min_K$ is the minimum temperature (K); SR/SRo is ratio between the incoming solar radiation and the clear sky solar radiation (MJ m^{-2} day^{-1}), which is less or equal to 1; and Ra is the extraterrestrial solar radiation (MJ m^{-2} day^{-1}). The other parameters of equation of ETo were determined as follows:

$$\varDelta = \frac{4098[0.6108\exp(17.27T/(T+237.3))]}{(T+237.3)^2} \tag{6}$$

$$e_s = \frac{\left[0.6108\ \exp\left(\frac{(17.27T\max_c)}{T\max_c+237.3}\right)\right] + [0.6108\exp((17.27T\min_c)/(T\min_c+273.3))]}{2} \tag{7}$$

$$e_a = \frac{\mathrm{RH}}{100} e_s \tag{8}$$

where $T\max_c$ is the maximum temperature (°C); $T\min_c$ is the minimum temperature (°C); and RH is the mean daily relative humidity, calculated from maximum and minimum values.

The following equation was used to the equation of a logarithmic wind speed profile to convert wind speed data obtained at height of 10 m to the standard height of 2 m.

$$U_2 = U_Z\left[\frac{4.87}{\ln(67.8z-5.42)}\right] \tag{9}$$

where z is the height of the wind speed measurement (=10 m).

(b) Hargreaves method (ET_Har): The Hargreaves method (Hargreaves and Samani 1985), which is a temperature based equation, estimates ETo (mm d^{-1}); using only the maximum and minimum temperatures, and is expressed by Eq. 10:

$$\mathrm{ETo} = C_0 Rs(T\max_c - T\max_c)^{0.5}(T+17.8) \tag{10}$$

where Rs is the extraterrestrial solar radiation, in mm day^{-1}; and C_o the conversion parameter (=0.0023).

(iii) Penman method (PET_PM): This method is still a mass-transfer-based method in estimating free water surface evaporation E because of its simplicity and reasonable accuracy. Penman (1948) proposed the following equation.

$$\mathrm{ETo} = 0.35\left(1+\frac{0.98}{100U_2}\right)(e_s - e_a) \tag{11}$$

where U_2 wind speed at 2 m high in miles day^{-1}; e_s the saturation vapor pressure at the temperature of the water surface; e_a the actual vapor pressure in the air.

(iv) Makking method (PET_Mak): For estimating potential evapotranspiration (mm d^{-1}) Makking (1957) proposed the following equation.

$$\mathrm{ETo} = 0.61\frac{\Delta}{\Delta+\gamma}\frac{R_s}{\lambda} - 0.12 \tag{12}$$

where R_s = the total solar radiation in cal cm^{-2} day^{-1}; Δ = the slope of saturation vapor pressure curve (in mb/^{8}C); γ = the psychrometric constant (in mb/^{8}C); λ = latent heat (in calories per gram); P = atmospheric pressure (in millibar).

(e) Thornthwaite method (PET_Thor): The Thornthwaite method is a temperature-based method for calculating PET can be expressed as (Thornthwaite 1948):

$$\mathrm{ETo} = \begin{cases} 0\ , & \mathrm{T_{avg}} < 0°\mathrm{C} \\ 16\left(\frac{10\ T_{\mathrm{avg}}}{I}\right)^a, & 0°\mathrm{C} \le \mathrm{T_{avg}} \le 26.5°\mathrm{C} \\ -0.43{T_{\mathrm{avg}}}^2 + 32.24T_{\mathrm{avg}} - 415.85\ , & \mathrm{T_{avg}} > 26.5°\mathrm{C} \end{cases} \tag{13}$$

$$I = \sum_{k=1}^{12}(0.2T_k)^{1.514} \tag{14}$$

$$a = 0.000000675I^3 - 0:0000771I^2 + 0.01792I + 0.49239 \tag{15}$$

where ETo = reference evapotranspiration estimated by Thornthwaite equation (mm month^{-1}), T_{avg} = mean monthly air temperature (°C), I = thermal index imposed by the local normal climatic temperature regime, and a = exponent being a function of I. The value of a varies from 0 to 4.25, while the thermal index I varies from 0 to 160.

(f) FAO Blaney Criddle method (FAO_BC): Blaney-Criddle equation (BC) is a simpler method comparing than other empirical methods and the method use only air temperature as an input data. The equation calculates evapotranspiration for a "reference crop" and this crop is an actively growing green grass with 8–15 cm high.

Blaney and Criddle (1950) proposed a very simplified calculating approach of the temperature-based equation.

$$\mathrm{ETo} = \mathrm{kp}\ (0.46T_a + 8.13) \quad (16)$$

where ETo = potential evapotranspiration from a reference crop, in mm, for the period in which p is expressed; T_a = mean temperature in °C; p = percentage of total daytime hours for the used period (daily or monthly) out of total daytime hours of the year (365 × 12); k = monthly consumptive use coefficient, depending on vegetation type, location and season, and for the growing season (May to October); k varies from 0.5 for orange tree to 1.2 for dense natural vegetation.

ANN method

Soft computing methods such as artificial neural networks (ANN) have been successfully employed to develop a new estimation model(s) for estimating the available model parameters. ANN is an information processing system that consists of three main layers as input, hidden, and output. ANN works in layers where send parallel-operated information with a series of processing elements called neurons. The function of these neurons provides various conversion functions for synaptic weights with their information. Training was occurred in this process. All neurons receive weighted inputs which run as interconnect between input variables or the outputs, add a bias term and pass the result by an activation function. The basis of this process can be formulized in the following equations.

$$I_J = \sum_{i=1}^{n} w_{ij}x_i + b_i \quad (17)$$

$$y_j = f(I_J) \quad (18)$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (19)$$

where I_j is the activation value of neuron j of the ith layer; w_{ij} is the weight of the ith input and the neuron j of the layer; x_i is the ith input value, b_i is the ith bias term, y_j is the output of the neuron j, and $f(x)$ is the activation function.

In ANN and MLR, the variables in dataset were normalized to increase the model performance. Min-max feature scaling (unity-based normalization) is used to bring all values into the range 0 and 1. The general form of normalization that is using in this study is presented in Eq. 20:

$$X_{\mathrm{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (20)$$

In this study, ANNs of the multi-layer perceptron (MLP) and radial basis function (RBF) were employed. The back propagation learning algorithm was used in MLP training process. A structure of MLP consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. Figure 3 represents a three-layer structure of MLP. Each neuron that uses a nonlinear activation functions except for the input layer. Every node is fully connected in MLP, and each node connects with a weight of w_{ij} and K_j from input layer to hidden layer and hidden layer to output layer, respectively.

Radial basic functions (RBF) calculate distance criteria with respect to the center, and the algorithm can be constructed accordingly. Figure 4 represents a RBF structure consisted of a three-layer structure namely (1) input layer, (2) hidden layer, and (3) output layer. The general construction is just like a MLP but there are some differences between MLP and RBF. The most characteristic feature of the RBF network is the activation function ($H_p(x)$ as networks neuron) in hidden layers using Gaussian Bell function that is the most widely used function of RBF (Fig. 4). This function calculates the distance between the neuron center in the hidden layer and the input vector for each neuron in the input layer. The final output is obtained by running sum of dot products of activation function and distance. Therefore, it describes the way that the unit responds to the total input.

We selected parameters of the input layer considering using correlation performance with the reference evapotranspiration (Table 2). Some monthly climate variables that are $T_{\max}$, RH_{avg}, and R_S were used in the input layer. The optimum hidden layer node numbers of the ANN models were obtained after trying different hidden layer network structures that errors can be minimized. The optimum iteration number of ANN networks was also tried. The training of the ANN models was stopped at 250 iterations due to the mean square error between the observed and estimated values decreased with increasing iteration numbers until this number of iterations. The learning process of the MLP and RBF was carried out with daily data series extracted from the 45 selected locations between January 1979 and January 2004 (70% of the whole data set). The data series from January 2004 to July 2014 (30% of the whole data set) were used for testing. The hyperbolic tangent and SoftMax activation functions were used for the hidden nodes for MLP and RBF models, respectively. It was found that the network structure of 3-5-1 in MLP and 3-9-1 in RBF leads to the best results. 3-5-1 denotes an MLP model comprising 3 inputs, 5 hidden, and 1 output node.

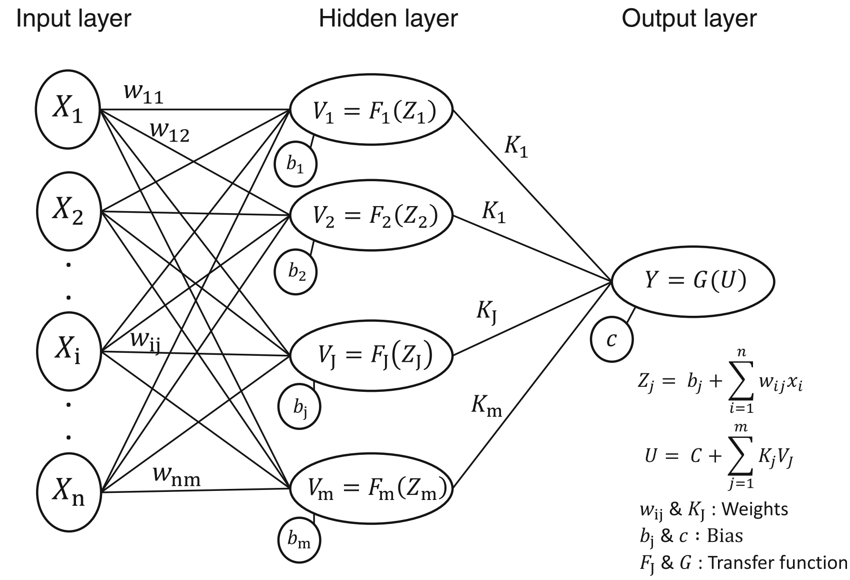

Fig. 3 General architecture of the MLR

MLR method

Regression analysis is one of the statistical tools, which can be considered the process as fitting a model to data. In a linear regression model, data and linear functions can be used to construct the relation that model real-world applications and output parameters are estimated from the data. MLR use several (two or more) explanatory (independent) variables to estimate the outcome of a response (dependent) variable with a linear equation to fitting a linear model. The independent variable x is associated with a value of the dependent variable y in MLR analysis. A typical MLR model expressed as in Eq. 21 below:

$$\widehat{Y} = a_0 + \sum_{j=1}^{m} a_j X_j \tag{21}$$

where $\hat{Y}$ is the model's output, X_j (from X_1 to X_m) is the independent input variables to the model, and a_j (from a_0 to a_m) is partial regression coefficients. The magnitude of each regression coefficient (a_j) in MLR model shows explanatory power of relationship between dependent and independent variables.

