Arab J Geosci (2016) 9: 202
DOI 10.1007/s12517-015-2273-x

ORIGINAL PAPER

Modeling daily evapotranspiration in hyper-arid environment using gene expression programming

A. A. Alazba [1,2] · M. A. Yassin [2] · M. A. Mattar [1]

Received: 5 September 2015 / Accepted: 10 December 2015 / Published online: 10 March 2016
© Saudi Society for Geosciences 2016

Abstract Accurate estimations of reference evapotranspiration (ET_{ref}) are extremely important for maximizing the beneficial use of water and hydrologic applications, particularly in arid and semiarid regions where water sources are so limited. The aim of this study is to develop mathematical models to calculate the daily ET_{ref} using a gene expression programming (GEP) technique. Eight GEP models (GEP-MOD1–8) were developed from combinations of climatic variables. The Penman-Monteith equation was considered the reference method, with the reference plant height varying from 5 to 105 cm in 5-cm increments. Daily climatic variables collected from 13 meteorological stations, one station from every region within the Kingdom of Saudi Arabia, covered the 1980 to 2010 period. Of the available climatic data, 65 % was used in the training process for the eight developed GEP models, and 35 % was used in the testing process. The accuracy of the eight developed GEP models to estimate ET_{ref} varied in significance depending on the climatic variables that were included. As more climatic parameters were input, the accuracy of the GEP model increased. For the testing process, the coefficient of determination (R^2) ranged from a low of 63.4 % for GEP-MOD1 to a high of 95.4 % for GEP-MOD8, and the root mean square error (RMSE) values ranged from 3.19 mm day^{-1} for GEP-MOD1 to 1.14 mm day^{-1} for GEP-MOD8. From the spatial evaluation, the values of RMSE ranged from 3.27 mm day^{-1} for GEP-MOD1 to 1.21 mm day^{-1} for GEP-MOD8. In addition, the respective RMSE values resulting from GEP-MOD8 for plant heights of 50 and 12 cm were 0.75 and 0.96 cm. This implies that the developed GEP-MOD8 can be used for any value of the reference plant height ranging from 5 to 105 cm with insignificant errors. Interestingly, solar radiation had an almost insignificant effect on ET_{ref} in the hyper-arid conditions. In contrast, wind speed and plant height had a large positive effect on increasing the accuracy of calculating ET_{ref}.

Keywords Evapotranspiration · Penman-Monteith model · Gene expression programming · Arid environment

✉ A. A. Alazba
alazba@ksu.edu.sa; alazba@hotmail.com

[1] Agricultural Engineering Department, King Saud University, Riyadh, Saudi Arabia

[2] Alamoudi Water Chair, King Saud University, Riyadh, Saudi Arabia

Introduction

Accurate quantification of evapotranspiration (ET) is necessary for different applications because it is generally the largest component of water and energy balance (Suleiman and Hoogenboom 2009). Direct or indirect measurement methods incorporating physical and empirical models can be used to estimate ET. The most accurate direct method is measurement with a lysimeter (Gavilan et al. 2007). However, this procedure is difficult, time-consuming, expensive, and unavailable in many locations (Kisi 2007; Wang et al. 2009). Estimating ET through indirect methods involves calculating the reference evapotranspiration (ET_{ref}) and then applying a suitable coefficient. ET_{ref} is considered a major component of the hydrologic process. Therefore, accurate estimation of ET_{ref} is crucial for studies on hydrologic water balance, irrigation system design and management, plant production, water resources planning and management, and environmental assessment (Irmak et al. 2003; Temesgen et al. 2005; Chattopadhyay et al. 2009; Kumar et al. 2011). Many factors affect ET_{ref}, including daily temperature, relative humidity, wind velocity, hours of sunshine, atmospheric pressure, soil heat flux, and

latitude. All of these factors cause estimation of ET_{ref} to be one of the most complex calculations in the meteorology and hydrology sciences because its estimation depends on the interaction of several climatic elements (Kumar et al. 2002). Several methods exist to compute ET_{ref} using climatological data that may be gathered from empirical relationships based on temperature, radiation, or mass transfer, or from combination methods based on physical processes. A typical combination method is the physically based, complex Penman (1948) equation. The combination approach links evaporation dynamics with the flux of net radiation and the aerodynamic transport characteristics of a natural surface. Monteith (1965) introduced a surface conductance term to account for the response of leaf stomata to their hydrological environment. This modified form of the Penman equation, widely known as the Penman-Monteith (PM) evapotranspiration model (Monteith 1973), is used by the Food and Agriculture Organization of the United Nations PM-FAO56 (Allen et al. 1998) and by the American Society of Civil Engineers PM-ASCE70 (Jensen et al. 1990; Walter et al. 2001; ASCE-EWRI 2005) as the only accurate method for calculating ET_{ref} and validating other equations. This equation incorporates thermodynamic and aerodynamic aspects, can be applied in a wide range of climatic contexts (Smith et al. 1991; Yin et al. 2008), and requires many climatic data inputs including solar radiation, wind speed, air temperature, and humidity. Its disadvantage is that all of the necessary daily climatic variables may not be available for a given location. This is especially true in developing countries where reliable weather data sets of radiation, relative humidity, and wind speed are limited (Trajkovic 2010; Tabari and Hosseinzadeh Talaee 2011). Furthermore, the installation and maintenance of weather station equipment can be expensive and complicated (Sentelhas et al. 2010). Many researchers acknowledge that the Penman-Monteith model is the most promising standardized method to estimate ET_{ref}; however, it requires a significant amount of climatic data, which may be unavailable or of low reliability in certain locations, especially within developing countries. In these cases, alternative methods that rely on fewer weather inputs are necessary (e.g., George et al. 2002; Xu and Singh 2002; Fooladmand et al. 2008; Sabziparvar and Tabari 2010; Tabari 2010; Trajkovic and Kolakovic 2010).

Over the past decade, intelligent computational models such as gene expression programming (GEP) have been applied as an alternative method for estimating ET_{ref}. This technique, an extension of the genetic programming (GP) of Koza (1992), was developed by Ferreira (2001b). GEP, which is a natural development of genetic algorithms, is defined as an algorithm used to implement symbolic regression in an attempt to find a mathematical function that fits a data set (Sakthivel et al. 2012). The method is used to solve problems including symbolic regression, multi-agent strategies, time series prediction, circuit design, and evolutionary neural networks (Samadianfard 2012). GEP has been applied in fields as diverse as artificial intelligence; artificial life; engineering and science; financial markets; industrial, chemical, and biological processes; and mechanical models. The capability of GEP to solve hydrological and hydraulic problem modeling has been examined by a number of researchers. Guven and Aytek (2009) used the GEP approach to model stage-discharge relationships and compared the obtained results with conventional methods. They found that the explicit algebraic formulations developed by GEP produced the best results. Similarly, Azamathulla et al. (2011) developed mathematical models for the estimation of stage-discharge relationships for the Pahang River based on the GP and GEP techniques. More recently, Azamathulla and Ahmad (2012) used GEP to predict the transverse mixing coefficient in open-channel flows. Furthermore, Fernando et al. (2012) introduced an innovative method for combining estimated outputs from a number of rainfall-runoff models using GEP to perform symbolic regression. The results showed that the GEP combination method produced a superior river flow forecast. Azamathulla (2013) used GEP to drive a new model to predict the friction factor for southern Italian rivers.

