Water Resour Manage (2013) 27:4815–4826
DOI 10.1007/s11269-013-0440-y

A Comparison Between Conventional and M5 Model Tree Methods for Converting Pan Evaporation to Reference Evapotranspiration for Semi-Arid Region

Ali Rahimikhoob · Maryam Asadi · Mahmood Mashal

Received: 26 October 2012 / Accepted: 22 August 2013 /
Published online: 10 September 2013

Abstract In this study, the performance of M5 model tree and conventional method for converting pan evaporation data (E_p) to reference evapotranspiration (ET_0) were assessed in semi-arid regions. Conventional method uses pan coefficient (K_p) as a factor to convert E_p to ET_0. Two common K_p equations for pans with dry fetch (Allen et al. 1998; Abdel-Wahed and Snyder in J Irrig Drain Eng 134(4):425–429, 2008) were considered for the comparison. The values of ET_0 derived using these three methods were compared to those estimated using the reference FAO Penmane Monteith (FAO-PM) method under semi-arid conditions of the Khuzestan plain (Southwest Iran). The results showed that the M5 model is the best one to estimate ET_0 over test sites (0.5 mm d^{-1} of root mean square error (RMSE) and 0.98 of coefficient of determination (R^2). Conversely, the performance of the two K_p equations was poor.

Keywords Reference evapotranspiration · Pan evaporation · M5 model tree · FAO-56 Penman–Monteith equation

1 Introduction

Accurate estimation of reference evapotranspiration (ET_0) is needed for water resources management, farm irrigation scheduling, and environmental assessment. A large number of methods have been developed for assessing ET_0 from meteorological data. The Penman Monteith (PM) method is recommended by FAO as the sole method to calculate reference evapotranspiration wherever the required input data are available (Allen et al. 1998). The PM is a physically based approach, which requires air temperature, relative humidity, solar radiation, and wind speed. The details of the PM equation are provided in the FAO's Irrigation and Drainage Paper Number 56 (FAO-56) (Allen et al. 1998). Unfortunately, there are a limited number of meteorological stations even in developed countries where these climatic variables are accurately measured. Empirical ET_0 models that require fewer

A. Rahimikhoob (✉) · M. Asadi · M. Mashal
Department of Irrigation and drainage Engineering, abouraihan Campus, University of Tehran, Tehran, Iran
e-mail: akhob@ut.ac.ir

variables exist. In the past decade, considerable attention has been focused on the evaluation of these models. For example, Trajkovic and Kolakovic (2009) evaluated five ET_0 estimation methods by comparing the estimates with results from the reference FAO-56 Penmane Monteith (FAO-PM) equation under humid conditions. They showed that Turc's method gave the best ET_0 estimates and ranking first, and other equations ranked in decreasing order are: Priestley–Taylor, Jensen–Haise, Thornthwaite, and Hargreaves. Tabari (2010) evaluated four simpler models based on monthly performance for Various Climates in Iran. The author reported that the Makkink and Priestley–Taylor models estimated ET_0 values less accurately than Turc and Hargreaves models for the all climates. Chauhan and Shrivastava (2009) compared the performance of four climate based methods and Artificial Neural Networks (ANNs) for estimation of ET_0 in India, when input climatic parameters are insufficient to apply FAO-PM method. They concluded that the ANN models were performed better than the climatic based methods.

Evaporation pans (class A pan, US Weather Bureau) are used extensively throughout the world to estimate ET_0. Evaporation pan (E_p) provides a measurement of the combined effect of temperature, humidity, wind speed and solar radiation on the reference crop evapotranspiration. This measurement can successfully be used to estimate ET_0 with a reasonable accuracy (Irmak et al. 2002). Numerous studies have shown that a high correlation between E_p and ET_0 can be obtained when evaporation pans are properly maintained (Jensen et al. 1961; Doorenbos and Pruitt 1977; Irmak et al. 2002). Conventional method uses Pan coefficient (K_p) as a factor to convert E_p into ET_0. Since the evaporation rate from the open pan and the ET_0 rate from the vegetated surface differ, ET_0 is computed by multiplying the E_p with K_p to account for differences between the grass and open water.

Doorenbos and Pruitt (1977) reported that the K_p values range from 0.40 to 0.85, depending on the prevailing upwind fetch distance (F) and climatic parameters such as wind speed at 2 m height (U_2) and air relative humidity (RH). Fetch is the horizontal distance that the wind blows over green vegetation or dry surface to reach the pan. So the ground cover in the station influences the K_p values. Two cases of evaporation pan sitting are considered (Allen et al. 1998): 1) the pan is sited on a short green vegetation cover (green fetch) and surrounded by fallow soil, and 2) the pan is sited on fallow soil (dry fetch) and surrounded by a green crop.