SVM method

Support vector machine (SVM), which is a well-known machine-learning method based on classification and regression

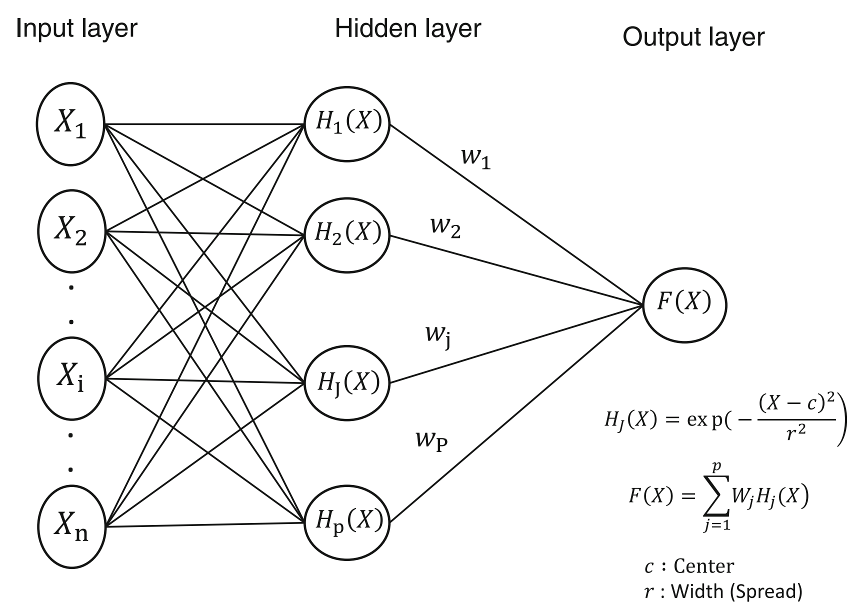

Fig. 4 General architecture of the RBF

analysis theory introduced by Vapnik (1995). The optimal support vector network automatically generated SVMs network architecture while ANN architecture generally involves manual trial-and-error procedures. The types of kernel functions namely linear, sigmoid, polynomial, radial basis function, and multi-layer perceptron are successful in explaining complex data sets. In this study, linear kernel function uses in SVM models. The kernel function is similar to a two-layer perceptron model of the neural network. Unlike the process in standard neural network, the weight of the network is found by solving a quadratic programming problem with linear constraints. In general architecture of the SVM (Fig. 5), the final output connected with hidden nodes are the support vectors (SVs) of the SVM and the weights of SVM network.

The relationship between a dependent variable (y) and a set of independent variables (x) is determined by $f(x)$ in SVM for regression, according to the following equation:

$$F(X) = \sum_{k=1}^{n} \overline{a}_n.K(x, x_n) + b \tag{22}$$

where $\overline{a}_n$ is the Lagrange multipliers, B is a bias term, and $K(x, s_n)$ is the kernel function which is based upon reproducing kernel Hilbert spaces. In this study, the input vectors (x_n) refer to the daily records of $T_{\max}$, RH_{avg}, and R_S while the target value (y) refers to ET_0 values calculated using the FAO-56 PM. In this study, the SVM (100, 10) model has the regularization constant = 100 and width of the RBF kernel = 10.

RF method

Random forest is one of the machine learning models that can be applied to both regression and classification problems. The algorithm uses decision trees using a CART-like procedure that uses a subset of observations through the bootstrap approach (Tsangaratos and Ilia 2017). It is necessary to understand the decision trees structure, which is the basic part of the model on the basis of the random forest. In the model, many individual trees are created by sampling the variables in the data set. Random forest aims to provide better accuracy by using these many decision trees to create a forest. The subsets of variables are generated in the method and each node in the decision tree is divided by the best of this subset of variables. Each variable is classified by each decision tree and thus contribution of variables is well determined to explaining the variance in the dependent variable. Breiman (2001) introduced that many regression trees in RF are installed on marginal functions which are dependent on random vector (Θ), indicator function (I), and specified numerical predictor $h_k(X)$. The marginal functions might be given as follows (Breiman 2001):

$$\mathrm{mg}(X, Y) = \mathrm{av}_k I(h_k(X) = Y) - \max_{j \neq Y} \mathrm{av}_k I(h_k(X) = j) \tag{23}$$

The overall result is given as the average of the sub-results from each tree. The average generalization error of RF can be given as follows:

$$\mathrm{PE}^*(\text{forest}) = P_{X,Y}(\mathrm{mg}(X, Y) < 0) \tag{24}$$

There are two theorems that can be given in RF algorithm.

Theorem 1 By the number of trees increases, we will have the following:

$$P_{X,Y}\left(P_\theta h(X, \theta) = Y) - \max_{j \neq Y} P_\theta(h, \theta) = j < 0\right) \tag{25}$$

The average generalization error of a tree will be as follows:

$$\mathrm{PE}^*(\text{tree}) = P_\theta P_{X,Y}(Y - h(X, \theta))^2 \tag{26}$$

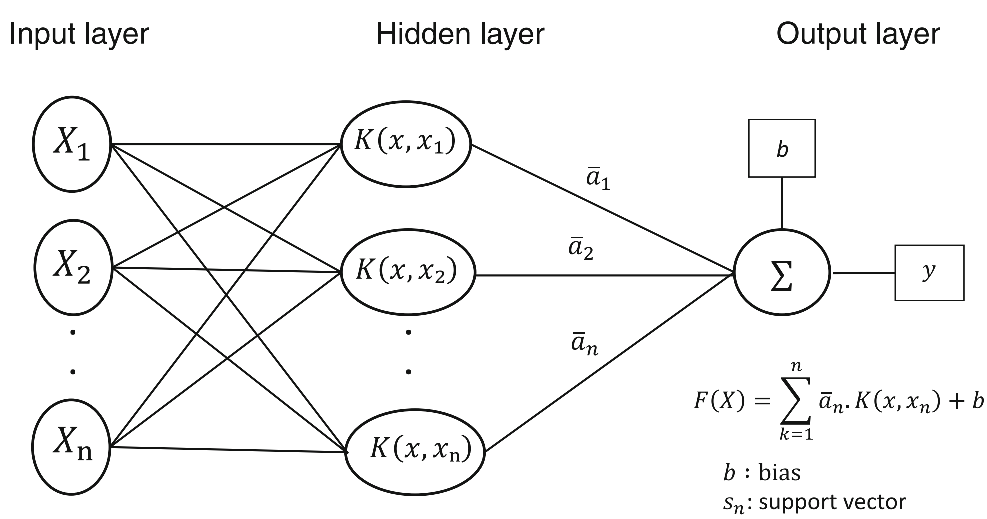

Fig. 5 General architecture of the SVM

Theorem 2 Suppose that PY = $P_X h(X, \theta)$ for all Θ, so:

$$\mathrm{PE}^*(\text{forest}) \leqslant \overline{\rho}.PE^*(\text{tree}) \tag{27}$$

where $\overline{\rho}$ is represented as a weighted correlation between the $Y - h(X, \theta)$ and $(Y - h(X, \theta'))$ (Breiman 2001).

Overfitting, which is one of the biggest problems of decision trees, is decreasing with training on different data sets in the random forest model. In addition, the chance of finding an outlier in subset of variables created by bootstrap method is reduced. The random forest training algorithm (for both classification and regression) applies bootstrap aggregating, or bagging, to tree learners. More details about random forest can be found in Breiman (2001). In this study, RF is used as regression model to estimate ETo. The important tunable parameters are the number of trees (n_{tree}) and the number of estimators in the random subset of each node (m_{try}). The default values of m_{try} (one-third of all estimator variables) were used in this study. The process of n_{tree} decision which affects the forecast performance was used during parameter optimization to yield the minimum error. An iterative evaluation and out-of-bag error (mean squared error for regression problems) were used as the selection criteria in n_{tree} defining. The number of trees was especially used in terms of parameter optimization to yield the minimum error in the study. In general, RMSE decreased with increasing n_{tree}, and r increased correspondingly. In this study, two number of trees were considered differently, for first forest 100 trees and second forest 30 trees. Since the 100-tree gives, the random forest with 100 trees is not included in the results section due to its results are very similar to the results of RBF (ANN). Thus, the random forest with 30 trees was considered in the evaluation of the study.

Performance criteria

Two performance criteria are used in this study to assess the goodness of fit of the models, which are R^2, root mean square error (RMSE), Nash Sutcliffe efficiency (NSE), the index of agreement (d), and percent bias (PBIAS) by using the following equations (Moriasi et al. 2015).

$$R^2 = \left[\frac{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)\left(P_i - \overline{P}\right)}{\sqrt{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2}\sqrt{\sum_{i=1}^{n}\left(P_i - \overline{P}\right)^2}}\right]^2 \tag{28}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \tag{29}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\left(O_i - \overline{O}\right)^2} \tag{30}$$

$$d = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}\left(\left|P_i - \overline{O}\right| + \left|O_i - \overline{O}\right|\right)^2} \tag{31}$$

$$PBIAS(\%) = \frac{\sum_{i=1}^{n} O_i - P_i}{\sum_{i=1}^{n} O_i} x100 \tag{32}$$

where O_i is the results of methods or model as ETo in mm d^{-1}; P_i is the ET_o in mm d^{-1}; $\overline{O}$ is the results of methods or model as ETo, and n is the total number of data.

Results and discussion

In the first step of the study, which denotes regional average, all approach and analysis were made by considering the average of data from 45 meteorological stations that represent the Central Anatolian Region. At this step, the stations were not compared, and the Central Anatolian Region was evaluated as a single station by taking the average of all stations. In the second step, 45 stations were divided into 4 groups according to their elevations, and the meteorological dataset of each stations are averaged within their elevation group. Thus, the performance of models and methods was analyzed according to four different groups in the second step.

The results of first step (regional average)

The mean, minimum, maximum, standard deviation, variation coefficient, and skewness of monthly statistical parameters of regional average dataset for the entire time series are given in Table 3. The statistical parameters of the training, testing, and whole data are shown in the table separately. The performance of ANN models was affected by skewness of the time series data (Zheng et al. 2018). It was shown that ET_o and all variables have quite low skewness values in the complete dataset. The precipitation shows higher skewed distribution comparing other parameters for each period (see SK values in Table 3). Accordingly, the skewness values for all data sets were seen to be roughly similar although the SK values of $T_{\min}$ quite differed from others for each period. The greater of CV

Table 3 Statistical parameter of climatic data and ET_o in training period, testing period, and entire period

Period	Stat. Para.	ET_o	T_{max}	T_{min}	T_{avg}	P	U_2	RH_{max}	RH_{min}	RH_{avg}	R_S
All	Mean	127.96	22.74	1.48	11.54	47.52	3.00	83.27	41.91	63.18	18.24
	Std.	83.47	10.08	10.97	10.47	41.60	0.61	13.35	17.32	16.35	7.58
	SK	0.52	−0.25	0.28	0.21	1.50	0.50	−0.91	0.71	−0.11	−0.06
	Min.	7.39	−1.48	−34.38	−22.23	0.00	1.34	27.03	5.06	10.00	3.86
	Max.	445.58	44.32	41.11	36.20	550.07	5.69	100.00	99.00	96.75	33.30
	CV	0.65	0.44	7.42	0.91	0.88	0.21	0.16	0.41	0.26	0.42
	R^2	-	0.84*	0.58	0.43	0.23	0.00	0.60	0.53	0.68*	0.79*
Training	Mean	119.27	23.31	2.97	11.03	42.77	2.90	82.26	42.11	62.14	18.24
	Std.	76.92	9.88	11.18	10.42	36.79	0.57	13.78	17.30	16.41	7.59
	SK	0.37	−0.24	0.32	0.28	1.16	0.59	−0.92	0.72	−0.17	−0.07
	Min.	7.77	−1.48	−32.90	−22.23	0.00	1.34	27.03	5.06	10.00	4.29
	Max.	401.47	44.32	41.11	36.20	311.92	5.58	100.00	99.00	96.56	32.49
	CV	0.64	0.42	3.76	0.94	0.86	0.20	0.17	0.41	0.26	0.42
	R^2		0.84*	0.57	0.38	0.24	0.00	0.57	0.51	0.65*	0.79*
Testing	Mean	131.70	21.40	−2.01	12.73	58.61	3.22	85.61	41.43	65.59	18.23
	Std.	85.86	10.40	9.61	10.48	49.34	0.65	11.96	17.36	15.97	7.56
	SK	0.55	−0.25	−0.16	0.06	1.56	0.21	−0.77	0.68	0.05	−0.06
	Min.	7.40	−0.46	−34.38	−16.53	0.00	1.46	41.62	5.57	30.14	3.86
	Max.	445.59	42.98	29.00	36.20	550.07	5.69	99.83	91.13	96.75	33.30
	CV	0.65	0.49	−4.78	0.82	0.84	0.20	0.14	0.42	0.24	0.41
	R^2		0.86*	0.65	0.62	0.23	0.00	0.68	0.59	0.76*	0.82*

Std., standard deviation; *SK*, skewness; *CV*, coefficient of variation; R^2, coefficients of determination with ET_o; * $p<0.05$

values, which is defined as the standard deviation divided by the mean, shows the greater level of dispersion around the mean. The mean ET_o (131.70 mm/month) in testing period set is quite higher than the mean ET_o in the training and whole data period (119.27 and 127.96 mm/month, respectively). As can be seen from the R^2 in whole series, T_{max} (R^2= 0.84, $p<0.05$), RH_{avg} (R^2= 0.68, $p<0.05$), and R_S (R^2= 0.79, $p<0.05$) are closely correlated with ET_o.