Only a few of the various published studies on the application of GEP in hydrological modeling have examined the applicability of GEP for modeling of the evapotranspiration process. Shiri et al. (2012) introduced a new GEP model to estimate daily ET_{ref} at four weather stations in northern Spain over a 5-year period (1999–2003). The GEP results were compared to the Adaptive Neuro-Fuzzy Inference System (ANFIS) and Priestley-Taylor and Hargreaves models using the PM-FAO56 equation as the reference. The comparisons showed that the GEP performed better than these three models. Moreover, Traore and Guven (2012) evaluated the ability of GEP to model the ET_{ref} of the tropical seasonally dry regions of West Africa using routing meteorological data from Burkina Faso. Their results revealed the GEP model to be a fairly promising approach with the advantages of successfully providing simple algebraic formulas and ease of use without recourse to the full set of meteorological data required for accurately estimating ET_{ref} in sub-Saharan Africa regions. However, other GP techniques have been applied in the modeling of the evapotranspiration process. Parasuraman et al. (2007) evaluated the performance of the GP model with artificial neural network (ANN) models and the traditional PM (hourly PM-FAO56) method for estimation of ET_{ref}, finding that both of the data-driven models, GP and ANN, performed better than the PM method, and the performance of the GP and ANN models was comparable. Moreover, Guven et al. (2007) presented GP as a new tool for the estimation of ET_{ref} by using daily climatic variables obtained from the CIMIS database. Compared to seven conventional ET_{ref} models, the new equation produced quite satisfactory results and can be used as an alternative to the conventional models. Kisi and Guven (2010) investigated the

accuracy of linear genetic programming (LGP), an extension of the GP technique, in daily ET_{ref} modeling using the PM-FAO56 equation. Comparison of the accuracy of the LGP model with those of support vector regression (SVR), ANN, and four empirical models showed that the LGP model performs better than the other models in estimating total ET_{ref}.

As indicated above, the previously developed GEP models for estimating ET_{ref} were based on the PM-FAO56 equation. The objective of this study, therefore, was to investigate the accuracy of developed GEP models to predict daily ET_{ref} using the ET_{ref} standard values calculated by the more generalized PM equation in which plant height varies. Thirty-one years of meteorological data (1980–2010) collected from 19 stations in Saudi Arabia as a representative hyper-arid region were used. In addition, the contribution ratio of each climate parameter used in the calculation of daily ET_{ref} was determined.

Materials and methods

Study area and climatic data

The study was carried out in the Kingdom of Saudi Arabia (KSA), a country situated in the far southwest corner of Asia (Fig. 1). KSA is located between latitudes 16° 22′ 46″ N and 32° 14′ 00″ N and longitudes 34° 29′ 30″ E and 55° 40′ 00″ E. KSA is divided into 13 provinces (Fig. 1). The study covered the 13 provinces, arranged in descending order by area in Table 1. The climate of KSA varies from region to region depending on the terrain. Generally, the climate is characterized by hot summers, cold winters, and winter rainfall. The central areas undergo hot and dry summers and cool and dry winters, and the humidity in the coastal areas is high.

For this study, climatic data were collected from 19 meteorological stations from the 13 provinces of KSA. Nearly half of the provinces were presented by two stations including eastern region, Al-Riyadh, Al-Madinah, Makkah, Tabuk, and northern borders (Fig. 1). The climate data provided by the Presidency of Meteorology and Environment covered 31 years of daily meteorological information recorded for the period from 1980 to 2010. Recorded data for all stations included maximum, minimum, and mean air temperature (T_x, T_n, and T_a); maximum, minimum, and mean relative humidity (Rh_x, Rh_n, and Rh_a); wind speed measured at a 2-m height (U_2); and solar radiation (R_s). Descriptions of the different meteorological stations and the average daily climatic data from each station over the study period appear in Table 1. The eight climatic variables, viz., T_x, T_n, T_a, Rh_x, Rh_n, Rh_a, U_2, and R_s, in addition to the reference plant height (h_c), which varied from 5 to 105 cm in increments of 5 cm, were the maximum input set variables represented by the GEP models. In the development of the GEP models, 65 % of the climatic data from the 13 stations (one station from each province) was used for the training process and 35 % was used for the testing process. In contrast, the climatic data from the other six stations were used for a spatial assessment of the eight developed GEP models, which was made with h_c varying from 5 to 105 cm and constant at 50 and 12 cm. The combinations of climatic data represented in Table 2 formed the eight GEP models for estimation of ET_{ref}. The GEP models were given the code GEP-MOD*k*, where *k* represents the model number. As an example, GEP-MOD1 represents the GEP model that only includes temperature, whereas GEP-MOD8 represents the GEP model that includes all climatic variables. Of note, two main variables including air temperature and plant height were used as inputs in all GEP models application (Table 2).

Output/targeted data of GEP models

Because ET_{ref} is the output variable of the GEP models, a reference value of ET_{ref} is needed during the development stage, in addition to being a value against which the GEP models must be compared during assessment. The PM method is considered the standard when no measured lysimeter data is available (Irmak et al. 2003; Gavilan et al. 2006). This method gives optimal results over all climatic zones (De Souza and Yoder 1994; Chiew et al. 1995; Hupet and Vanclooster 2001; Naoum and Tsanis 2003; Irmak et al. 2003; Alazba 2004; Gavilan et al. 2006) and has two advantages over many other mathematical equations. First, it can be used globally with no local calibration required due to its physical basis. Second, it is a well-documented equation validated with a significant amount of lysimeter data (Trajkovic 2010). Many researchers (Kumar et al. 2002; Trajkovic and Kolakovic 2010; Kisi and Ozturk 2007; Zanetti et al. 2007) have used the PM equation as a reference and standard equation to evaluate the results of their mathematical models. Therefore, for this study, the daily ET_{ref} values obtained with the PM equation were used as output/target variables in the GEP models. A generalized form of the PM equation can be written as (Alazba 2004)

$$ET_{ref} = \lambda^{-1}\left[\frac{\Delta}{\Delta+\gamma^*}(R_n - G) + \frac{\gamma}{\Delta+\gamma^*}K(e_s - e_a)\right] \quad (1)$$

where

ET_{ref} = reference evapotranspiration (mm day^{-1})
λ = latent heat of vaporization (MJ kg^{-1})
Δ = slope of the saturation vapor pressure-temperature curve at mean air temperature (kPa °C^{-1})
γ = psychometric constant (kPa °C^{-1})
R_n = net radiation (MJ m^{-2} day^{-1})
G = soil heat flux (MJ m^{-2} day^{-1})
γ^* = modified psychometric constant (kPa °C^{-1})
K = parameter equal to $1.854 \times 10^5 \frac{\lambda/r_a}{T+273}$ (MJ m^{-2} day^{-1}°kPa^{-1})