The K_p values were first published by Jensen (1974) and subsequently tabulated by FAO-24 (Doorenbos and Pruitt 1977). Doorenbos and Pruitt (1977) suggested K_p values for the two cases of evaporation pan siting in tabular form for a number of fetch distance under different wind speed and relative humidity conditions. The values for F are presented quantitatively but those of U_2 and RH are presented as classifications in their table. When the K_p values were first reported no computers were available. Later on when computers and data loggers were developed and when electronic data transmission became possible, automatic conversion of E_p to ET_0 and the elimination of search operations became possible (Snyder 1992). Since then, several empirical equations to calculate daily values of K_p have been developed based on Doorenbos and Pruitt (1977) table using linear, nonlinear, and indicator regression techniques (Frevert et al. 1983; Cuenca 1989; Snyder 1992; Allen and Pruitt 1991 and Raghuwanshi and Wallender 1998).

The fundamental question of which equation predicts K_p most accurately has been considered in several studies. Irmak et al. (2002) evaluated the techniques of Frevert et al. (1983) and Snyder (1992) to convert E_p to ET_0 in the humid climate of Gainesville, Florida. Results of Irmak et al. (2002) showed that ET_0 calculated using the daily K_p values from Equation of Frevert et al. (1983) provided more accurate daily, monthly, and annual total estimates compared to the ET_0 calculated using K_p values from Equation of Snyder (1992) when the FAO-PM method was used as a reference for this climatic condition. The Snyder (1992) method tended to overestimate ET_0 calculated by the FAO-PM method, especially in summer (Irmak et al. 2002). In another study, Sabziparvar et al. (2010) compared seven exiting pan models to estimate K_p values for two different climates of Iran. They showed that, for the cold semi-arid climate condition, the best K_p models for estimation of ET_0 were Orang and Raghuwanshi–Wallender, respectively. Also, the Snyder and Orang models were best fitted models for warm arid climate, respectively. Trajkovic and Kolakovi (2010) evaluated the reliability of simplified pan-based approaches for estimating ET_0. In this study, three pan-based (FAO-24 pan, Snyder ET_0, and Ghare ET_0) equations were compared against lysimeter measurements of grass evapotranspiration using daily data from Policoro, Italy. Based on summary statistics, the Snyder ET_0 equation ranked first with the lowest RMSE value.

The above K_p equations were presented for pans with green fetch and only two equations were presented for pans with dry fetch. Allen and Pruitt (1991) developed a non-linear K_p equation for a Class A pan type with fallow soil surrounding condition. This equation was presented by Allen et al. (1998) in FAO-56. Abdel-Wahed and Snyder (2008) reported that the equation to calculate K_p developed by Allen et al. (1998) was somewhat complex and as a result, they proposed a simpler equation to calculate daily K_p values for a pan placed in a dry fallow area. Evaporation pans are placed in dry fallow area at most weather stations in Iran, especially in arid and semi-arid environment, so it is desirable to select the appropriate K_p equations. Therefore, the first objective of this study was to compare the Allen et al. (1998) and Abdel-Wahed and Snyder (2008) equations to estimate ET_0 by comparing them against the FAO-PM method using data collected in a semi arid climate of Iran. The FAO-PM method was chosen as a standard for testing the accuracy of the K_p equations in this study because there were no measured ET_0 data at this location. This method was accepted as a standard method for estimating ET_0 by the FAO (Allen et al. 1998).

Recently, M5 model trees have been used successfully for flood forecasting (Solomatine and Xue 2004), water level-discharge relationship (Bhattacharya and Solomatine 2005), rainfall-runoff modeling (Solomatine and Dulal 2003), sedimentation modeling (Bhattacharya and Solomatine 2006), and estimation of ET_0 (Pal and Deswal 2009). Pal and Deswal (2009) investigated the potential of M5 model tree based regression approach to model daily ET_0 using four inputs including solar radiation, average air temperature, average relative humidity, and average wind speed. Results from their study suggested that M5 model tree could successfully be employed in modeling the ET_0. The second objective of this study was to examine the potential of this approach for converting E_p to ET_0. A comparison between conventional approach and M5 model tree was the last objective of this study.

2 Materials and Methods

2.1 Study Area and Data

The area under study was Khuzestan province, which lies between latitudes 29.95°N and 32.9°N and between longitudes 47.6°E and 50.6°E. Khuzestan province is in the south-west of Iran, borders Iraq and the Persian Gulf, and covers an area of 63,238 km^2. On the basis of the Koppen climate classification, Khuzestan province is categorized as having a semi-arid climate. The average annual rainfall ranges from 320 mm in the east to 145 mm in the west and occasionally reaches as high as 400 mm in the east. Based on the climatic data from meteorological stations, the maximum annual rainfall is experienced during winter and late fall. The air temperature reaches its maximum in August and its minimum in January. According to the climatic data from meteorological stations, the average annual temperature along the Khuzestan province has varied from 21.5 °C in the north to 25.3 °C in the south over the past decade. The warmest temperature of the warmest month ranges from 38 °C to 47 °C, while that of the coldest month ranges from 1.5 °C to 4 °C. Daily mean relative humidity ranges from 13 to 92 % with an annual average of 54 %. The highest wind speed of approximately 259 km day^{-1} usually occurs in December. Wind speed is usually lowest from June through September, ranging from 47 to 145 km day^{-1} and averaging 96 km day^{-1}.