In this study, monthly ET was estimated using six different data-driven models including ANN (MLP and RBF), SVR, RF, and MLR. The ET_MLP, ET_RBF, ET_SVM, ET_RF, and ET_MLR models use the same input variables. The Penman-Monteith FAO-56 equation (ET_o) was accepted as the reference equation; and other empirical equations (Hargreaves-Samani, Penman, Makking, Thornthwaite, Blaney Criddle), data-driven models (ET_MLP, ET_RBF, ET_SVM, ET_RF), and statistical model (ET_MLR) were compared with ET_o. Table 4 shows the results of all models and equations based on the MAD, RMSE, *d*, and NSE calculations in training and testing period. Generally, considering its high R^2 and low MAD and RMSE, the ET_RBF model and ET_PM formula produced better results in the field of this study within all equation methods, while the worst performance belongs to the ET_Mak in data-driven models. This result is similar in the Kingdom of Saudi Arabia where the Makking equations perform worse than different selected methods (Islam et al. 2020). It is clear from Table 4 that the ET_MLP and SVM model outperformed all other models in terms of all performance criteria in training period. ET_RF and ET_MLR equation results are close to each other, based on their high R^2 and low RMSE in training period.

It is apparent that all of the methods and models performed well in training and testing periods, and the values of RMSE, *d*, and NSE had very small difference between training and testing periods, and all R^2 were also greater than 0.85. In testing periods, it is apparent that MLP (R^2=0.999, $p<0.05$) and SVM models (R^2=0.998, $p<0.05$) were better than others in testing period for ET_o estimation, (Table 4). Therefore, ET_MLP and ET_SVM were selected as the best fit models for estimating the ET_o in training and testing period. The performance of the MLP and SVM model on the testing dataset showed that the MLP and SVM models can be used to provide accurate and reliable ET_o estimations. Based on the results of Table 4, Penman method (ET_PM) whose input combinations were U, actual and saturation vapor pressure had the highest value of R^2 (0.989; $p<0.05$), NSE (0.99), and *d* (0.99), than other empirical equations in the training period. The results of performance evaluation showed that ET_PM also performs clearly better than other empirical methods in testing period based on R^2 (0.988), RMSE (20.74 mm/month), *d* (0.98), and

Table 4 The performance statistics of the models and equations in training and testing period

Methods	Training period					Testing period				
	R^2	MAD	RMSE	*d*	NSE	R^2	MAD	RMSE	*d*	NSE
ET_Har	0.924*	69.13	91.09	0.85	0.63	0.930*	69.27	91.56	0.85	0.62
ET_PM	0.989*	14.65	18.23	0.99	0.99	0.988*	17.42	20.74	0.98	0.98
ET_Mak	0.874*	29.32	35.45	0.96	0.94	0.854*	32.00	38.46	0.95	0.93
ET_Thor	0.941*	58.23	63.02	0.91	0.82	0.938*	55.30	59.64	0.92	0.84
ET_BC	0.960*	26.10	31.28	0.97	0.96	0.951*	25.14	30.71	0.97	0.96
ET_MLP	0.999*	1.58	2.02	1.00	1.00	0.999*	1.69	2.22	1.00	1.00
ET_RBF	0.956*	12.62	17.14	0.99	0.89	0.956*	12.70	17.47	0.99	0.99
ET_SVM	0.999*	9.50	10.18	1.00	0.81	0.998*	8.49	9.31	1.00	1.00
ET_RF	0.978*	30.04	33.03	0.97	−2.74	0.977*	30.49	33.25	0.97	0.95
ET_MLR	0.982*	8.87	10.61	1.00	0.72	0.979*	9.45	11.43	0.99	0.99

*p<0.05

NSE (0.98). In both periods, it was found that the ET_MLP method provides best accuracy (R^2=0.998), highest *d* value (1.00), and lowest RMSE value (2.02 mm/month) in all methods. Malik et al. (2017) reported better performances (RMSE = 0.214 mm/month) by multi-layer perceptron neural network to estimate monthly pan-evaporation (EPm) in Indian central Himalayas. This indicates that the accuracy of the models may vary according to the climate of the research site, the type of climatic data, and the sample size. In this study, ET_RBF and ET_RF models have almost same R^2, and both models performed worse than ET_SVM and ET_MLP models in testing period. As can be seen from the Table 4, all performance statistics illustrated a reasonably better performance for all data-driven models than empirical methods. These results are parallel with previous studies (Karimaldini et al. 2011; Tabari and Talaee 2013) which indicate that the performances of data-driven models were better than local calibrated physical model or conventional methods. It is evident that all data-driven models and statistical method (ET_MLR) are rather simple in terms of input parameter, and its difference from empirical methods is that it contains RH_{avg} in the input parameters group. The results of models show that the models, in which T_{max}, RH_{avg}, and R_S are needed, performed well in reference to ETo modeling and could be used with limited weather data. The results of performance show that the presence or absence of critical input significantly impacted the performances of equation methods. However, the performance values can vary with model dynamics (numbers of hidden nodes, epoch values, type of activation functions used, etc.) in data-driven models with the same input set.

The comparison of the ET_o values calculated by FAO PM-56 and the values estimated by different empirical methods and data-driven models in testing period was shown in Figs. 6 and 7, in the form of line graphs, scatter plots, and residual graphs. The slope of regression lines ranged from 0.23 to 1.08 in empirical methods while in data-driven models, these values ranged from 0.89 to 0.99. The ETo values estimated by the ET_MLP, ET_SVM, and ET_PM were close to that calculated using the ET_o values and followed the same trend as in ET_o. It was clearly shown from the figures that the ET_MLP, ET_SVM, and ET_PM models closely follow the corresponding ET_o values and less scattered estimates compared to other methods. Therefore, these methods are considered as best alternatives for estimating monthly averages of monthly ET_o based on the values of R^2. The slope of regression lines for each method was <1.0 except for the ET_PM and ET_RF method, indicating that ET_PM and ET_o methods had strong relationships with the ET_o among all empirical methods and data-driven models, respectively. However, in general, the estimated ET_o in empirical methods cannot catch the observed values and produce less accurate results than the data-driven methods including ANN, SVM, RF, and MLR in testing period based on R^2. For example, the R^2 of the ET_MLP, ET_RBF, ET_SVM, ET_RF, and ET_MLR models varies from 0.956 to 0.999 (Fig. 7); the R^2 of the ET_HAR, ET_PM, ET_Mak, ET_Thor, and ET_BC models slightly decreases and varies from 0.854 to 0.988 (Fig. 6) in testing period, respectively. These results indicate that types and number of input variables affect better efficiency in the ETo estimation. In equation methods, a radiation-based model (ET_Mak) compared to other empirical methods was not satisfactory, with R^2 value of 0.854. It is seen that the Hargraeves method shows less predictive accuracy when considering the peak values of estimated ET_o values in equation methods in Fig. 6. An evaluation that only base on R^2 may not be sufficient to decide since R^2 is oversensitive to extreme values and insensitive to both additive and proportional differences between observed and model-estimated values (Legates and

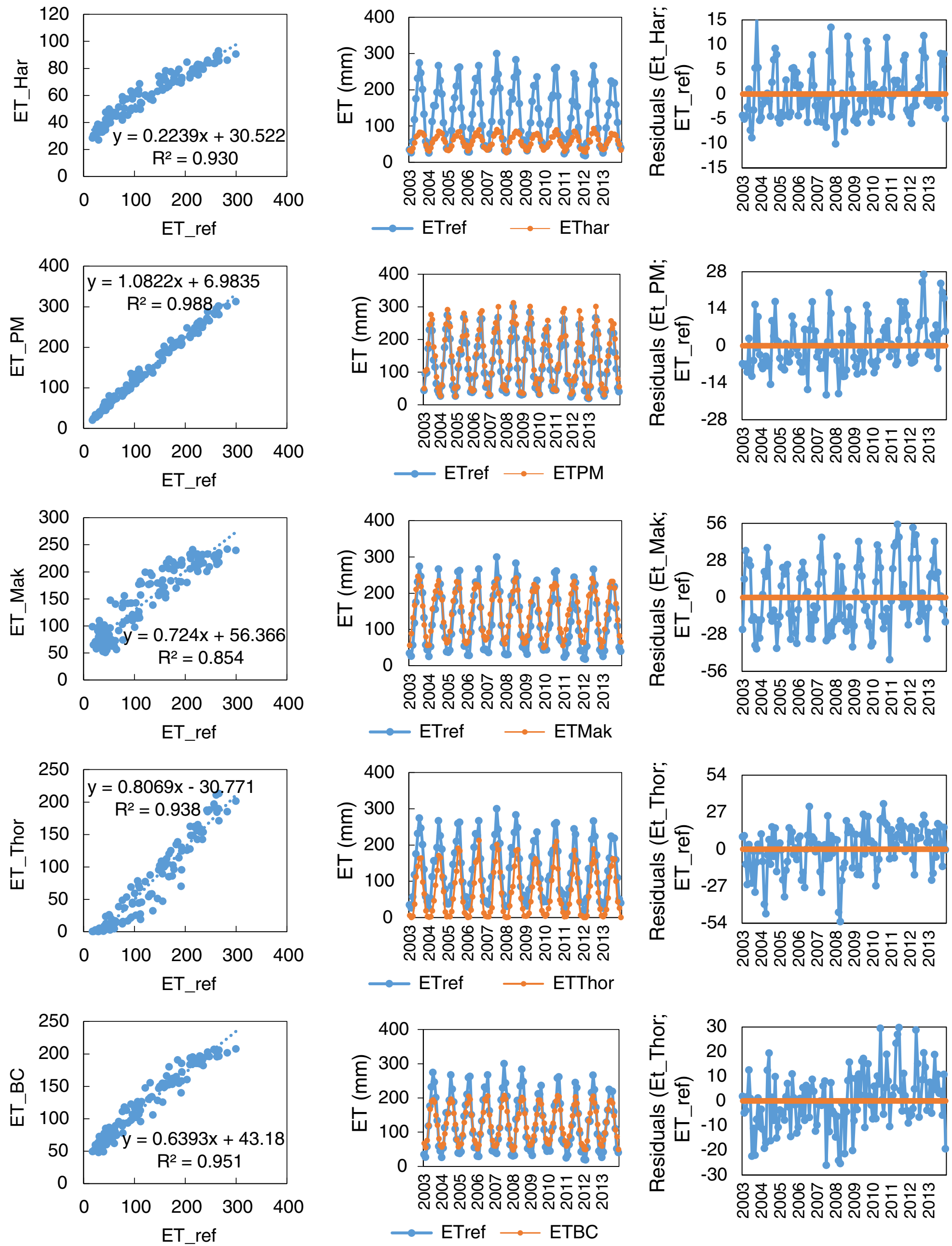

Fig. 6 Relationship between results of empirical methods and reference evapotranspiration with scatter plot (left) (**a**), time series (middle) (**b**), and temporal residual graph (**c**) in testing period