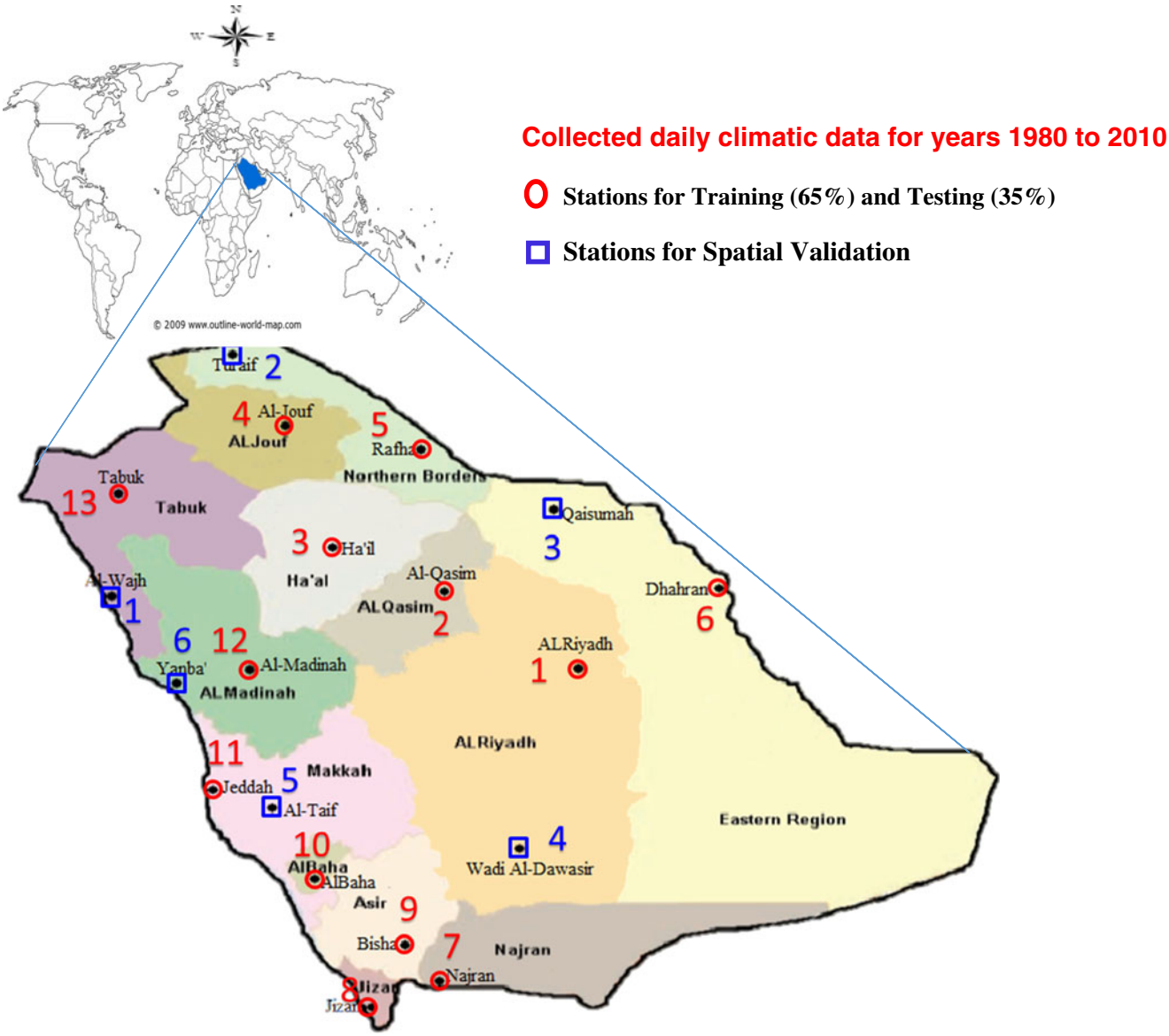

Fig. 1 Map of Saudi Arabia showing provinces and meteorological sites

r_a = aerodynamic resistance (s m^{-1}), a function of U_2 and h_c
T = air temperature (°C)
e_s = saturation vapor pressure at air temperature (kPa)
e_a = actual vapor pressure (kPa)

Description of GEP

GEP is a new evolutionary artificial intelligence technique developed by Ferreira (2001a). There are two main factors in GEP (Ferreira 2006): chromosomes, which are usually composed of more than one gene of equal length, and expression trees or programs, which are the expressions of the genetic information encoded in chromosomes. The chromosomes are composed of multiple genes, each gene encoding a smaller subprogram. In GEP, linear structures, or the chromosomes, represent the genotype, and branched structures, or expression trees, represent the phenotype (Ferreira 2001b). GEP computer programs are all encoded in linear chromosomes, which are later translated or expressed in expression trees. These programs are usually developed to solve a particular problem and are selected according to their ability to solve that problem (Guven and Aytek 2009).

Development of the GEP model

Initially, the training set is selected from the whole data, and the testing set uses the remaining data. The modeling

Table 1 Meteorological station sites and average daily values of the climatic parameters

Provinces	Areas[a] (km^2)	Stations	Location coordinates			Climatic parameters							
			Longitude (°)	Latitude (°)	Altitude (m)	T_x (°C)	T_n (°C)	T_a (°C)	Rh_x (%)	Rh_n (%)	Rh_a (%)	U_2 (m s^{-1})	R_s (MJ m^{-2} day^{-1})
Eastern region	540	Qaisumah	46.13	28.31	355	32	19	25	77	30	50	2.6	21
		Dhahran	50.20	26.30	17	33	20	26	75	29	52	4.2	20
Al-Riyadh	380	Riyadh (north)	46.72	24.93	614	33	20	26	38	16	31	3.9	15
		Wadi Al-Dawasir	45.20	20.50	617	35	22	28	35	17	26	3.4	18
Al-Madinah	150	Al-Madina	39.60	24.47	619	33	25	19	56	29	44	4.2	26
		Yanba'	38.10	24.10	1	29	22	17	78	23	50	3.2	29
Makkah	137	Jeddah	39.17	21.40	12	34	28	22	81	37	60	2.6	23
		Al-Ta'if	40.50	21.50	1449	35	29	23	60	29	39	3.2	27
Tabuk	136	Tabuk	36.58	28.38	770	29	14	22	53	17	32	2.9	33
		Al-Wajh	36.50	26.20	20	28	10	18	70	22	45	2.2	29
Najran	130	Najran	44.40	17.60	1214	35	29	25	60	33	44	3.5	28
Ha'il	120	Ha'il	41.70	27.40	1013	34	28	22	81	37	60	2.3	14
Northern borders	104	Turaif	38.65	31.68	854	35	29	23	60	29	39	3.3	29
		Rafha	43.50	29.60	447	29	14	22	53	17	32	2.9	22
Al-Jouf	85	Al-Jouf	40.10	29.80	689	30	14	22	48	18	31	3.11	25
Asir	80	Bisha	42.60	20.00	1157	33	17	25	47	15	29	2.4	28
Al-Qasim	73	Al-Qasim	43.80	26.30	650	32	18	25	44	30	18	2.9	27
Jizan	13	Jizan	42.60	16.88	3	36	30	25	61	34	44	3.3	36
Al-Bahah	12	Al-Baha	41.60	20.30	1656	29	16	22	56	22	38	1.3	28

[a] Saudi Geological Survey (2012), Kingdom of Saudi Arabia: Facts and Numbers, edition 1

Table 2 The input combinations for forming GEP models

Models	Input parameters								
	Air temperature (°C)			Relative humidity (%)			Wind speed (m s^{-1})	Solar Radiation (MJ m^{-2} day^{-1})	Crop Height (cm)
	T_x	T_n	T_a	Rh_x	Rh_n	Rh_a	U_2	R_s	h_c
GEP-MOD1	✓	✓	✓						✓
GEP-MOD2	✓	✓	✓	✓	✓	✓			✓
GEP-MOD3	✓	✓	✓				✓		✓
GEP-MOD4	✓	✓	✓					✓	✓
GEP-MOD5	✓	✓	✓	✓	✓	✓	✓		✓
GEP-MOD6	✓	✓	✓	✓	✓	✓		✓	✓
GEP-MOD7	✓	✓	✓				✓	✓	✓
GEP-MOD8	✓	✓	✓	✓	✓	✓	✓	✓	✓

comprises five major steps in the use of GEP, the details of which can be found in Ferreira (2001a, b). The main purpose in developing the GEP models is to generate the mathematical functions for the prediction of ET_{ref}. A set of preparatory model runs were carried out to test the performance of the models using four possible function sets. One function set was selected for continued use. All of these procedures were performed on GEP-MOD1, which contained the least number of variables, using the root mean square error (RMSE) fitness function and the addition linking function. The results of the investigation of function selection presented in Table 3 show that the operator function set F4 outperformed all of the other structures. The superiority of the F4 function set confirms the results of Shiri et al. (2012) and Kisi and Guven (2010), who concluded that the GeneXpro default operator function set (F4) performed better than other applied function sets for estimating the daily ET_{ref} at four weather stations in the Basque Country, northern Spain, and for estimating the daily suspended sediment load at two stations on the Cumberland River, USA. Despite these results, the F3 function set was used in the GEP models to develop less complicated mathematical equations and avoid the use of trigonometric functions (sin(x), cos(x), and arctan(x)). There were only slight differences between the F3 and F4 function sets. The parameters used in training the eight models are given in Table 4. The program was run until there was no longer a significant improvement in the performance of the models. The maximum number of generations and best fitness generated during training are listed in Table 5. In this study, the general formulation of the GEP models takes the following form:

$$ET_{ref(\text{GEP-MOD}k)} = \sum_{j=1}^{N} \text{gene}j \tag{2}$$

where N is the number of genes.