Measured weather data were obtained from eight weather stations across the study area with varying latitudes, longitudes, and elevations. The spatial distribution of selected stations is shown in Fig. 1. The stations belong to the meteorological organization of Iran. Information regarding the sites and mean annual values of relevant weather variables are given in Table 1. The dataset consist of daily records of 12 years (1997–2008) of maximum and minimum air temperature, T_x and T_n respectively, (°C), relative humidity, RH, (%), wind speed, U, (m s^{-1}), bright sunshine hours, n, (hours) and class A pan evaporation, E_{pan}, (mm d^{-1}). Monthly means of these daily data were used for estimating K_p and ET_0 on a monthly basis. Measurements were made at a height of 2 m (air temperature and relative humidity) and 10 m (wind speed) above the soil surface. Wind speeds at 2 m (U_2) were obtained from those taken at 10 m using the log-wind profile equation. The Class-A pan evaporimeters (USWB) were 0.25 m deep and 1.21 m diameter were made of galvanized steel. The bottoms of the pans were supported 0.15 m above the ground level on open-frame wooden platforms. The water level in the pans was maintained between 5.0 and 7.5 cm from the rim. E_p values were measured on the stations daily at 7.00 AM (local time).

2.2 Conventional Method of Estimating ET_0

The basic form of the conventional method as described by FAO-24 (Doorenbos and Pruitt 1977) is $ET_0 = K_p \times E_p$. In this study two K_p equations proposed by Allen et al. (1998) and

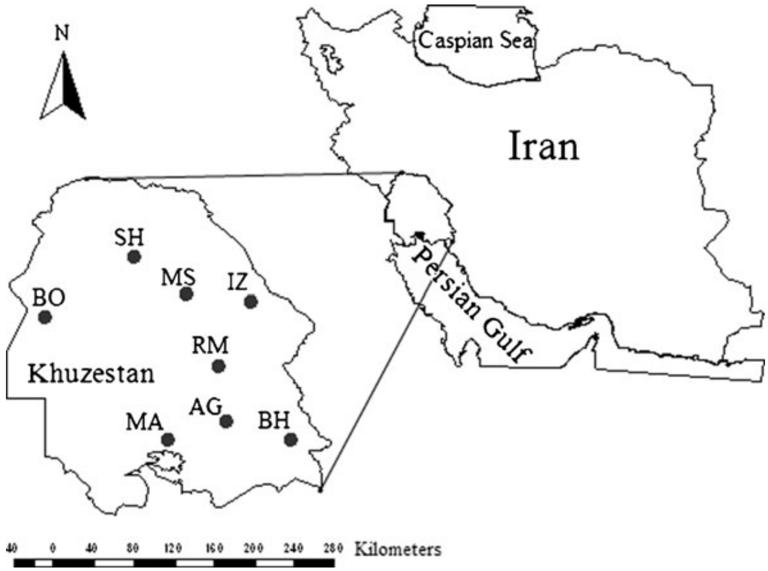

Fig. 1 Study area and location of the weather stations

Table 1 Summary of weather stations used in the study

Station	Code	Lat. (°N)	Alt. (m)	T_{max} (°C)	T_{min} (°C)	RH (%)	U (m/s)
Masjedsoliman	MS	31.93	320.5	32.1	19.6	39.1	1.4
Bostan	BO	31.72	7.8	33.3	16.5	45.1	2.6
Shushtar	SH	32.05	67	33.0	20.3	37.5	2.2
Ramhormoz	RA	31.27	150.5	32.7	19.7	38.2	1.6
Izeh	IZ	31.85	767	28.4	13.9	37.4	1.4
Behbahan	BH	30.60	313	32.7	17.3	40.7	1.0
Mahshahr	MA	30.55	6.2	32.4	18.6	45.9	2.7

Abdel-Wahed and Snyder (2008) were evaluated (Table 2). The K_p equations are functions of daily mean relative humidity, RH (%), daily mean wind speed, U_2 (m s^{-1}), and fetch distance, F (m), as defined by Doorenbos and Pruitt (1977). All the stations used in this study are surrounded by dry fallow land. In the K_p calculations, F was taken as 1,000 m since the weather stations were surrounded by dry fallow land.