McCabe 1999). The error term calculations based on goodness-of-fit indicators (*d*, RMSE, and MAD) are also suitable for model evaluation than R^2 as they calculate the deviation or error between each pair of observed and estimated values based on the measurement uncertainty. Thus, *d*, RMSE, and MAD were used in addition to evaluate the performance of all techniques and these values are shown in Table 4. Also shown in Figs. 6 and 7 is graphical representation of temporal variation between observed and estimated monthly ET_o values by empirical methods and data-driven models during testing period. Initially, ETo values of cooler months were observed as low and then increased gradually when number of high-temperature months increased in all trends. The record shows marked fluctuations between winter and summer, which implies that changes in climatic conditions that alter evapotranspiration, could easily affect balance and interaction with surface and subsurface water. However, it can perceptibly be seen in Fig. 6 that the Hargraves method did not accurately estimate the evapotranspiration values of the high-temperature months. ET_Har method was not good

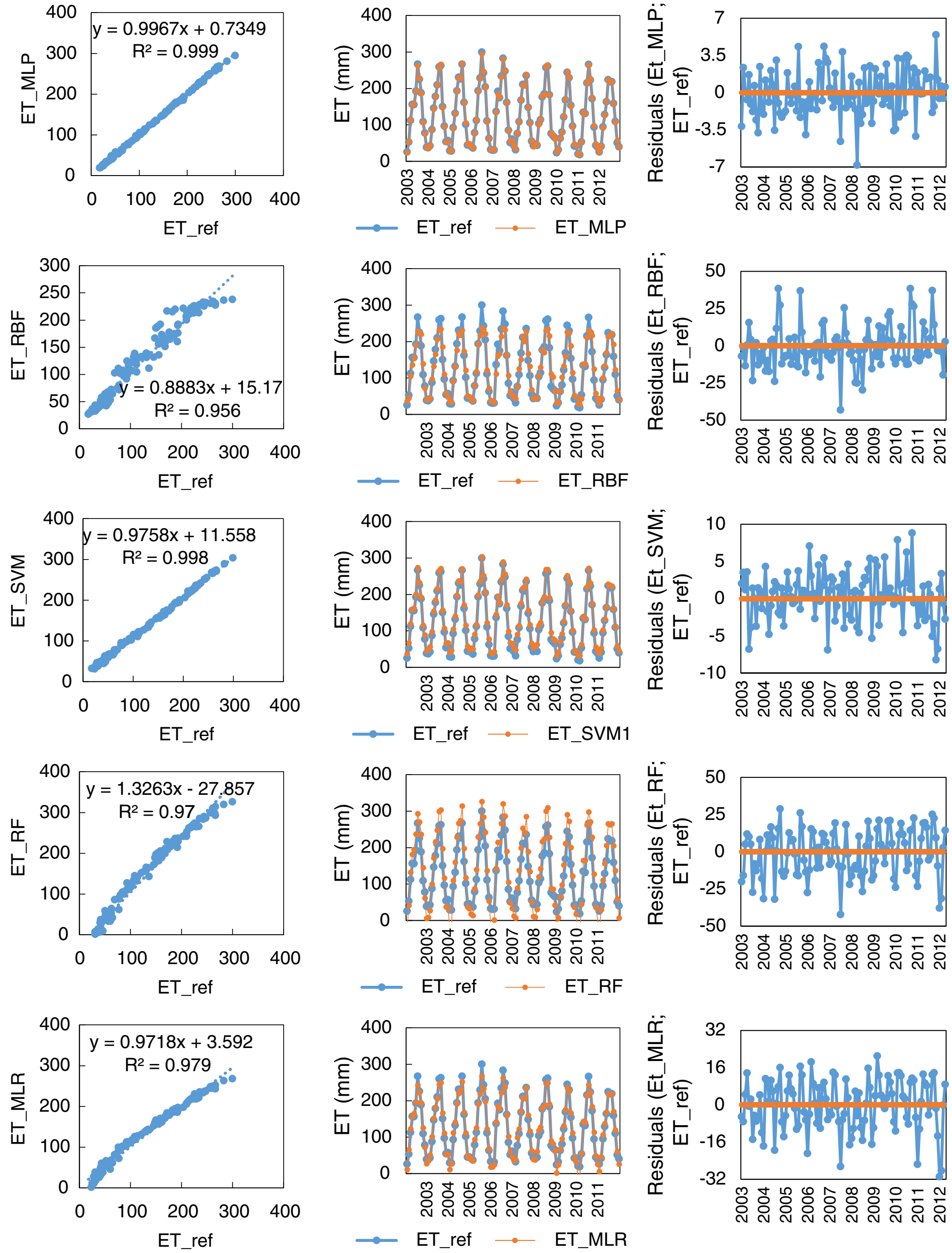

Fig. 7 Relationship between results of data-driven models and reference evapotranspiration with scatter plot (left) (**a**), time series (middle) (**b**), and temporal residual graph (**c**) in testing period

enough in forecasting peak ET_o values. This could be due to the fact that the study area is characterized by a semi-arid continental climate of mild cold winter and hot dry summer, where atmospheric conditions other than temperature, and R_S are more favorable to evaporation and transpiration. Therefore, peak ET_o values inefficiency could be caused by the formulation used in the ET_Har method. Likewise, the scatters of the ET_PM, which base on a combination technique using U and vapor pressure of input parameters, based models are less dispersed, generally overestimating the ET_o with very low errors. Generally, ET_PM models indicate overestimation while values of ET_Mak, ET_Thor, and ET_BC models remain under peak of ET_o values for the Central Anatolian Region. The result of empirical methods, which indicates the superiority of the ET_PM models on the combination-based one, could be considered as a reliable alternative method for ETo estimation among empirical methods. The Penman method uses vapor pressure deficit, actual vapor pressure, and an empirical U function. ET_PM method run underestimates ET_o little. Lee et al. (2004)

reported that this difference derived from empirical wind function used in the equation and the function takes many different forms in literature. The estimation results of ET_o using data-driven methods for the regional average values in Central Anatolian Region have revealed that the SVM and MLP models can achieve reliable estimates. In the study, data-driven models estimated peak ET_o values more accurately. The data preprocessing such as normalization of the data in data-driven models enabled a finest accuracy for capturing peak magnitudes (Demirel et al. 2009). In particular, among data-driven models, the SVM and MLP-based models used in this study were found to have better performances than the RBF, RF, and MLR models (statistical method); they increased the estimation accuracy by up to 98% in regional average dataset. The obtained results were in well agreement with some previous studies (Sayyadi et al. 2009; Rahimikhoob 2010; Traore et al. 2010) that all reported the application of MLP model and their superior accuracy compared to other methods for ET_o estimation in different climates around the Earth. As far as the performance of the ET_SVM model is concerned, the results appeared to be quite satisfactory, and similar results were obtained by Tabari et al. (2012) and Mohammadrezapour et al. (2019) for semi-arid environment. The selection of kernel function type is responsible for performance of SVM model for estimating of ET_o (Seifi and Riahi 2020). As mentioned earlier, a very satisfactory performance has been obtained by using the linear kernel function of SVM model. However, Tabari et al. (2012) had found RBF is the best kernel function among the other functions of SVM models. With regard to the overall performance of the applied all empirical methods and data-driven models in testing period, the hierarchical performance for regional average in Central Anatolian Region follows the order: ET_MLP > ET_SVM > ET_PM > ET_MLR > ET_RF > ET_RBF > ET_BC > ET_Thor > ET_Har > ET_Mak, respectively.

All regression model residuals as a function of observed ET_o of testing period and month of year were also examined in Figs. 6 and 7. These graphs explain the vertical distance between the actual data point and the estimated point on the line. Figures show an example of model residuals versus observed ET_o for all (regional average) dataset. The seasonality can be seen in the residuals at all methods or models, which is more clearly pronounced at some models such as ET_Mak (Fig. 6) and ET_MLR (Fig. 7). The magnitude of seasonality, which increases with increasing estimated ET_o magnitude, is particularly pronounced for notable residuals of ET_RBF and ET_RF models. The reason of these results can be explained by considering that the RH_{avg} is a seasonally dynamic property. In other words, this parameter leads the seasonality more pronounced in residuals. Besides, the seasonal magnitude (the difference between the maximum and minimum value) of seasonally varying RH_{avg} values explained as a percentage is considerably higher than the other parameters. Therefore, the residuals in all models using RH_{avg} as an input parameter were found higher than others that did not use this parameter. For example, the residuals for ET_Har model are not strongly related to ET_o magnitude or month of year. The residuals also show relatively unbiased situations for the models. According to the results of equation methods depicted in Fig. 6, the ET_Har, ET_PM, and ET_Mak methods tended to overestimate observed ET_o while the ET_Thor method tended to underestimate ET_o. In the ET_BC method, residuals generally showed a balanced distribution by years. As can be seen from Fig. 8, the all methods were found to be mostly positive residuals after 2010 and 2011 (Fig. 8a, b). The graphs of cumulative average residuals clearly depict cumulative underestimate and overestimate estimations (Fig 8c, d). Before 2009, the ET_MLP model, which is the best data-driven model, estimated cumulative overestimate values in 2006, while ET_PM, which is the best empirical method, tended to overestimate after 2011. As has been shown in Fig 8c, d, cumulative residual plots may display a tendency to overestimate and underestimate with relation to control of wet and cold biases in considered years of the study area.

The results of the second step (elevation groups)

The statistics given in Table 5 illustrate the difference between some selected data characteristics in the 4 different elevation groups. T_{min} of all group except for G4 in whole data shows a significantly greater level of dispersion around the mean compared with other CV of all groups. The parameter with the highest CV value in the whole data of G4 is T_{max}. It is seen in Table 5 that the CV values of other parameters are also close to the value in T_{max} of G4. The precipitation has higher skewed distribution in all groups, just as in complete dataset shown in Table 5. Another important statistical characteristic of the selected climate data is the highest R^2 found between the ET_o and T_{max} in training period of all four groups in the ranges 0.84 and 0.86 (p<0.05) and the lowest R^2 between the ET_o and U_2 in training period of all groups ranges within an interval of 0.00 and 0.04 (p>0.05)

Test results of the six different optimal data-driven models for each station are provided in training period (Table 6) and testing period (Table 7) using long-term monthly data of elevation-based groups. In training/testing period, it is clear from the Tables 6 and 7 that the RMSE values of empirical methods in training period are considerably higher than the RMSE results of testing period. For the ET_RBF, ET_RF, and ET_MLR models, the maximum RMSE (15.12, 29.17, and 12.93 mm/month) values were found for the G1, respectively. The maximum R^2 of all models in G1 were found in ET_MLR (R^2= 0.998, p<0.05) and ET_SVM (R^2= 0.998, p<0.05). These models presented the highest d value (1.00; 1.00) and NSE equal to 1.00 and 0.99, respectively. For the ET_Thor,