Performance criteria of the GEP models

Several statistical parameters can be used for comparison between referenced and estimated ET_{ref} values. The following statistical parameters were used in this study:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(E_i - C_i)^2}{n}} \tag{3}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|E_i - C_i|}{n} \tag{4}$$

$$R^2 = \frac{\left(\sum_{i=1}^{n}\left(E_i - \overline{E}\right)\left(C_i - \overline{C}\right)\right)^2}{\sum_{i=1}^{n}\left(E_i - \overline{E}\right)^2 . \sum_{i=1}^{n}\left(C_i - \overline{C}\right)^2} \tag{5}$$

Table 3 Preliminary selection of basic functions for the expression tree using the scatter index

	Functions	RMSE (mm day^{-1})	R^2 (%)
F1	+, −, ×, ÷	3.45	61.3
F2	+, −, ×, ÷, sqrt, exp	3.23	63.2
F3	*+, −, ×, ÷, sqrt, exp, Ln, X^2*	*3.10*	*64.4*
F4	+, −, ×, ÷, −, ×, ÷, sqrt, 3Rt, exp, Ln, X^2, X^3, sin(x), cos(x), arctan(x)	2.94	65.7

Italic values are for the least, moderate and most accurate GEP-MODs

Table 4 Parameters of the GEP models

Parameter	Description of parameter	Parameter setting
P1	Number of chromosomes	30
P2	Number of genes	3 (GEP-MOD1 to GEP-MOD7), 4 (GEP-MOD8)
P3	Head size	8 (GEP-MOD1 to GEP-MOD7), 10 (GEP-MOD8)
P4	Function set	+, −, ×, ÷, sqrt, exp, Ln, X^2
P5	Linking function	Addition
P6	Fitness function	Mean squared error
P7	Mutation rate	0.00206
P8	Inversion rate	0.00546
P9	One-point recombination rate	0.00277
P10	Two-point recombination rate	0.00277
P11	Gene recombination rate	0.00277
P12	Gene transposition rate	0.00277

where

E_i = value of ET_{ref} calculated by the PM model
C_i = corresponding ET_{ref} value estimated by the GEP models
n = number of observations
$\bar{E}$ = average of the calculated values
$\overline{C}$ = average of the estimated values

Results and discussion

This section discusses the eight developed GEP models with additional details oriented to three specific GEP models, GEP-MOD1, GEP-MOD8, and GEP-MOD5. The first model consisted of only one climatic parameter, temperature, whereas the second model contained all climatic parameters. The latter model produced results in close agreement with GEP-MOD8, even though the R_s parameter was not included. The spatial assessment of the eight developed GEP models is also discussed. The results of the importance analysis of each of the input parameters are presented.

Developed GEP models

Figures 2 and 3, which depict the eight developed GEP models for the training and testing stages, respectively, clearly illustrate the good agreement between the training and testing GEP models. The statistical parameters shown in Table 5

Table 5 Statistical performance of the developed GEP models for the training and testing processes

Model	Input combination	Training			Testing		
		RMSE (mm day^{-1})	MAE (mm day^{-1})	R^2 (%)	RMSE (mm day^{-1})	MAE (mm day^{-1})	R^2 (%)
GEP-MOD1 (209,208, 782)	T_x, T_n, T_a, h_c	*3.10*	*2.29*	*64.4*	*3.19*	*2.40*	*63.6*
GEP-MOD2 (260,672, 794)	T_x, T_n, T_a, Rh_x, Rh_n, Rh_a, h_c	2.85	2.07	72.2	2.94	2.17	71.0
GEP-MOD3 (207,489, 791)	T_x, T_n, T_a, U_2, h_c	2.64	1.98	76.1	2.65	1.98	76.8
GEP-MOD4 (170,031, 805)	T_x, T_n, T_a, R_s, h_c	2.92	2.11	68.3	2.99	2.21	67.9
GEP-MOD5 (202,825, 831)	T_x, T_n, T_a, Rh_x, Rh_n, Rh_a, U_2, h_c	*1.98*	*1.46*	*89.7*	*2.09*	*1.53*	*89.3*
GEP-MOD6 (104,139, 814)	T_x, T_n, T_a, Rh_x, Rh_n, Rh_a, R_s, h_c	2.48	1.73	77.5	2.52	1.81	77.6
GEP-MOD7 (199,602, 845)	T_x, T_n, T_a, U_2, R_s, h_c	2.20	1.64	82.2	2.21	1.63	82.6
GEP-MOD8 (124,547, 915)	T_x, T_n, T_a, Rh_x, Rh_n, Rh_a, U_2, R_s, h_c	*1.12*	*0.83*	*95.5*	*1.14*	*0.83*	*95.4*

The numbers between parentheses are the maximum number of generations and the best fitness, respectively
Italic values are for the least, moderate and most accurate GEP-MODs

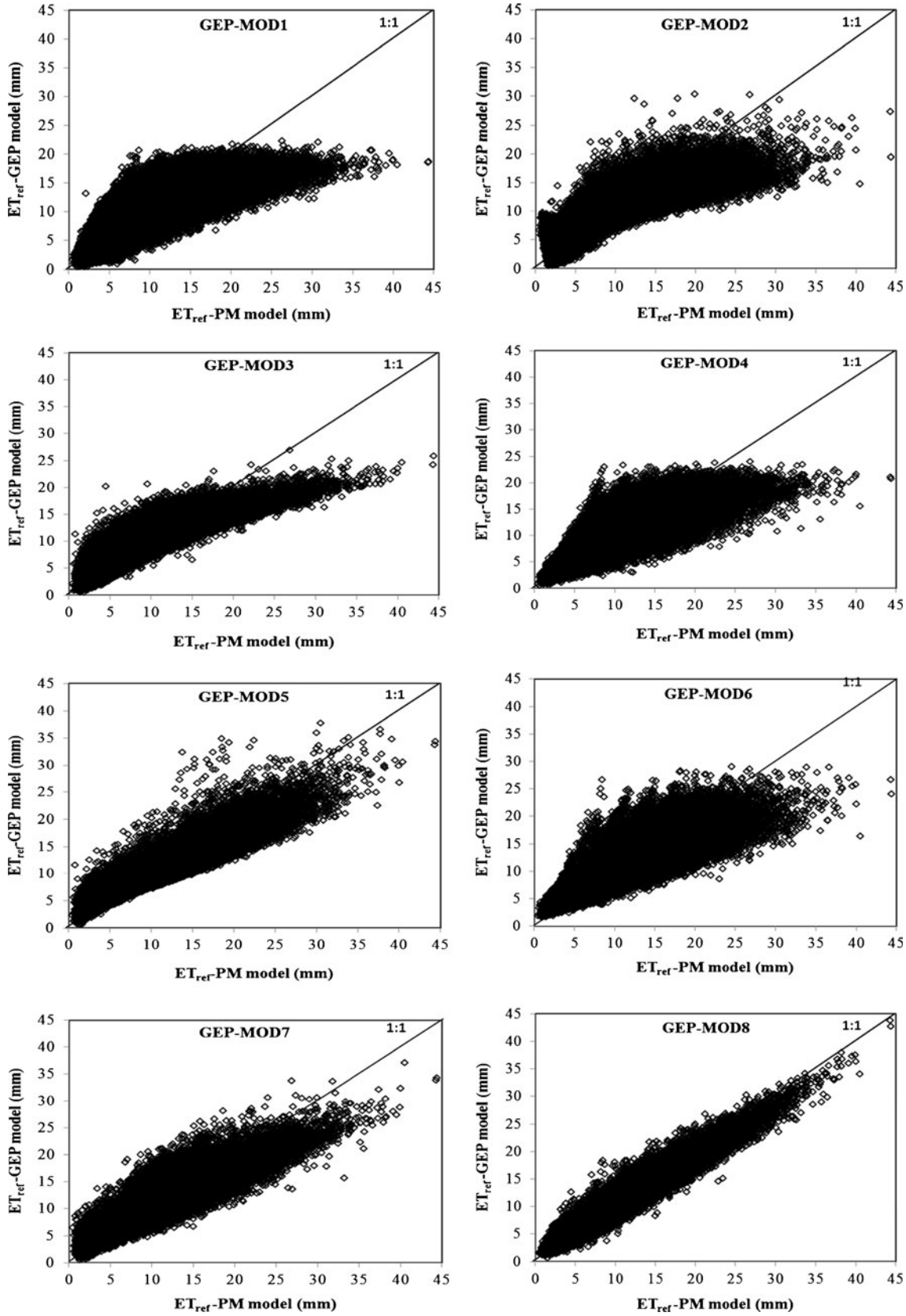

Fig. 2 Estimated daily ET_{ref} by GEP versus corresponding calculated daily ET_{ref} by the generalized PM model, training process