2.3 M5 Model Tree

M5 model tree was first presented by Quinlan (1992). The model is based on a binary decision tree having linear regression functions at the terminal (leaf) nodes, which develops a relationship between independent and dependent variables. Unlike decision tree which is used for categorical data, it can also be used for quantitative data (Quinlan 1992; Mitchell 1997). M5 model tree generation requires two different stages (Quinlan 1992; Solomatine and Xue 2004). The first stage involves splitting of the data into subsets to create a decision tree. The splitting criterion is based on treating the standard deviation of the class values that reach a node as a measure of the error at that node, and calculating the expected reduction in this error as a result of testing each attribute at that node. The formula for computing the standard deviation reduction (SDR) is defined as follows (Pal and Deswal 2009):

$$SDR = sd(\mathrm{T}) - \sum \frac{|\mathrm{T_i}|}{|\mathrm{T}|} sd(\mathrm{T_i}) \tag{3}$$

where T denotes a set of examples that reaches the node; T_i denotes the subset of examples that have the ith outcome of the potential set; sd denotes the standard deviation (Wang and Witten 1997). Due to the splitting process, the standard deviation of the data in child nodes (lower nodes) is less than that at the parent node. After examining all the possible splits, the one that maximizes the expected error reduction was chosen. However, this division often

Table 2 K_p equations in the evaluation analysis

Authors (year)	K_p equations	
Allen et al. (1998)	$K_p = 0.61 + 0.00341 \times RH - 0.000162 \times U_2 \times RH - 0.00000959 \times U_2 \times F + 0.00327 \times U_2 \times ln\ F - 0.00289 \times U_2 \times ln\ (86.4 \times U_2) - 0.0106 \times ln(86.4 \times U_2) \times ln\ F + 0.00063 \times [ln\ F]^2 \times ln(86.4 \times U_2)$	(1)
Abdel-Wahed and Snyder (2008)	$K_p = 0.62407 - 0.02660\ \ln F - 0.00028 \times U_2 + 0.00226 \times RH$	(2)

produces a large tree-like structure which may cause over fitting or poor generalization. To overcome this problem, in second stage the overgrown tree is pruned and then pruned sub-trees are replaced with linear regression functions. This technique of generating the model tree substantially increases the accuracy of estimation (Quinlan 1992). Figure 2a shows splitting the input space X1 × X2 (independent variables) into six subspaces (leaves) by M5 model tree algorithm. A linear regression function was built at the leaves, labeled LM1 through LM6. Figure 2b shows its relations in form of tree diagram, in which LM1 to LM6 is in leave level. Further details of the M5 model tree can be found in Quinlan (1992).

In this study, pan evaporation data (mm d^{-1}) with relative humidity (%) and daily mean wind speed (m s^{-1}) were selected as inputs to the M5 model tree for estimating reference evapotranspiration. The whole data of Mahshahr, Ramhormoz, Izeh and Bostan stations (from 1997 to 2008) were collected into one group in order to create the M5 model tree that has a higher regional capacity that could be applied to estimate ET_0 for different locations in Khuzestan. After the creating process, the whole data of Aghajari, Behbahan, Masjedsoliman and Shushtar stations (from 1997 to 2008) were used to test the created model.

2.4 The FAO Penman–Monteith (FAO-PM)

In this study, the performance of empirical methods and M5 model tree were compared with the conventional FAO Penman–Monteith method. Although in practice, the best way to test the performance of the empirical methods would be to compare their performances against lysimeter-measured data; this type of data set is not available in the study area. The following equation was applied for the PM (Allen et al. 1998):

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T_a + 273}U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \tag{4}$$

where ET_0 is reference crop evapotanspiration (mm d^{-1}), R_n is the daily net radiation (MJ m^{-2} d^{-1}), G is the daily soil heat flux (MJ m^{-2} d^{-1}), T_a is the mean daily air temperature at a height of 2 m (°C), U_2 is the daily mean wind speed at a height of 2 m (m s^{-1}), e_s is the saturation vapor pressure (kPa), e_a is the actual vapor pressure (kPa), Δ is the slope of the saturation vapor pressure versus the air temperature curve (kPa $°C^{-1}$), and γ is the psychrometric constant (kPa $°C^{-1}$). The terms in the numerator on the right-hand side of the equation are the radiation term and aerodynamic term, respectively.

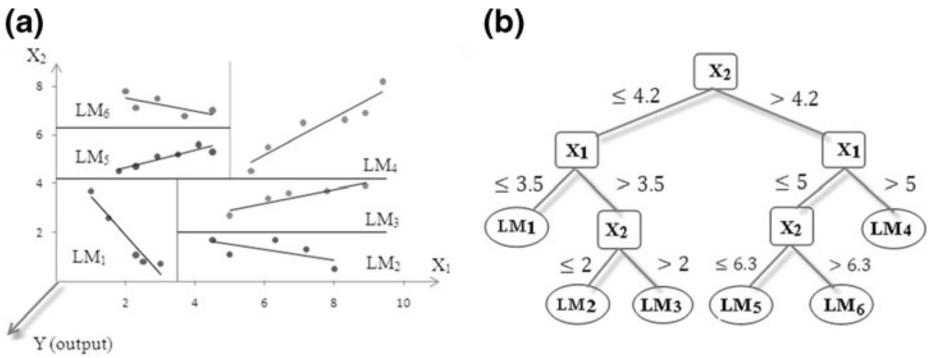

Fig. 2 Example of M5 model tree, **a** splitting the input space X1 × X2 by M5 model tree algorithm, **b** diagram of model tree with six linear regression models at the leaves