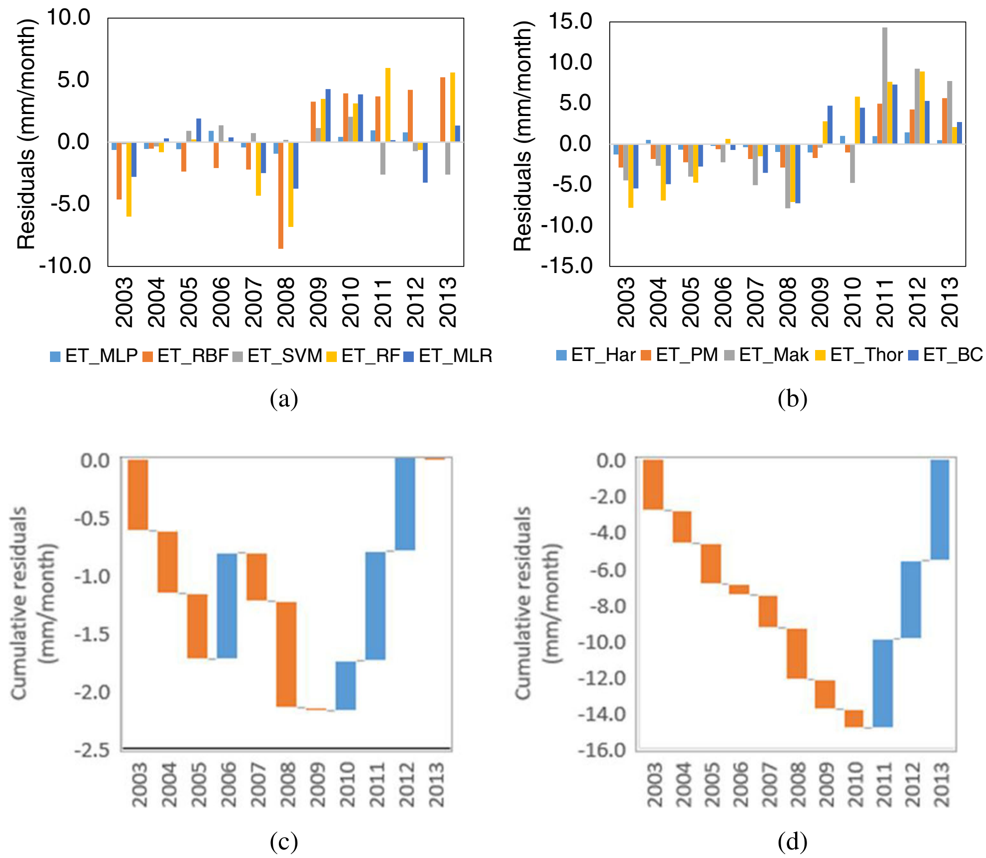

Fig. 8 Residuals and cumulative residuals graphs for testing period. **a** Annual average residuals of empirical methods. **b** Annual average residuals of data-driven models. **c** Cumulative residuals graph of ET_MLP. **d** Cumulative residuals graph of ET_PM

however, the maximum RMSE value was found to be 3.50 mm/month in the G2. It can be clearly seen in Table 7, the G2 group shows already better performance for all performance criteria than the other groups in testing period. Therefore, it has been determined that the models used for ET_o estimation in the Central Anatolian Region can be used most effectively at an altitude between 850 and 1000 m. The values of performance are similar for all elevation groups in training period (Table 7). Once again, the ET_Har method performed the worst in G3, due to significant underestimations, with a RMSE value of 78.95 mm/month, NSE value of −0.11. The values *d*, NSE, and R^2 shown in Table 7 indicate that the ET_MLP was the best simple method for estimating ET_o in G4 ($R^2 = 0.75$). It is clearly seen from Table 6 that the accuracy of the ET_MLP is generally better than the other models in ET_o estimation. In four groups, the ET_MLP model has the best accuracy. The ET_SVM and ET_PM models respectively also performed well in all groups while the ET_Mak yielded the worst estimation in all groups in testing period (Table 7). Estimated ET_o values by models are lower than the observed ET_o values since 2009.

PBIAS (%) indicates the model performance with overestimate (PBIAS < 0) or underestimate (PBIAS > 0) of ET_o, and values of the PBIAS nearer to 0 suggest a model or method with more predictive skill. Safeeq and Fares (2012) emphasize that value of PBIAS more than 15% and less than 25% was considered an indicator of average performance; however, a value between 10% and 15% indicates a good performance, and a value less than 10% indicates a very good performance. As it can be seen, model efficiency using PBIAS is higher for data-driven models as compared to the use of empirical methods in both training and testing period of each group (Fig. 9). The model performance of ET_MLP and ET_RBF in the entire training and testing period is considered "very good" on the basis of the PBIAS values vary between 0 and −2%, respectively (Fig. 9). Celestin et al. (2020) found that the World Meteorological Organization (WMO) and the Mahringer (MAHR) models performed well with monthly data compared to the PM FAO-56 model with PBIAS of −2.5% and −2.6% after the calibration period, respectively. With regard to PBIAS, The ET_Thor method provided the highest PBIAS values in all group in both periods. From the MAD, *d*, and NSE perspective, ET_Thor shows acceptable performance in both training and testing periods (Tables 4, 5, and 6). However, the maximum RMSE and MAD values are exhibited by ET_Har method in non-group in both periods

Table 5 Statistical parameter of climatic data and ET_o in four elevation groups

Group	Period	Stat. Para.	ET_o	T_{max}	T_{min}	T_{avg}	P	U_2	RH_{max}	RH_{min}	RH_{avg}	R_s	Group
G1	**All**	**Mean**	128.05	24.08	2.91	11.52	42.34	2.74	84.86	42.82	65.68	17.39	G3
		Std. De.	85.70	9.81	9.91	10.16	34.00	0.48	11.44	14.67	13.81	7.57	
		SK	0.55	−0.24	0.23	0.34	1.02	0.42	−0.82	0.58	−0.01	−0.06	
		Min.	7.40	0.64	−27.28	−13.06	0.00	1.34	47.29	10.49	34.67	4.46	
		Max.	424.37	44.32	29.20	35.99	220.42	4.92	99.74	89.50	95.89	31.49	
		CV	0.67	0.41	3.41	0.88	0.80	0.17	0.13	0.34	0.21	0.44	
		R^2	-	0.84*	0.71	0.36	0.25	0.03	0.68	0.59	0.79*	0.80*	
	Training	**Mean**	124.83	24.00	2.48	11.72	40.23	2.66	85.08	43.29	66.07	17.47	
		Std. De.	83.19	9.87	9.73	10.20	33.01	0.46	11.19	14.51	13.59	7.59	
		SK	0.50	−0.24	0.23	0.32	1.11	0.62	−0.80	0.55	−0.04	−0.06	
		Min.	7.40	0.64	−27.28	−11.71	0.00	1.34	47.29	10.49	35.11	4.57	
		Max.	405.55	44.32	29.10	35.99	220.42	4.92	99.72	87.99	95.89	31.49	
		CV	0.67	0.41	3.92	0.87	0.82	0.17	0.13	0.34	0.21	0.43	
		R^2	-	0.85*	0.70	0.37	0.23	0.04	0.64	0.58	0.77*	0.82*	
	Testing	**Mean**	136.21	24.30	3.98	11.03	42.34	2.74	84.86	42.82	65.68	17.39	
		Std. De.	91.21	9.65	10.28	10.04	34.00	0.48	11.44	14.67	13.81	7.57	
		SK	0.59	−0.25	0.20	0.37	1.02	0.42	−0.82	0.58	−0.01	−0.06	
		Min.	9.43	1.91	−24.00	-13.06	0.00	1.34	47.29	10.49	34.67	4.46	
		Max.	424.37	41.43	29.20	34.86	220.42	4.92	99.74	89.50	95.89	31.49	
		CV	0.67	0.40	2.58	0.91	0.80	0.17	0.13	0.34	0.21	0.44	
		R^2	-	0.83*	0.71	0.35	0.32	0.02	0.78	0.62	0.84*	0.80*	
G2	**All**	**Mean**	131.43	23.68	3.03	11.94	39.97	2.82	80.82	41.62	60.38	18.34	G4
		Std. De.	84.13	9.74	11.25	10.65	34.32	0.54	14.55	17.78	16.89	7.53	
		SK	0.56	−0.22	0.43	0.20	1.12	0.59	−0.85	0.75	−0.14	−0.07	
		Min.	7.59	−1.48	−29.04	−22.23	0.00	1.33	27.03	5.06	10.00	4.43	
		Max.	445.59	44.14	41.11	36.20	246.74	5.19	100.00	99.00	96.04	32.37	
		CV	0.64	0.41	3.71	0.89	0.86	0.19	0.18	0.43	0.28	0.41	
		R^2	-	0.83*	0.53	0.38	0.23	0.00	0.56	0.49	0.64	0.80*	
	Training	**Mean**	117.89	23.62	4.97	10.02	38.52	2.87	80.07	42.54	59.64	18.47	
		Std. De.	71.32	9.76	11.71	10.45	34.36	0.55	15.03	18.54	17.30	7.56	
		SK	0.26	−0.22	0.35	0.36	1.19	0.53	−0.83	0.73	−0.17	−0.07	
		Min.	7.59	−1.48	−28.74	−22.23	0.00	1.53	27.03	5.06	10.00	4.43	
		Max.	300.20	43.06	41.11	36.20	246.74	5.19	100.00	99.00	96.04	32.37	
		CV	0.60	0.41	2.36	1.04	0.89	0.19	0.19	0.44	0.29	0.41	
		R^2	-	0.84*	0.51	0.37	0.23	0.00	0.52	0.48	0.59	0.79*	
	Testing	**Mean**	136.86	23.84	−1.81	16.73	43.54	2.69	82.69	39.31	62.21	18.01	
		Std. De.	88.16	9.69	8.15	9.59	33.91	0.49	13.10	15.48	15.68	7.44	
		SK	0.56	−0.21	−0.43	−0.05	0.98	0.71	−0.80	0.64	0.03	−0.08	
		Min.	8.45	0.92	−29.04	−3.03	0.00	1.33	40.03	11.31	29.38	4.53	
		Max.	445.59	44.14	15.38	35.87	219.34	4.93	99.74	89.89	95.90	31.91	
		CV	0.64	0.41	−4.50	0.57	0.78	0.18	0.16	0.39	0.25	0.41	
		R^2	-	0.89*	0.71	0.88*	0.22	0.00	0.67	0.65	0.82*	0.88*	

Table 5 (continued)