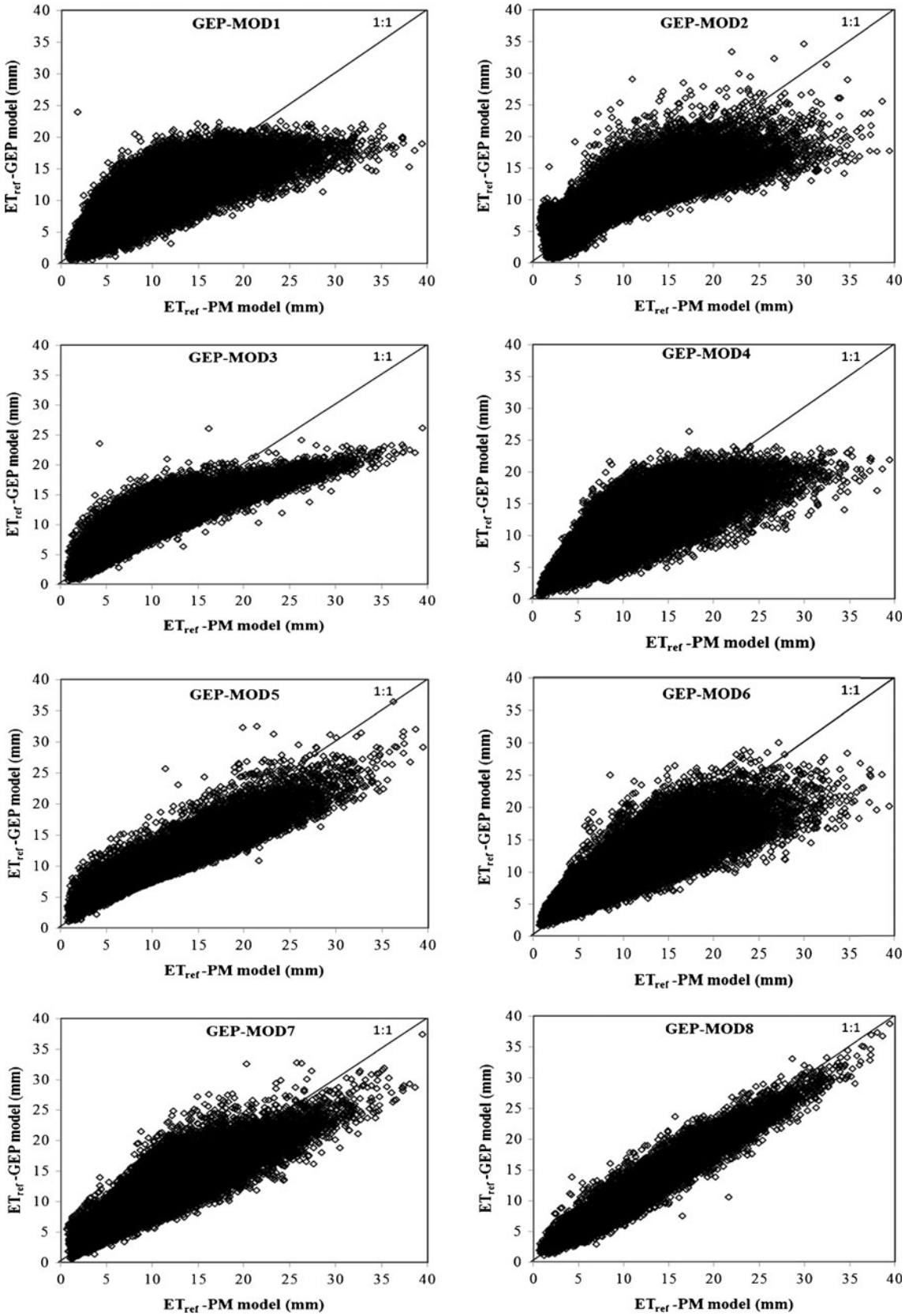

Fig. 3 Estimated daily ET_{ref} by GEP versus corresponding calculated daily ET_{ref} by the generalized PM model, testing process

indicate that the values of R^2 for each GEP model during the training process are nearly identical to those of the testing process. The R^2 magnitudes in the training process range from 64.4 % for GEP-MOD1 to 95.5 % for GEP-MOD8, whereas they equal 63.3 and 95.4 %, respectively, for the training process. The same can be stated for the values of the RMSE and mean absolute error (MAE) parameters. The values of RMSE range from as low as 1.12 and 1.14 mm day^{-1} to as high as 3.10 and 3.19 mm day^{-1} for the training and testing processes, respectively (Table 5). Also from Table 5, the MAE values range from 2.29 to 0.83 mm day^{-1} and from 2.40 to 0.83 mm day^{-1} for the respective training and testing processes. The RMSE and MAE values for the training and testing processes clearly show that the GEP models can successfully estimate ET_{ref} with close agreement that depends on the climatic parameters included in the GEP model.

As indicated in Fig. 2 and Table 5, the least accurate GEP model for the prediction of ET_{ref} is GEP-MOD1. This is not surprising as temperature is the only climatic parameter included in the model. In contrast, the most accurate GEP model for ET_{ref} estimation is GEP-MOD8, implying that when all climatic parameters are included, GEP can be used with precision. The inclusion of R_s is not essential to predict ET_{ref} using GEP modeling, as shown with GEP-MOD5. The one-to-one curve depicted in Fig. 2 and the values of R^2, RMSE, and MAE listed in Table 5 indicate that GEP-MOD5 is in good agreement with GEP-MOD8. This result may lead to the interesting conclusion that ET_{ref} can be estimated highly accurately in hyper-arid conditions even when R_s is unavailable.

Because the ultimate outcome is to predict ET_{ref} with the GEP models and for the previously stated justification, the next section presents the mathematical expressions of three GEP models, GEP-MOD1, GEP-MOD5, and GEP-MOD8.

Expression trees and mathematical equations of three GEP models

As shown in Fig. 4, GEP-MOD1 and GEP-MOD5 each contain three genes and GEP-MOD8 contains four genes, and each gene has constants that differ in number and magnitude. These coefficients and their magnitudes are shown in Table 6. From Fig. 4, the algebraic formulations for GEP-MOD1, GEP-MOD5, and GEP-MOD8 can easily be attained.