In this study, the daily values of Δ, R_n, e_s and e_a were calculated using the equations given by Allen et al. (1998). For R_n, an albedo of 0.23 (green vegetation surface) was used. Since G is usually small compared with R_n and is difficult to measure, it was assumed to be zero over the calculation time step period (daily and monthly) (Allen et al. 1998). The measured RH, T_x and T_n values were used to calculate e_a and e_s. The daily solar or shortwave radiation (R_s) was calculated using the Angstrom formula, which relates solar radiation to extraterrestrial radiation and relative sunshine duration. Equation (39) in Allen et al. (1998) was used to calculate the net outgoing longwave radiation.

2.5 Statistical Analysis

The comparison between the models (M5 and two K_p equations) and the FAO-PM model was carried out using: (1) a linear regression equation (Y = mX + c), through least square regression, between ET_0 computed by FAO-PM equation and ET_0 estimated from the above mentioned three methods (m and c are the slope and the intercept of the regression equation, respectively); (2) the coefficient of determination (R^2); (3) the Root Mean Square Error (RMSE). In the case of a perfect correlation with no bias, c=0 and m=1, R^2=1 and RMSE=0.

3 Results and Discussion

To assess the estimation capacity of the Kp equations and to express the interactions between the different variables a correlation matrix for two training and testing data set was prepared (Table 3). Using a 95 % level F test, nearly all variables are significantly intercorrelated. It can be observed from Table 3 that the linear correlation between E_p and ET_0 is high (0.98 and 0.96 for training and testing data set, respectively) implying that any model built using E_p will certainly be able to compute the ET_0 satisfactorily. The relationship between K_p equations and ET_0 FAO-PM shows a statistically significant correlation as well. The model's accuracy can be improved by incorporating K_p variables that account for aerodynamic effects on ET_0, such as humidity and wind speed in addition to E_p. As seen from these results, the correlation coefficients of two K_p equations and ET_0 FAO–PM are negative, which indicates a decrease in K_p values, the ET_0 rate will increase. This could be attributed

Table 3 Correlation matrix between ET0–PM, relative humidity (RH), wind velocity (U_2), K_p equations and pan evaporation (E_p) for two training and testing data set

Variables	Training data set						Testing data set					
	U_2	RH	K_p (Eq. 1)	K_p (Eq. 2)	E_p	ET_0-PM	U_2	RH	K_p (Eq. 1)	K_p (Eq. 2)	E_p	ET_0
U_2 (m s^{-1})	1						1					
RH (%)	−0.29	1					−0.52	1				
K_p (Eq. 1)	−0.77	0.82	1				−0.80	0.92	1			
K_p (Eq. 2)	−0.30	0.92	0.83	1			−0.53	0.98	0.92	1		
E_p (m d^{-1})	0.66	−0.80	−0.89	−0.80	1		0.62	−0.89	−0.88	0.90	1	
ET_0 (mm d^{-1})	0.69	−0.81	−0.92	−0.82	0.98	1	0.76	−0.87	−0.92	−0.87	0.96	1

to the fact that the decrease in Kp values is associated with a reduction in aerodynamic resistance to ET_0, greater ET_0 resulting in lower relative humidity and higher wind speed. Among the two K_p equations, the Allen et al. (1998) equation shows a high correlation coefficient (r=−0.92 for the both data set) with ET_0 FAO–PM.

All monthly K_p data calculated from the two K_p equations were averaged over the 12 years to obtained mean monthly estimated K_p. The comparisons of calculated monthly K_p values using Eqs. 1 and 2 for all the stations are given in Fig. 3. The evolution of monthly values of K_p were nearly similar for all equations. Equation 1 gave a lower value, whereas Equation 2 gave a higher value of K_p for all months.

For building model tree, based on creating data set, the Weka software (Witten and Frank 2005) was used. The model tree generated by M5 algorithm is shown in Fig. 4. As can be seen, four rules (LM1 to LM4) were generated. Figure 5 shows the scatter plot between ET_0 estimated by the FAO-PM method and M5 model estimated ET_0 for all creating data set. As seen from the fit line equation there is a very good agreement (m=1.0, with c=−0.006 and R^2=0.99) and less scatter between the points.