Group	Period	Stat. Para.	ET_o	T_{max}	T_{min}	T_{avg}	P	U_2	RH_{max}	RH_{min}	RH_{avg}	R_s
Group	Period	Stat. Para.	ET_o	T_{max}	T_{min}	T_{avg}	P	U_2	RH_{max}	RH_{min}	RH_{avg}	R_s
G1	**All**	**Mean**	119.81	21.44	−2.39	13.34	57.09	3.19	85.26	40.66	64.80	18.64
		Std. De.	76.50	10.20	9.34	10.27	47.91	0.64	12.19	17.09	16.13	7.58
		SK	0.37	−0.22	−0.23	0.04	1.56	0.35	−0.74	0.72	0.06	−0.07
		Min.	7.77	−0.49	−32.90	−13.5	0.00	1.48	44.40	9.24	30.16	4.29
		Max.	398.90	40.71	26.82	33.66	550.07	5.58	100.00	90.53	96.56	32.49
		CV	0.64	0.48	−3.90	0.77	0.84	0.20	0.14	0.42	0.25	0.41
		R^2	-	0.86*	0.70	0.69	0.25	0.00	0.67	0.59	0.76*	0.83*
	Training	**Mean**	121.03	21.44	−2.39	13.34	57.09	3.19	85.26	40.66	64.80	18.64
		Std. De.	77.92	10.20	9.34	10.27	47.91	0.64	12.19	17.09	16.13	7.58
		SK	0.41	−0.22	−0.23	0.04	1.56	0.35	−0.74	0.72	0.06	−0.07
		Min.	9.40	−0.49	−32.90	−13.5	0.00	1.48	44.40	9.24	30.16	4.29
		Max.	398.90	40.71	26.82	33.66	550.07	5.58	100.00	90.53	96.56	32.49
		CV	0.64	0.48	−3.90	0.77	0.84	0.20	0.14	0.42	0.25	0.41
		R^2	-	0.86	0.71	0.63	0.25	0.01	0.70	0.59	0.77	0.82
	Testing	**Mean**	117.41	23.06	−2.47	15.80	53.82	3.07	85.36	41.99	65.98	17.49
		Std. De.	73.52	10.11	8.50	9.73	42.66	0.62	11.16	16.20	14.32	7.49
		SK	0.28	−0.29	−0.61	−0.08	0.99	0.31	−0.92	0.63	−0.01	−0.03
		Min.	7.77	−0.10	−34.38	−4.20	0.00	1.59	41.62	12.46	31.38	4.60
		Max.	294.05	42.98	13.93	36.20	265.37	5.41	99.70	87.78	95.81	31.75
		CV	0.63	0.44	−3.44	0.62	0.79	0.20	0.13	0.39	0.22	0.43
		R^2	-	0.88*	0.70	0.88*	0.24	0.00	0.62	0.61	0.73*	0.85*
G2	**All**	**Mean**	121.93	20.49	−1.12	10.13	59.09	3.30	85.34	40.86	65.05	18.61
		Std. De.	79.94	22.21	−0.82	8.45	52.90	3.31	88.24	35.97	63.17	19.25
		SK	0.41	10.44	10.51	10.45	49.48	0.66	12.20	17.88	16.67	7.59
		Min.	9.42	−0.24	−0.02	0.25	1.44	0.10	−0.69	0.74	0.13	−0.07
		Max.	401.47	−0.46	−30.41	−16.5	0.00	1.46	42.40	5.57	30.14	3.86
		CV	0.66	1.08	0.73	0.83	0.89	1.01	1.03	1.13	0.97	1.03
		R^2	-	0.86*	0.62	0.52	0.24	0.00	0.71	0.58	0.77*	0.80*
	Training	**Mean**	123.03	0.51	−9.41	1.03	0.84	0.20	0.14	0.44	0.26	0.41
		Std. De.	82.88	10.56	11.19	9.98	51.54	0.64	12.18	18.51	16.89	7.63
		SK	0.44	−0.25	−0.05	0.45	1.46	0.05	−0.71	0.66	0.09	−0.08
		Min.	9.42	−0.46	−30.19	−16.5	0.00	1.74	42.40	5.57	30.68	3.86
		Max.	401.47	39.17	29.00	31.85	518.68	5.69	99.83	91.13	96.75	33.30
		CV	0.67	0.53	29.95	1.28	0.82	0.19	0.14	0.44	0.25	0.41

Table 5 (continued)

Group	Period	Stat. Para.	ET_o	T_{max}	T_{min}	T_{avg}	P	U_2	RH_{max}	RH_{min}	RH_{avg}	R_s
		R^2	-	0.86*	0.63	0.47	0.26	0.00	0.73	0.60	0.78*	0.79*
	Testing	**Mean**	119.83	21.50	−4.07	14.82	51.92	3.11	83.41	38.10	62.32	18.88
		Std. De.	73.68	10.13	8.22	9.78	44.28	0.64	12.00	16.20	15.87	7.49
		SK	0.31	−0.22	−0.76	−0.05	1.28	0.23	−0.72	0.88	0.18	−0.06
		Min.	11.26	0.96	−30.41	−4.19	0.00	1.46	46.83	10.35	30.14	5.68
		Max.	294.52	39.98	8.83	33.41	320.58	5.38	99.65	90.35	95.19	32.53
		CV	0.61	0.47	−2.02	0.66	0.85	0.21	0.14	0.43	0.25	0.40
		R^2	-	0.89*	0.68	0.88	0.19	0.00	0.70	0.55	0.79*	0.85*

Std., standard deviation; *SK*, skewness; *CV*, coefficient of variation; R^2 , coefficients of determination with ET_o; *$p<0.05$

Table 6 Comparison of performances of all techniques for ET_o estimation of four elevation groups in training period

Method	G1					G2					G3					G4				
	R^2	MAD	RMSE	*d*	NSE	R^2	MAD	RMSE	*d*	NSE	R^2	MAD	RMSE	*d*	NSE	R^2	MAD	RMSE	*d*	NSE
ET_Har	0.912	76.62	100.81	0.60	−0.52	0.914	74.42	96.95	0.62	−0.48	0.933	59.33	78.95	0.72	−0.10	0.919	63.22	84.57	0.70	−0.15
ET_PM	0.991	12.42	15.28	0.99	0.97	0.988	19.26	23.47	0.99	0.91	0.989	11.05	14.50	0.99	0.96	0.988	8.89	12.51	1.00	0.97
ET_Mak	0.876	28.46	34.01	0.97	0.83	0.879	28.48	34.22	0.98	0.82	0.869	31.49	38.26	0.97	0.74	0.851	34.09	41.63	0.96	0.72
ET_Thor	0.936	63.36	70.19	0.76	0.26	0.934	63.38	68.60	0.78	0.26	0.946	48.20	51.90	0.87	0.52	0.943	54.31	59.07	0.83	0.44
ET_BC	0.950	31.60	38.64	0.96	0.78	0.956	28.39	35.00	0.97	0.81	0.963	22.68	26.06	0.98	0.88	0.958	24.84	29.55	0.98	0.86
ET_MLP	0.999	2.34	3.16	1.00	1.00	0.999	2.01	2.53	1.00	1.00	1.00	1.93	2.44	1.00	1.00	1.00	4.00	5.73	1.00	0.99
ET_RBF	0.946	13.91	19.13	0.99	0.95	0.964	12.33	16.75	0.99	0.96	0.94	13.73	18.63	0.99	0.94	0.94	14.95	19.13	0.99	0.94
ET_SVM	0.998	8.93	9.99	1.00	0.99	0.999	6.31	7.69	1.00	0.99	1.00	14.79	15.22	0.99	0.96	1.00	10.95	12.11	1.00	0.98
ET_RF	0.964	29.24	32.70	0.98	0.84	0.984	25.79	28.32	0.98	0.87	0.97	35.67	39.49	0.96	0.72	0.97	35.48	38.92	0.96	0.74
ET_MLR	0.969	12.08	14.44	0.99	0.97	0.986	8.34	10.25	1.00	0.98	0.98	9.00	11.09	1.00	0.98	0.97	10.93	13.45	1.00	0.97

Table 7 Comparison of performances of all techniques for ET_o estimation of four elevation groups in testing period

Method	G1					G2					G3					G4				
	R^2	MAD	RMSE	*d*	NSE	R^2	MAD	RMSE	*d*	NSE	R^2	MAD	RMSE	*d*	NSE	R^2	MAD	RMSE	*d*	NSE
ET_Har	0.911	0.06	0.62	0.59	−0.54	0.924	0.04	0.49	0.62	−0.48	0.937	58.91	78.25	0.72	−0.11	0.924	0.05	0.55	0.70	−0.15
ET_PM	0.990	0.12	1.28	0.99	0.95	0.988	0.14	1.50	0.99	0.90	0.987	13.74	17.01	0.99	0.95	0.986	0.13	1.40	0.99	0.96
ET_Mak	0.866	0.22	2.40	0.97	0.81	0.856	0.22	2.44	0.97	0.78	0.847	34.18	41.55	0.96	0.69	0.829	0.24	2.60	0.96	0.67
ET_Thor	0.935	0.29	3.18	0.79	0.34	0.933	0.32	3.50	0.81	0.34	0.939	45.15	48.86	0.89	0.57	0.930	0.34	3.76	0.86	0.50
ET_BC	0.945	0.09	1.03	0.96	0.78	0.947	0.07	0.80	0.97	0.81	0.953	22.45	25.89	0.98	0.88	0.947	0.05	0.54	0.98	0.87
ET_MLP	0.998	2.57	3.52	1.00	1.00	0.999	0.00	0.04	1.00	1.00	0.999	2.14	2.76	1.00	1.00	0.998	4.00	5.71	1.00	0.99
ET_RBF	0.940	15.12	20.91	0.99	0.94	0.960	0.04	0.39	0.99	0.95	0.943	13.20	18.12	0.99	0.94	0.938	14.92	19.13	0.99	0.94
ET_SVM	0.997	7.73	9.17	1.00	0.99	0.999	0.04	0.45	1.00	0.99	0.997	13.68	14.30	0.99	0.96	0.996	10.95	12.09	1.00	0.98
ET_RF	0.963	29.17	32.60	0.98	0.85	0.982	0.26	2.83	0.98	0.87	0.973	36.05	39.63	0.96	0.71	0.971	35.47	38.91	0.96	0.74
ET_MLR	0.967	12.93	15.41	0.99	0.97	0.984	0.08	0.89	1.00	0.98	0.975	9.81	12.35	1.00	0.97	0.974	10.92	13.39	1.00	0.97

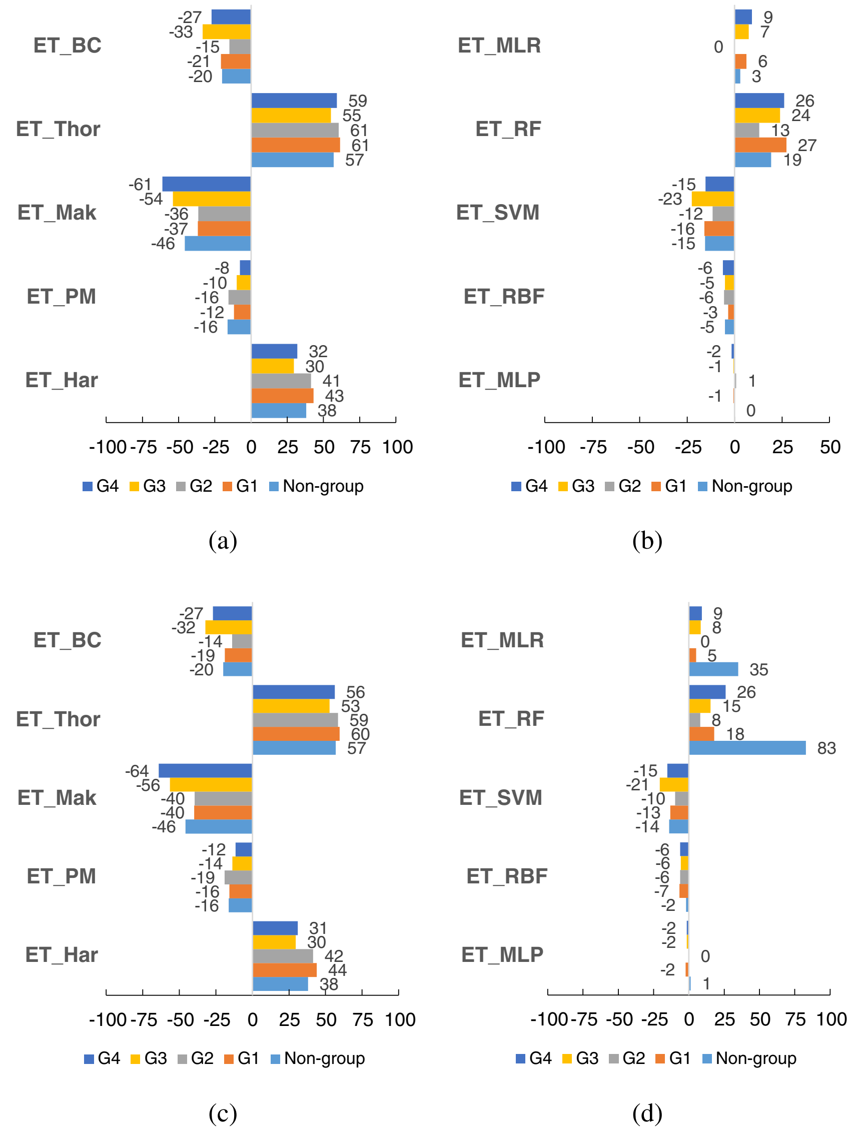

Fig. 9 PBIAS (%) graphs. **a** Empirical methods in training period. **b** Data-driven models in training periods. **c** Empirical methods in testing period. **d** Data-driven models in testing period

(Table 4). ET_BC, ET_Mak, ET_PM, ET_SVM, and ET_RBF are underestimating the reference evapotranspiration for all elevation groups and regional average (non-group) in training and testing period. ET_MLP in regional average and ET_MLR in G2 group are equally suitable with 0% PBIAS, and they could also be used satisfactorily to estimate reference evapotranspiration for the study area.