The algebraic gene expressions for GEP-MOD1 are

$$\text{gene1} = \frac{\left(\frac{T_x}{C_2} - T_n\right) - (T_n + T_a)}{(C_3 - C_1) + \frac{T_x}{h_c}} \tag{6}$$

$$\text{gene2} = \frac{[(h_c - T_x) - T_x]}{\frac{C_1}{h_c}} + (C_3 - h_c) + \frac{T_a}{C_2} \tag{7}$$

$$\text{gene3} = \frac{[(C_3 - T_x) - C_1] - (T_a + h_c)}{(C_2 - h_c) + (T_a - C_1)} \tag{8}$$

Those for GEP-MOD5 are

$$\text{gene1} = \frac{[(U_2 - \text{Rh}_a) \times T_x]/(\text{Rh}_x + C_2)}{\left(\frac{C_1}{U_2}\right) \times \text{Rh}_a} \tag{9}$$

$$\text{gene2} = \sqrt{T_x - \frac{(U_2 \times C_1) \times (h_c \times T_a)}{\text{Rh}_n}} \tag{10}$$

$$\text{gene3} = \left[\left(\left(h_c \times U_2\right) - C_2\right) - \frac{C_1}{U_2}\right] + \frac{2T_n}{\text{Rh}_x} \tag{11}$$

and those for GEP-MOD8 are

$$\text{gene1} = \frac{[C_1 + (h_c \times T_a)] - (U_2 + R_s)}{C_2 - 2U_2} \tag{12}$$

$$\text{gene2} = \frac{(T_a \times C_1) + C_3}{(C_4 + \text{Rh}_n) \times (C_2 - \text{Rh}_x)} + U_2 \tag{13}$$

$$\text{gene3} = C_1 - \frac{U_2}{C_2^2/[(\text{Rh}_x - T_x) - R_s]} \tag{14}$$

$$\text{gene4} = \frac{(T_a \times h_c) \times T_n}{[(\text{Rh}_a - U_2) - C_1] + (T_a - U_2)} \tag{15}$$

Recalling Eq. 2, the mathematical equations for ET_{ref} estimates for these three GEP models can be expressed as follows.

For GEP-MOD1,

$$ET_{ref} = \frac{\left(\frac{T_x}{C_2} - T_n\right) - (T_n + T_a)}{(C_3 - C_1) + \frac{T_x}{h_c}} + \frac{[(h_c - T_x) - T_x]}{\frac{C_1}{h_c}} + (C_3 - h_c) + \frac{T_a}{C_2} + \frac{[(C_3 - T_x) - C_1] - (T_a + h_c)}{(C_2 - h_c) + (T_a - C_1)} \tag{16}$$

and Eq. 16 can be written by replacing the values of the constants listed in Table 6 as follows:

$$ET_{ref} = \frac{\left(\frac{T_x}{-5.24} - T_n\right) - (T_n + T_a)}{(5.13 - 5.58) + \frac{T_x}{h_c}} + \frac{[(h_c - T_x) - T_x]}{\frac{-6.13}{h_c}} + (0.0687 - h_c) + \frac{T_a}{3.57} + \frac{[(8.98 - T_x) - 2.17] - (T_a + h_c)}{(-0.69 - h_c) + (T_a + 2.17)} \tag{17}$$

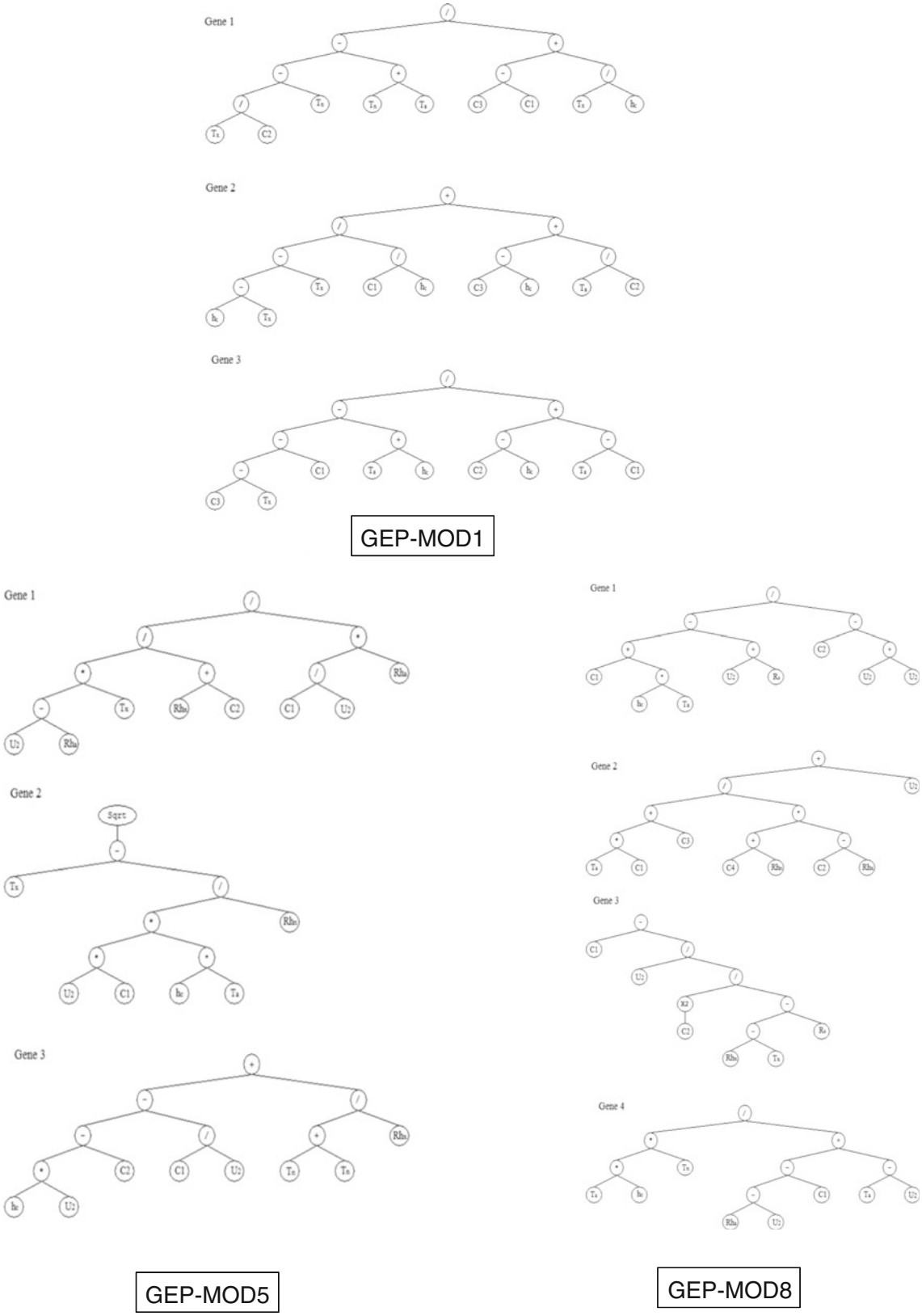

Fig. 4 Expression tree for three developed GEP models for daily ET_{ref} estimates

Table 6 Constants for GEP-MOD5 and GEP-MOD8 formulations

Constant	Gene1	Gene2	Gene3	
GEP-MOD1				
C1	5.58	−6.13	−2.17	
C2	−5.24	3.57	−0.69	
C3	5.13	0.0687	8.98	
GEP-MOD5				
C1	−4.70	−8.46	0.37	
C2	−0.78	–	1.63	
C3	–	–	–	
GEP-MOD8				
C1	10.40	5.30	1.80	−14.00
C2	−6.35	1.80	8.66	–
C3	–	−1.6	–	–
C4	–	7.40	–	–

For GEP-MOD5,

$$\mathrm{ET_{ref}} = \frac{[(U_2 - \mathrm{Rh_a}) \times T_\mathrm{x}]/(\mathrm{Rh_x} + C_2)}{\left(\frac{C_1}{U_2}\right) \times \mathrm{Rh_a}} + \sqrt{T_\mathrm{x} - \frac{(U_2 \times C_1) \times (h_\mathrm{c} \times T_\mathrm{a})}{\mathrm{Rh_n}}} + \left[\left(\left(h_\mathrm{c} \times U_2\right) - C_2\right) - \frac{C_1}{U_2}\right] + \frac{2T_\mathrm{n}}{\mathrm{Rh_x}} \tag{18}$$

and by replacing the values of the constants in Table 6, Eq. 18 takes the following form:

$$\mathrm{ET_{ref}} = \frac{[(U_2 - \mathrm{Rh_a}) \times T_\mathrm{x}]/(\mathrm{Rh_x} \pm 0.78)}{\left(\frac{-4.70}{U_2}\right) \times \mathrm{Rh_a}} + \sqrt{T_\mathrm{x} - \frac{(U_2 \times -8.46) \times (h_\mathrm{c} \times T_\mathrm{a})}{\mathrm{Rh_n}}} + \left[\left(\left(h_\mathrm{c} \times U_2\right) - 1.63\right) - \frac{0.37}{U_2}\right] + \frac{2T_n}{Rh_x} \tag{19}$$