The ET_0 estimates of developed M5 model tree and conventional K_p equations for the data set of test locations are illustrated in Fig. 6 in the form of scatterplot. It is clear from the scatterplots that the M5 estimates are closer to the corresponding FAO-PM ET_0 values than those of the Two K_p equations. As seen from the fit line equations in the scatterplots that the m and c coefficients for the M5 model are closer to the 1 and 0 with a higher R^2 value than

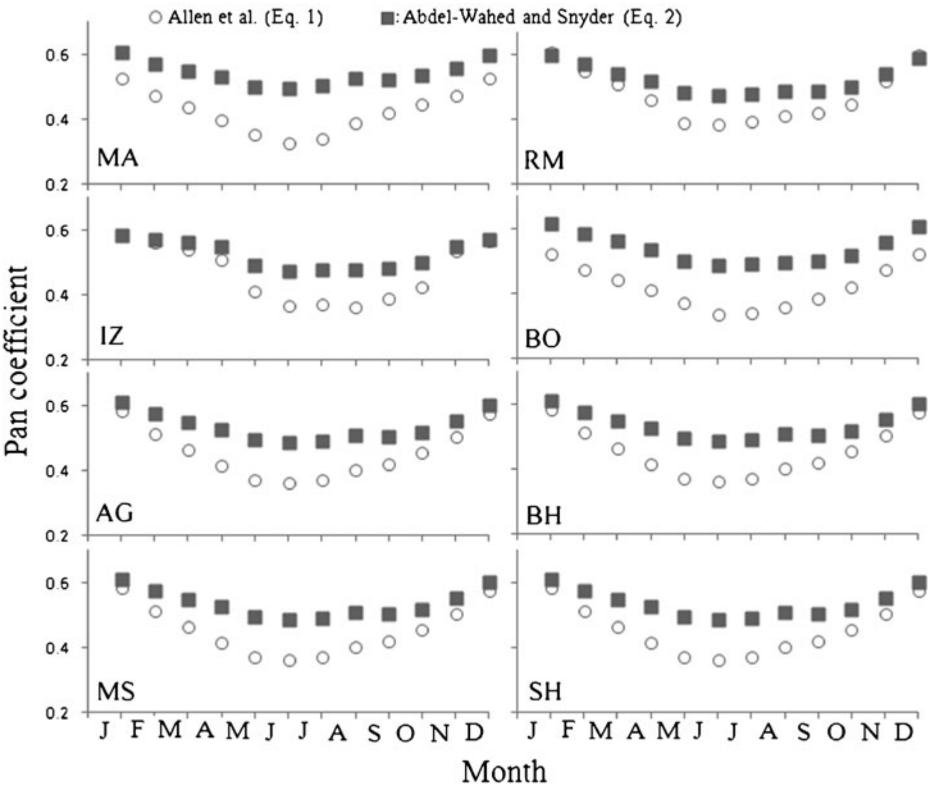

Fig. 3 Calculated monthly Kp values using the Kp equations

```
M5 pruned model tree:

        EPAN <= 7.953:
|          EPAN <= 3.955: LM1
|          EPAN > 3.955: LM2
        EPAN > 7.953:
|          EPAN <= 12.43: LM3
|          EPAN > 12.43: LM4
LM num: 1
      ETPM = 0.0228 * U2 - 0.0085 * HUM + 0.1168 * RA+ 0.4211 * EPAN - 0.0613
LM num: 2
      ETPM = 0.175 * U2 - 0.0242 * HUM+ 0.1847 * RA  + 0.2879 * EPAN + 0.1645
LM num: 3
      ETPM = 0.3115 * U2 - 0.033 * HUM + 0.1792 *   RA + 0.2512 * EPAN+ 0.6208
LM num: 4
      ETPM = 0.5013 * U2 - 0.0335 * HUM + 0.1793 * RA + 0.2265 * EPAN + 0.4662
```

Fig. 4 Linear models generated by M5 model tree

those of the other Kp equations. The slope of the fitted line is nearly close to one (lying on 1:1 line) for each station. This shows that the M5 model produces well for estimating ET_0 in the scatter plots the slope of straight line (m) varies between 0.98 and 1.13 with an average of 1.04. As seen from the scatter plots, Allen et al. (1998) and Abdel-Wahed and Snyder (2008) equations compared less favorably with FAO-PM values than the M5 method. Estimates by the Kp equations overestimated the ET_0 at all locations. This overestimation was constant throughout the study area.