For further analysis, the developed predictive models of reference evapotranspiration are examined by Taylor diagram (Taylor 2001). Taylor diagram classifies the results of methods or models by using standard deviation and the R^2 of observed and simulated data. The radial coordinate shows the value of standard deviation; the concentric semi-circles represent the magnitude of standard deviation, and the angular coordinate indicates the values of R^2. Estimated ET_o by different methods and models that run it with observed ET_o will lie nearest to the point marked "reference" on the *x*-axis. Figures 10 and 11 display the standard deviation and R^2 (with observed ET_o) for the results of different equation methods and models calculated from the various inputs, respectively. The data-driven models (Fig. 11), in general, are produced more accurately than empirical methods (Fig. 10), with the latter having a relatively low *d* and NSE values (Table 7). On the basis of the results shown in the Taylor diagram, four elevation groups for the variables are determined by concentric analysis, which falls in the range of 78–86, with respect to ideal model points of both empirical methods and data-driven models in testing period. The ET_PM for the Taylor diagram is composed of the models that perform highly for estimated ET_o (Fig. 10). Taylor diagram analysis reveals that ET_MLP

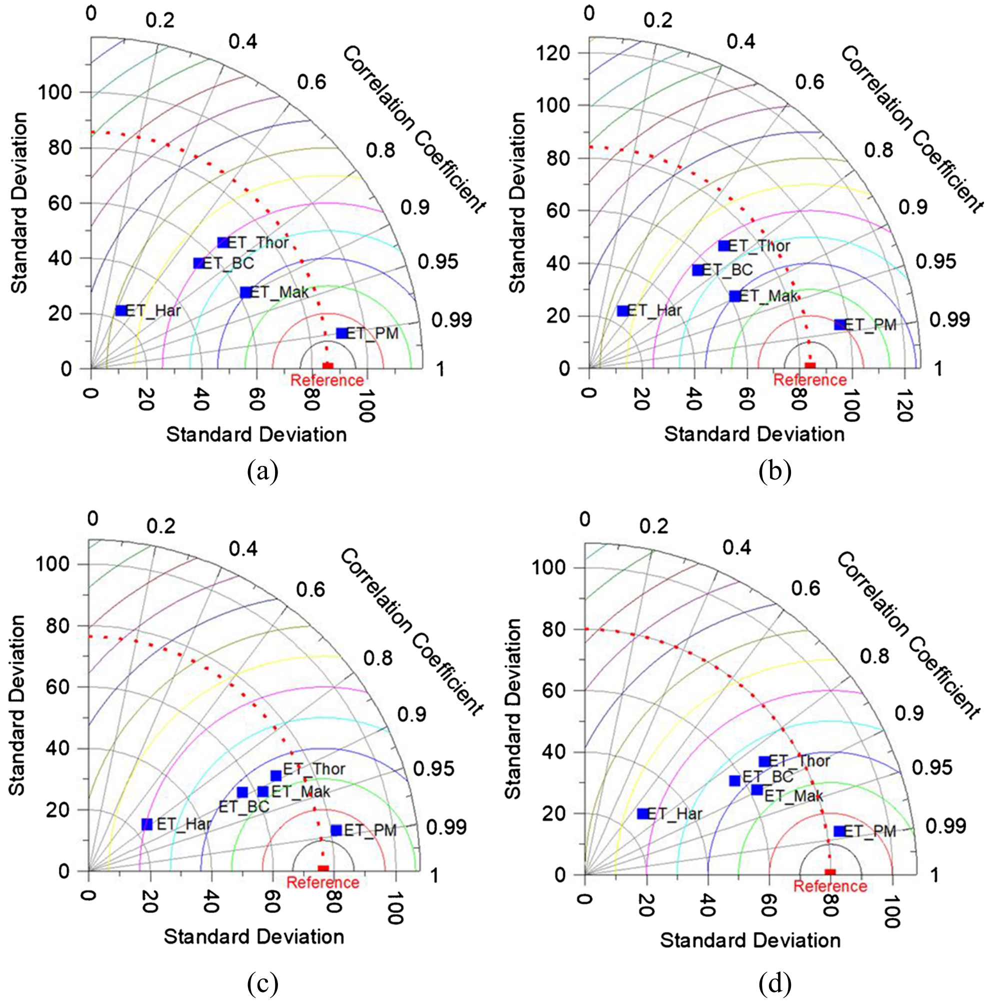

Fig. 10 Taylor diagram of the correlation coefficient (*r*), the centered root mean square difference, and standard deviation between estimated ETo of different groups (**a** G1, **b** G2, **c** G3, and **d** G4) by using empirical methods and ET_o values in testing period

has the R^2 (range between 0.997 and 0.999), lowest standard deviation (range between 74 and 80 mm/month), and smallest RMSE (range between 0.04 and 5.71, and captures observations better than all data-driven models in all group.

Conclusion

The performances of ET_o, developed based on two main approaches (regional average and elevation group) to the estimated ET_o produced by the five different empirical methods (ET_Har, ET_PM, ET_Mak, ET_Thor, and ET_BC) and the six different data-driven models (ET_MLP, ET_RBF, ET_SVM, ET_RF, and ET_MLR), were assessed for the Central Anatolian Region of Turkey. The performances of the empirical methods and data-driven models are reported to provide evidence for suitable techniques for estimating ET_o values.

Monthly selected climatic data variables of 45 meteorological stations, over a period of 35 years (1979–2013) were used in this study. This study conducted by two stages of data preparation. In the first stage, the average of all parameter values obtained from 45 meteorology stations was evaluated. In the second step, the data set was divided into 4 elevation groups. Correlation of the parameters with ET_o was taken into account in the selection of input parameters. Climatic variables considered in all stations showed that ET_o is strongly and positively correlated with T_{max}, RH_{avg}, and R_S, with a R^2 equal to 0.84, 0.68, and 0.79, respectively. It has been found that these three variables can be effective in modeling evapotranspiration in a semi-arid region. Therefore, these variables should be included in long-term monitoring programs, especially in agricultural planning and water resources management in semi-arid regions due to evapotranspiration is an essential factor that causes a great change in the water budget, especially in fragile semi-arid ecosystems.

Based on the performance of a grouping result evaluations, it is found that the MLP and SVM models in G2 (850–1100 m) can be employed successfully in modeling the monthly mean ET_o, because both approaches yield better estimates

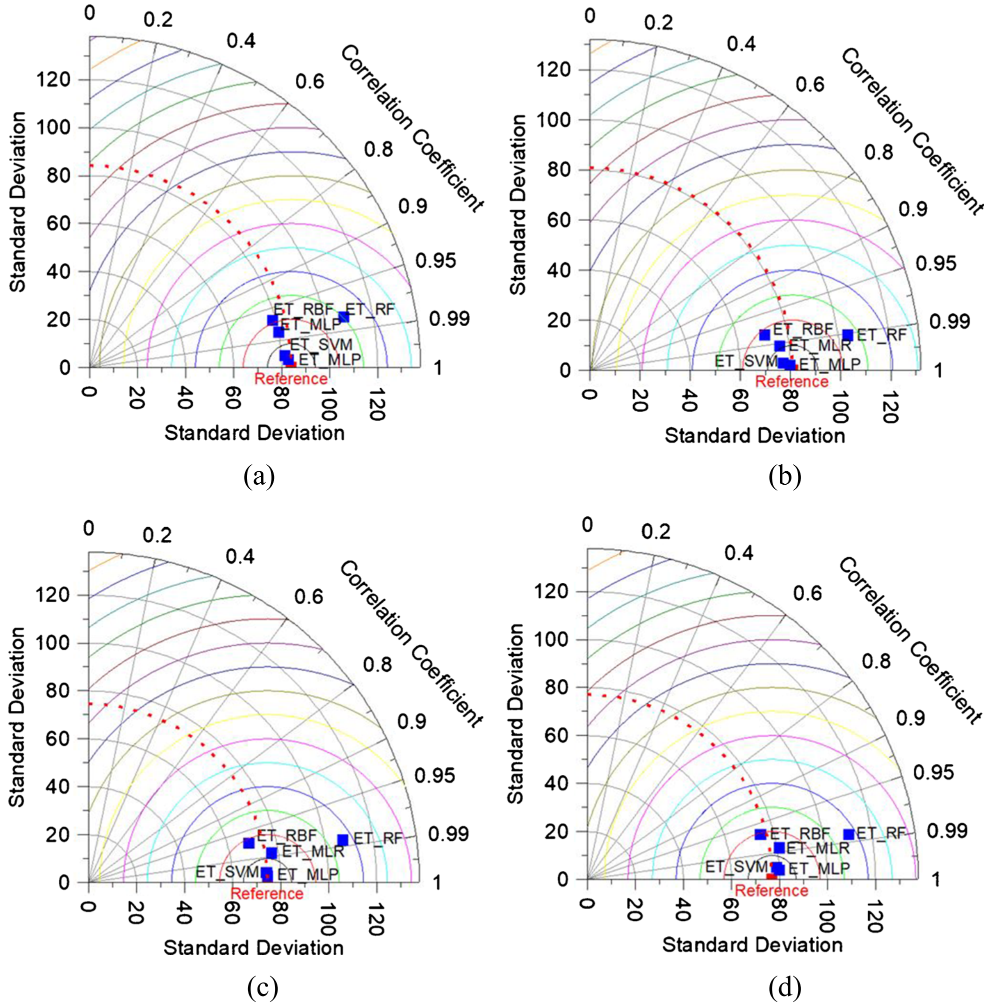

Fig. 11 Taylor diagram of the correlation coefficient (*r*), the centered root mean square difference (RMSD), and standard deviation (STD) between estimated ETo of different groups (**a** G1, **b**:G2, **c**:G3, and **d** G4) by using data-driven methods and ET_o values in testing period

with high value of R^2, compared to other empirical methods and yet MLP being slightly more successful than SVM. Therefore, this research suggests that a reference evapotranspiration in semi-arid region can be modeled using only a few input parameters with the help of a simple but effective data-driven models. We find that Penman method has achieved the highest accuracy in terms of all performance criteria among the empirical methods. The Penman method is suitable for estimating the reference evapotranspiration, and it can be used reliably in semi-arid areas.