For GEP-MOD8,

$$\mathrm{ET_{ref}} = \frac{[C_1 + (h_\mathrm{c} \times T_\mathrm{a})] - (U_2 + R_\mathrm{s})}{C_2 - 2U_2} + \frac{(T_\mathrm{a} \times C_1) + C_3}{(C_4 + \mathrm{Rh_n}) \times (C_2 - \mathrm{Rh_x})} + U_2 + C_1 - \frac{U_2}{C_2^2/[(\mathrm{Rh_x} - T_\mathrm{x}) - R_\mathrm{s}]} + \frac{(T_\mathrm{a} \times h_\mathrm{c}) \times T_\mathrm{n}}{[(\mathrm{Rh_a} - U_2) - C_1] + (T_\mathrm{a} - U_2)} \tag{20}$$

and by replacing the values of the constants given in Table 6, Eq. 20 can be rewritten as:

$$\mathrm{ET_{ref}} = \frac{[10.40 + (h_\mathrm{c} \times T_\mathrm{a})] - (U_2 + R_\mathrm{s})}{-6.35 - 2U_2} + \frac{(T_\mathrm{a} \times 5.30) - 1.60}{(7.40 + \mathrm{Rh_n}) \times (1.80 - \mathrm{Rh_x})} + U_2 + 1.80 - \frac{U_2}{8.66^2 \big/ [(\mathrm{Rh_x} - T_\mathrm{x}) - R_\mathrm{s}]} + \frac{(T_\mathrm{a} \times h_\mathrm{c}) \times T_\mathrm{n}}{[(\mathrm{Rh_a} - U_2) + 14] + (T_\mathrm{a} - U_2)} \tag{21}$$

Equation 21, which includes all climatic parameters in addition to h_c, is the best function obtained using GEP to estimate $\mathrm{ET_{ref}}$.

Spatial assessment of the GEP models

As previously mentioned, the GEP models were spatially evaluated using climatic data collected from six meteorological stations other than the 13 meteorological stations used for the development of the GEP models. The values of the statistical parameters RMSE, MAE, and R^2 are presented in Table 7, which shows the values for $h_{\mathrm{c5-105}}$, h_{c50}, and h_{c12}. In other words, the eight developed GEP models were evaluated for spatial assessment with h_c varying from 5 to 105 cm in 5-cm increments and with h_c equal to the two constant values ($h_\mathrm{c} = 12$ cm and $h_\mathrm{c} = 50$ cm). The constant h_c values were selected considering the two well-known PM formulations of PM-ASCE70 and PM-FAO56. The first considers h_c equal to 50 cm, and the latter uses h_c equal to 12 cm. With respect to the RMSE, MAE, and R^2 statistics shown in Table 7, the eight GEP models can perfectly predict $\mathrm{ET_{ref}}$ at different values of h_c regardless of the accuracy of each GEP model. Nevertheless, the best GEP model is GEP-MOD8 because it gives accurate $\mathrm{ET_{ref}}$ estimates (Table 7) at different values of h_c. This can be seen more clearly by referring to Fig. 4, which depicts the relationship of $\mathrm{ET_{ref}}$ calculated by Eq. 1, the generalized PM formula, with $\mathrm{ET_{ref}}$ predicted by the GEP-MOD8 model.

Equation 21 indicates that the variable h_c exists in the first and fourth gene expressions. Accordingly, Eq. 21 can be expressed for the two well-known PM-type formulas of PM-ASCE70 and PM-FAO56 by simply replacing h_c with 50 and 12 cm, respectively, in both genes. Along with arithmetic elimination and algebraic manipulation, Eq. 21 can be

Table 7 Statistical performance of GEP models for the assessment process in six weather stations within the period from 1980 to 2010

Model	Input combination	h_{c5-105}			h_{c50} [a]			h_{c12} [b]		
		RMSE (mm day^{-1})	MAE (mm day^{-1})	R^2 (%)	RMSE (mm day^{-1})	MAE (mm day^{-1})	R^2 (%)	RMSE (mm day^{-1})	MAE (mm day^{-1})	R^2 (%)
GEP-MOD1	T_x, T_n, T_a, h_c	*3.27*	*2.35*	*64.0*	*2.95*	*2.21*	*58.9*	*1.73*	*1.29*	*66.1*
GEP-MOD2	T_x, T_n, T_a, Rh_x, Rh_n, Rh_a, h_c	3.02	2.04	73.6	2.55	1.81	71.5	1.75	1.27	69.3
GEP-MOD3	T_x, T_n, T_a, U_2, h_c	3.04	2.31	71.3	2.70	2.18	72.4	1.45	1.12	78.3
GEP-MOD4	T_x, T_n, T_a, R_s, h_c	3.17	2.26	65.4	2.86	2.16	60.8	1.62	1.16	70.8
GEP-MOD5	T_x, T_n, T_a, RH_x, RH_n, RH_a, U_2, h_c	*2.15*	*1.62*	*87.3*	*1.81*	*1.42*	*88.3*	*1.28*	*1.02*	*85.4*
GEP-MOD6	T_x, T_n, T_a, Rh_x, Rh_n, Rh_a, R_s, h_c	2.54	1.74	81.0	2.30	1.62	78.4	1.46	1.05	82.0
GEP-MOD7	T_x, T_n, T_a, U_2, R_s, h_c	2.63	1.97	78.4	2.25	1.73	78.9	1.77	1.26	76.5
GEP-MOD8	T_x, T_n, T_a, RH_x, RH_n, RH_a, U_2, R_s, h_c	*1.21*	*0.89*	*95.6*	*0.75*	*0.55*	*97.6*	*0.96*	*0.76*	*91.9*

[a] Reference values for ET_{ref} were calculated using PM-ASCE70

[b] Reference values for ET_{ref} were calculated using PM-FAO56

Italic values are for the least, moderate and most accurate GEP-MODs

written for alfalfa-based and grass-based regions as follows:

$$ET_{ref} = 1.8 + U_2 + 0.013U_2[T_x + R_s - Rh_x] + \frac{5.3T_a - 1.6}{(7.40 + Rh_n) \times (1.80 - Rh_x)} - \frac{0.5T_nT_a}{2U_2 - T_a - Rh_a - 14} - \frac{10.4 + 0.5T_a - U_2 - R_s}{6.35 + 2U_2} \quad (22)$$