The statistical results are reported in Table 4. According to these results, the M5 method seems to be the best one to calculate ET_0 in the Khuzestan plain (semi-arid climate). The coefficient of determination (R^2) and the slope are close to 1 and the value of RMSE= 0.50 mm d^{-1} can be also considered acceptable with regard to the average value of ET_0 (5.35 mm). In contrast to M5 model, the performance of the conventional methods (Eqs. 1

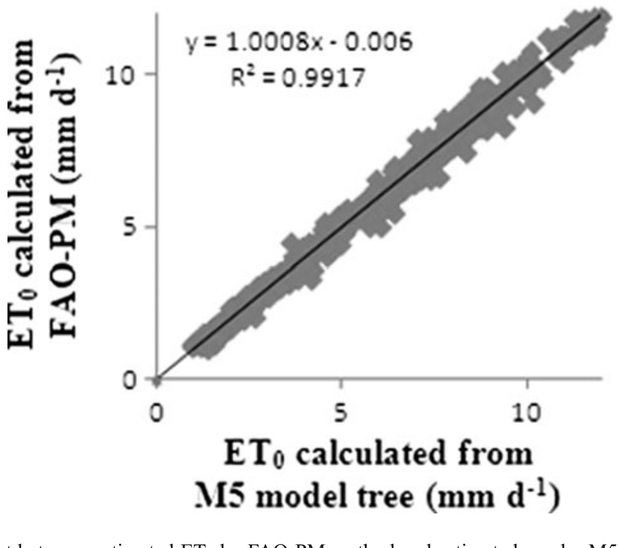

Fig. 5 Scatter plot between estimated ET_0 by FAO-PM method and estimated one by M5 model tree, using creating data set

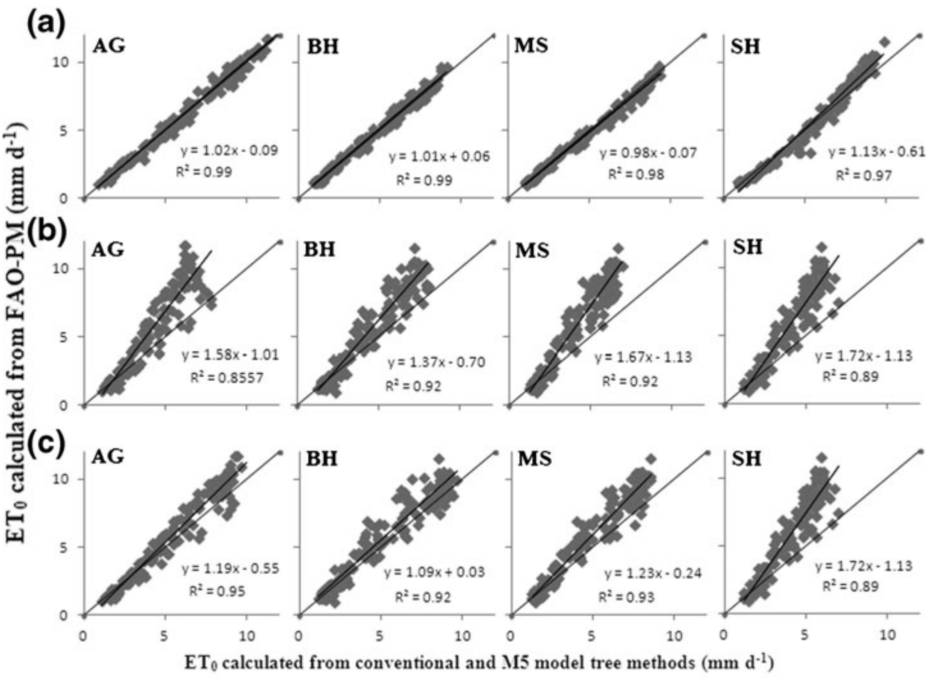

Fig. 6 Comparison between the values of ET_0 calculated by FAO-PM method and those by three methods at four test weather stations. **a** M5 model tree, **b** Allen et al. (1998) and **c** Abdel-Wahed and Snyder (2008)

and 2) was poor, the corresponding RMSE were 1.90 and 1.1 mm d^{-1} for Allen et al. (1998) and Abdel-Wahed and Snyder (2008), respectively (see Table 4 for other statistical analysis).

Table 4 Statistical values of the comparison between ET_0 estimated by FAO-PM method against those obtained by the three methods

Estimated method	Station code	Slope	Y-intercept mm d^{-1}	RMSE mm d^{-1}	R^2
M5	AG	1.02	−0.09	0.4	0.99
	BH	1.01	0.06	0.6	0.99
	MS	0.98	−0.07	0.4	0.98
	SH	1.13	−0.61	0.6	0.97
	Average	1.04	−0.18	0.5	0.98
Allen et al. (1998)	AG	1.58	−1.01	1.8	0.86
	BH	1.37	−0.7	1.5	0.92
	MS	1.67	−1.13	2.1	0.92
	SH	1.72	−1.13	2.3	0.89
	Average	1.59	−0.99	1.9	0.90
Abdel-Wahed and Snyder (2008)	AG	1.19	−0.55	0.9	0.95
	BH	1.09	0.03	1.0	0.92
	MS	1.23	−0.24	1.3	0.93
	SH	1.72	−1.13	1.2	0.89
	Average	1.31	−0.47	1.1	0.92