From this study, it can be concluded that in case a single climatic variable such as *U* or sunshine duration is missing, the alternative models can be used for computing accurate PM FAO-56 model semi-arid environments. The results are encouraging and suggest an easy-to-use and accurate estimate to assess reference evapotranspiration model as an alternative to empirical approaches, because the advantage of the soft computational methods lies in the possibility of having improvements in the performance criteria by modifying the important tunable parameters.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

Alexandris S, Stricevic R, Petkovic S (2008) Comparative analysis of reference evapotranspiration from the surface of rainfed grass in central Serbia, calculated by six empirical methods against the Penman-Monteith formula. Euro Water 21:17–28

Allen RG, Pereira LS, Raes D, Smith M (1998) In: FAO (ed) Crop evapotranspiration —guidelines for computing crop water requirements- FAO Irrigation and drainage paper 56. Food and Agriculture Organization, Rome, pp 1–326

Blaney HF, Criddle WD (1950) Determining water requirement in irrigated areas from climatological data. Soil Conservation Service Technical Publication No. 96, US Department of Agriculture, Washington DC

Breiman L (2001) Random forests. Mach Learn 45:5–32

Celestin S, Qi F, Li R, Yu T, Cheng W (2020) Evaluation of 32 simple equations against the Penman–Monteith method to estimate the reference evapotranspiration in the Hexi Corridor, Northwest China. Water 12:2772

Chia MY, Huang YF, Koo CH (2020) Support vector machine enhanced empirical reference evapotranspiration estimation with limited meteorological parameters. Comput Electron Agric 175:105577

Currie DJ (1991) Energy and large-scale patterns of animal- and plant-species richness. Am Nat 137:27–49

da Silva HJ, dos Santos MS, Junior JBC, Spyrides MH (2016) Modeling of reference evapotranspiration by multiple linear regression. Journal of Hyperspectral Remote Sensing 6:44–58

Demirel MC, Venancio A, Kahya E (2009) Flow forecast by SWAT model and ANN in Pracana basin, Portugal. Adv Eng Softw 40: 467–473

Dile YT, Srinivasan R (2014) Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: an application in the Blue Nile River Basin. J Am Water Resour Assoc 50:1226–1241

Douglas EM, Jacobs JM, Sumner DM, Ray RL (2009) A comparison of models for estimating potential evapotranspiration for Florida land cover types. J Hydrol 373:366–376

Duane WJ, Pepin NC, Losleben ML, Hardy DR (2008) General characteristics of temperature and humidity variability on Kilimanjaro, Tanzania. Arct Antarct Alp Res 40:323–334

Efthimiou N, Alexandris S, Karavitis C, Mamassis N (2013) Comparative analysis of reference evapotranspiration estimation between various methods and the FAO56 Penman-Monteith procedure. Euro Water 42:19–34

Feng Y, Cui N, Gong D, Zhang Q, Zhao L (2017) Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. Agr Water Manag 193:163–173

Fisher DK, Pringle HC III (2013) Evaluation of alternative methods for estimating reference evapotranspiration. Agric Sci 4:51–60

Fuka DR, Walter MT, MacAlister C, Degaetano AT, Steenhuis TS, Easton ZM (2014) Using the climate forecast system reanalysis as weather input data for watershed models. Hydrol Process 28:5613–5623

Hadria R, Benabdelouhab T, Lionboui H, Salhi A (2021) Comparative assessment of different reference evapotranspiration models towards a fit calibration for arid and semi-arid areas. J Arid Environ 184: 104318

Hargreaves GH, Samani ZA (1985) Reference crop evapotranspiration from temperature. Appl Eng Agric 1:96–99

Hashemi M, Sepaskhah AR (2020) Evaluation of artificial neural network and Penman–Monteith equation for the prediction of barley standard evapotranspiration in a semi-arid region. Theor Appl Climatol 139: 275–285

Islam S, Abdullah RAB, Badruddin IA, Algahtani A, Shahid S (2020) Calibration and validation of reference evapotranspiration models in semi-arid conditions. Appl Ecol Environ Res 18:1361–1386

Issaka AI, Paek J, Abdella K, Pollanen M, Huda AKS, Kaitibie S, Goktepe I, Haq MM, Moustafa AT (2017) Analysis and calibration of empirical relationships for estimating evapotranspiration in Qatar: case study. J Irrig Drain Eng 143:05016013

Käfer PS, da Rocha NS, Diaz LR, Kaiser EA, Santos DC, Veeck GP, Robérti DR, Rolim SBA, de Oliveira GG (2020) Artificial neural networks model based on remote sensing to retrieve evapotranspiration over the Brazilian Pampa. J Appl Remote Sens 14:038504

Karimaldini F, Teang Shui L, Ahmed Mohamed T, Abdollahi M, Khalili N (2011) Daily evapotranspiration modeling from limited weather data by using neuro-fuzzy computing technique. J Irrig Drain Eng 138:21–34

Kisi O (2014) Comparison of different empirical methods for estimating daily reference evapotranspiration in Mediterranean climate. J Irrig Drain Eng 140:04013002

Lang D, Zheng J, Shi J, Liao F, Ma X, Wang W, Chen X, Zhang M (2017) A comparative study of potential evapotranspiration estimation by eight methods with FAO Penman–Monteith method in southwestern China. Water 9:1–18

Lee TS, Najim MMM, Aminul MH (2004) Estimating evapotranspiration of irrigated rice at the West Coast of the Peninsular of Malaysia. J Appl Irrig Sci 39:103–117

Legates DR, McCabe GJ (1999) Evaluating the use of “goodness-of-fit” measures inhydrologic and hydroclimatic model validation. Water Resour Res 35:233–241

Li S, Kang S, Zhang L, Zhang J, Du T, Tong L, Ding R (2016) Evaluation of six potential evapotranspiration models for estimating crop potential and actual evapotranspiration in arid regions. J Hydrol 543: 450–461

Lu J, Sun G, McNulty SG, Amatya DM (2005) A comparison of six potential evapotranspiration methods for regional use in the Southeastern United States 1. J Am Water Resour As 41:621–633

Ma YJ, Li XY, Liu L, Yang XF, Wu XC, Wang P, Lin H, Zhang GH, Miao CY (2019) Evapotranspiration and its dominant controls along an elevation gradient in the Qinghai Lake watershed, northeast Qinghai-Tibet Plateau. J Hydrol 575:257–268

Maes WH, Gentine P, Verhoest NEC, Miralles DG (2019) Potential evaporation at eddy-covariance sites across the globe. Hydrol Earth Syst Sci 23:925–948

Makking GF (1957) Testing the Penman formula by means of lysimeters. J Inst Water Eng 11:277–288

Malik A, Kumar A, Kisi O (2017) Monthly pan-evaporation estimation in Indian central Himalayas using different heuristic approaches and climate based models. Comput Electron Agric 143:302–313

Middleton N, Thomas D (1997) World atlas of desertification, 2nd edn. UNEP, London

Mohammadrezapour O, Piri J, Kisi O (2019) Comparison of SVM, ANFIS and GEP in modeling monthly potential evapotranspiration in an arid region (Case study: Sistan and Baluchestan Province, Iran). Water Supply 19:392–403

Moriasi DN, Gitau MW, Pai N, Daggupati P (2015) Hydrologic and water quality models: performance measures and evaluation criteria. T Asabe 58:1763–1785

Penman HL (1948) Natural evaporation from open water, bare soil, and grass. Proc Royal Soc Lond A 193:120–146

Rácz C, Nagy J, Dobos AC (2013) Comparison of several methods for calculation of reference evapotranspiration. Acta Silv et Lignaria Hungarica 9:9–24

Rahimikhoob A (2010) Estimation of evapotranspiration based on only air temperature data using artificial neural networks for a subtropical climate in Iran. Theor Appl Climatol 101:83–91

Rahimikhoob A, Behbahani MR, Fakheri J (2012) An evaluation of four reference evapotranspiration models in a subtropical climate. Water Resour Manag 26:2867–2881

Safeeq M, Fares A (2012) Hydrologic response of a Hawaiian watershed to future climate change scenarios. Hydrol Process 26:2745–2764

Sayyadi H, Oladghaffari A, Faalian A, Sadraddini AA (2009) Comparison of RBF and MLP neural networks performance for estimation of reference crop evapotranspiration. Water Soil Sci 19: 1–12

Schemmel F, Mikes T, Rojay B, Mulch A (2013) The impact of topography on isotopes in precipitation across the Central Anatolian Plateau (Turkey). Am J Sci 313:61–80

Seifi A, Riahi H (2020) Estimating daily reference evapotranspiration using hybrid gamma test-least square support vector machine,

gamma test-ANN, and gamma test-ANFIS models in an arid area of Iran. J Water Clim Change 11:217–240
Sentelhas PC, Gillespie TJ, Santos EA (2010) Evaluation of FAO Penman–Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. Agr Water Manag 97:635–644
Serengil Y (2018) Climate change and carbon management. UNDP, Ankara
Stephens JC, Stewart EH (1963) A comparison of procedures for computing evaporation and evapotranspiration. Publication 62:123–133
Sudheer KP, Gosain AK, Ramasastri KS (2003) Estimating actual evapotranspiration from limited climatic data using neural computing technique. J Irrig Drain Eng 129:214–218
Sun JY, Sun XY, Hu ZY, Wang GX (2020) Exploring the influence of environmental factors in partitioning evapotranspiration along an elevation gradient on Mount Gongga, eastern edge of the Qinghai-Tibet Platea, China. J Mt Sci 17:384–396
Tabari H, Talaee PH (2013) Multilayer perceptron for reference evapotranspiration estimation in a semiarid region. Neural Comput & Applic 23:341–348
Tabari H, Kisi O, Ezani A, Talaee PH (2012) SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. J Hydrol 444:78–89
Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106:7183–7192
Tellen VA (2017) A comparative analysis of reference evapotranspiration from the surface of rainfed grass in Yaounde, calculated by six empirical methods against the penman-monteith formula. Earth Perspect 4:1–8
Thornthwaite CW (1948) An approach toward a rational classification of climate. Geogr Rev 38:55–94
Trajkovic S, Todorovic B, Stankovic M (2003) Forecasting reference evapotranspiration by artificial neural networks. J Irrig Drain E 129:454–457
Traore S, Wang YM, Kerh T (2010) Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. Agr Water Manag 97:707–714
Tsangaratos P, Ilia I (2017) Applying machine learning algorithms in landslide susceptibility assessments. In: Samui P, Sekhar S, Balas VE (eds) Handbook of neural computation. Academic Press, London, pp 433–457
Tukimat NNA, Harun S, Shahid S (2012) Comparison of different methods in estimating potential evapotranspiration at Muda Irrigation Scheme of Malaysia. J Agr Rural Dev Trop 113:77–85
Türkeş M, Tatlı H (2011) Use of the spectral clustering to determine coherent precipitation regions in Turkey for the period 1929–2007. Int J Climatol 31:2055–2067
Vapnik VN (1995) The nature of statistical learning theory. Springer Verlag, NewYork
Vicente-Serrano SM, Lanjeri S, López-Moreno JI (2007) Comparison of different procedures to map reference evapotranspiration using geographical information systems and regression-based techniques. Int J Climatol 27:1103–1118
Wang S, Lian J, Peng Y, Hu B, Chen H (2019) Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China. Agr Water Manag 221:220–230
Wang LH, He XB, Steiner JF, Zhang DW, Wu JK, Wang SY, Ding YJ (2020) Models and measurements of seven years of evapotranspiration on a high elevation site on the Central Tibetan Plateau. J Mt Sci 17:3039–3053
Wen X, Si J, He Z, Wu J, Shao H, Yu H (2015) Support-vector-machine-based models for modeling daily reference evapotranspiration with limited climatic data in extreme arid regions. Water Resour Manag 29:3195–3209
Xu CY, Singh VP (2002) Cross comparison of empirical equations for calculating potential evapotranspiration with data from Switzerland. Water Resour Manag 16:197–219
Yirga SA (2019) Modelling reference evapotranspiration for Megecha catchment by multiple linear regression. Model Earth Syst Environ 5:471–477
Yurtseven I, Zengin M (2013) Neural network modelling of rainfall interception in four different forest stands. Ann For Res 56:351–362
Zanetti SS, Sousa EF, Oliveira VP, Almeida FT, Bernardo S (2007) Estimating evapotranspiration using artificial neural network and minimum climatological data. J Irrig Drain Eng 133:83–89
Zheng F, Maier HR, Wu W, Dandy GC, Gupta HV, Zhang T (2018) On lack of robustness in hydrological model development due to absence of guidelines for selecting calibration and evaluation data: Demonstration for data-driven models. Water Resour Res 54: 1013–1030