4 Conclusions

This study investigated the ability of M5 model tree for converting pan evaporation data to reference evapotranspiration under dry fetch condition in a semi- arid environment of Iran. The accuracy of M5 model tree has been compared to those of the two common K_p equations (Allen et al. in FAO irrigation and drainage paper number 56, 1998; Abdel-Wahed and Snyder in J Irrig Drain Eng 134(4):425–429, 2008). The monthly climatic data of eight weather stations in Khuzestan, are used for the model simulations. The Penman-Monteith method as recommended by FAO (Allen et al. 1998) was assumed as a standard in evaluating the above methods. The study demonstrated that modelling of reference evapotranspiration is possible through the use of M5 Model tree technique (RMSE of 0.4 to 0.6 mm d^{-1} for mean daily ET_0 of 4.5 to 5.7 mm d^{-1}) from pan evaporation, relative humidity, wind speed and extraterrestrial radiation data. The comparison results show that the M5 model tree approach works well in estimating reference evapotranspiration in comparison with conventional method that uses K_p equations. However, it would be suitable to consider for more humid and fetch distance to confirm this result.

References

Abdel-Wahed MH, Snyder RL (2008) Simple equation to estimate reference evapotranspiration from evaporation pans surrounded by fallow soil. J Irrig Drain Eng 134(4):425–429

Allen RG, Pruitt WO (1991) FAO-24 reference evapotranspiration factors. J Irrig Drain Eng 117(5):758–773

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration. Guidelines for computing crop water requirements. In: FAO irrigation and drainage paper, no 56. FAO, Roma, Italy

Bhattacharya B, Solomatine DP (2005) Neural networks and M5 model trees in modeling water level–discharge relationship. Neurocomputing 63:381–396

Bhattacharya B, Solomatine DP (2006) Machine learning in sedimentation modelling. Neural Netw 19(2):208–214

Chauhan S, Shrivastava RK (2009) Performance evaluation of reference evapotranspiration estimation using climate based methods and artificial neural networks. Water Resour Manag 23(5):825–837

Cuenca RH (1989) Irrigation system design: an engineering approach. Prentice-Hall, Englewood Cliffs

Doorenbos J, Pruitt WO (1977) Guidelines for prediction of crop water requirements. FAO Irrig Drain. Paper No. 24, Rome

Frevert DK, Hill RW, Braaten BC (1983) Estimation of FAO evapotranspiration coefficients. J Irrig Drain Eng 109(2):265–270

Irmak S, Haman D, Jones W (2002) Evaluation of class A pan coefficients for estimating reference evapotranspiration in humid location. J Irrig Drain Eng 128(3):153–159

Jensen ME (1974) Consumptive use of water and irrigation water requirements. Irrig Drain Div Rep ASCE N Y 89:15–41

Jensen MC, Middleton JE, Pruitt WO (1961) Scheduling irrigation from pan evaporation. Circular 386, Washington Agricultural Experiment Station

Mitchell TM (1997) Machine learning. The McGraw-Hill Companies, Inc., New York, NY, pp 414

Pal M, Deswal S (2009) M5 model tree based modelling of reference evapotranspiration. Hydrol Process 23:1437–1443

Quinlan JR (1992) Learning with continuous classes. In Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence, Hobart, Australia, 16–18 November, World Scientific, Singapore: 343–348

Raghuwanshi NS, Wallender WW (1998) Converting from pan evaporation to evapotranspiration. J Irrig Drain Eng 118(6):977–980

Sabziparvar AA, Tabari H, Aeini A, Ghafouri M (2010) Evaluation of Class A Pan Coefficient Models for Estimation of Reference Crop Evapotranspiration in Cold Semi-Arid and Warm Arid Climates. Water Resour Manag 24(5):909–920

Snyder RL (1992) Equation for evaporation pan to evapotranspiration conversions. J Irrig Drain Eng 118(6):977–980

Solomatine DP, Dulal KN (2003) Model trees as an alternative to neural networks in rainfall-runoff modelling. Hydrol Sci J 48(3):399–411

Solomatine DP, Xue Y (2004) M5 model trees compared to neural networks: application to flood forecasting in the upper reach of the Huai River in China. J Hydrol Eng 9(6):491–501

Tabari H (2010) Evaluation of reference crop evapotranspiration equations in various climates. Water Resour Manag 24:2311–2337

Trajkovic S, Kolakovi S (2010) Comparison of Simplified Pan-Based Equations for Estimating Reference Evapotranspiration. J Irrig Drain Eng 136(2):137–140

Trajkovic S, Kolakovic S (2009) Evaluation of reference evapotranspiration equations under humid conditions. Water Resour Manag 23:3057–3067

Wang Y, Witten IH (1997) Induction of model trees for predicting continuous lasses. In: Proceedings of the Poster Papers of the European Conference on Machine Learning. University of Economics, Faculty of Informatics and Statistics, Prague

Witten IH, Frank E (2005) Data Mining: Practical Machine Learning Tools and Technique. Morgan Kaufmann Publishers, San Francisco, pp 560