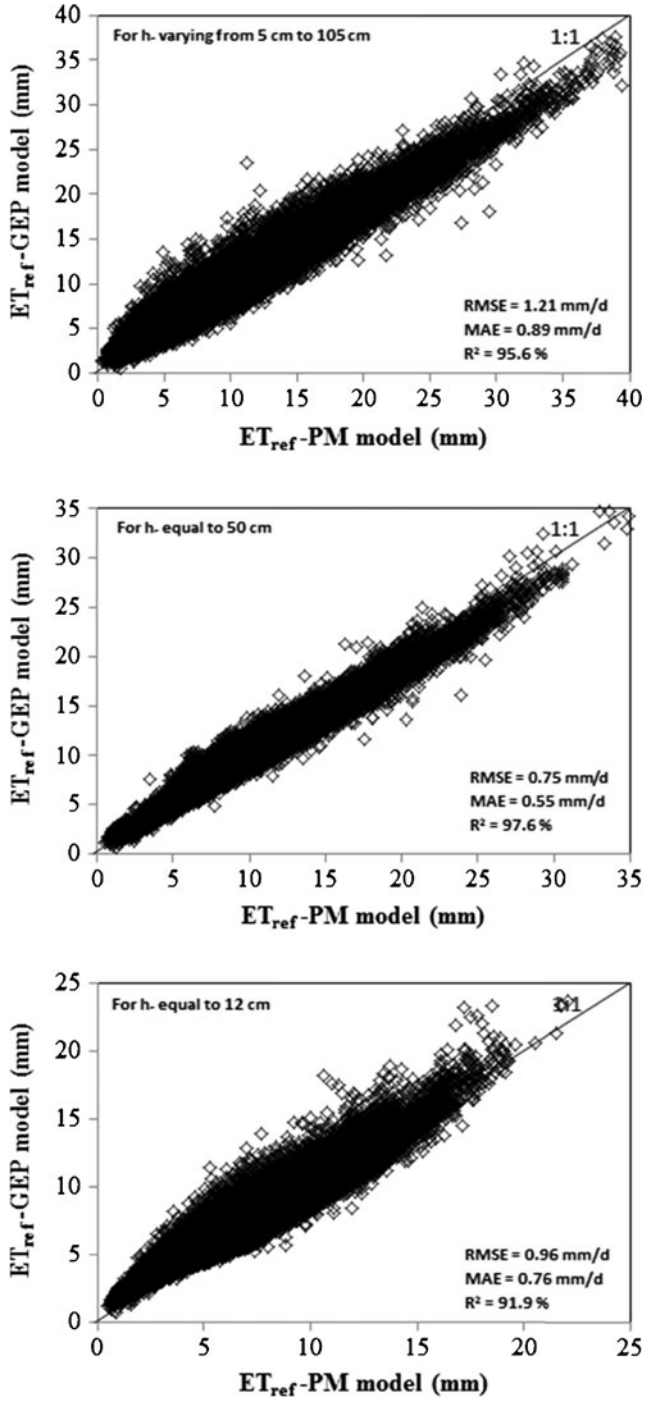

Fig. 5 Spatially assessed ET_{ref} at different h_c considerations for GEP-MOD8

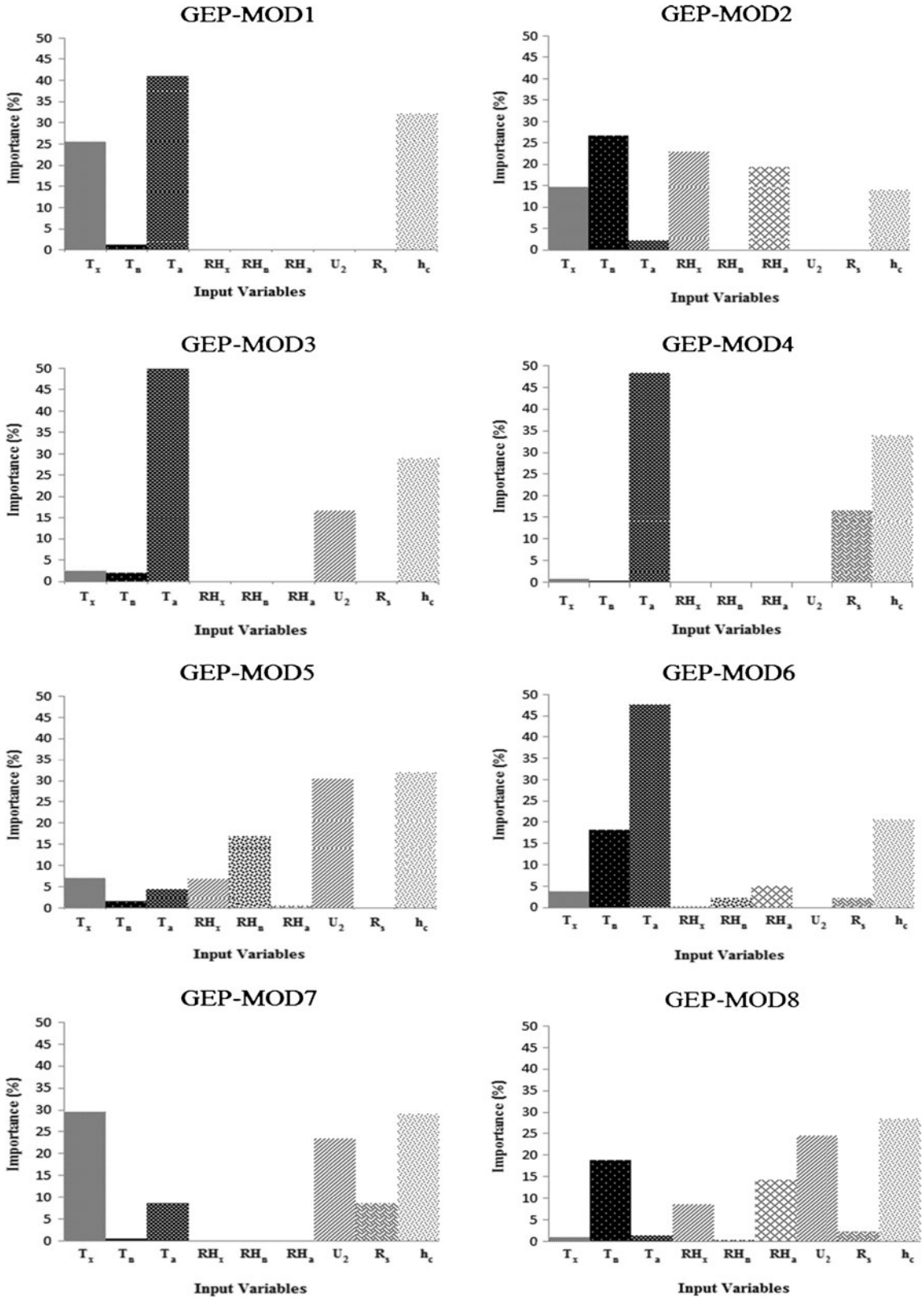

Fig. 6 Importance ratio analysis of input variables, climatics, and h_c, for GEP models

for alfalfa-based and

$$\mathrm{ET_{ref}} = 1.8 + U_2 + 0.013U_2[T_\mathrm{x} + R_\mathrm{s} - \mathrm{Rh_x}] + \frac{5.3T_\mathrm{a} - 1.6}{(7.40 + \mathrm{Rh_n}) \times (1.80 - \mathrm{Rh_x})} - \frac{0.12T_\mathrm{n}T_\mathrm{a}}{2U_2 - T_\mathrm{a} - \mathrm{Rh_a} - 14} - \frac{10.4 + 0.12T_\mathrm{a} - U_2 - R_\mathrm{s}}{6.35 + 2U_2} \quad (23)$$

for grass-based regions.

The one-to-one curves in Fig. 5 show good agreement of the $\mathrm{ET_{ref}}$ values obtained by GEP-MOD8 and by the generalized PM method for the three cases of h_c, 5 to 105, 50, and 12 cm.

Importance ratio analysis of input variables

Importance ratio analysis was also performed using GeneXproTools software to examine the climatic parameters and the h_c variable with the most influence on the eight GEP models. Figure 6 shows that h_c has an almost identical effect on $\mathrm{ET_{ref}}$ for all GEP models at a rate of approximately 15–30 %. The temperature effect on $\mathrm{ET_{ref}}$ is higher when the Rh parameters and/or U_2 are included, either as in the case of GEP-MOD2, GEP-MOD3, GEP-MOD6, and GEP-MOD7 or when both are included, as with GEP-MOD1 and GEP-MOD2. In general, Fig. 6 indicates that the parameters of Rh and R_s have the least effect on $\mathrm{ET_{ref}}$. In contrast, the U_2 parameter has a significant effect on $\mathrm{ET_{ref}}$, particularly with the absence of the R_s parameter in GEP-MOD5. A final conclusion on the effects of the input parameters on $\mathrm{ET_{ref}}$ may not be reached without keeping some parameters constant while varying others and/or without the use of special statistical tools such as cross-product or suitably and appropriately similar tools. Therefore, an ultimate judgment on the effect of input parameters on $\mathrm{ET_{ref}}$ will require further investigation, which is beyond the scope of this work.

Conclusions

Eight GEP models were developed to estimate the daily $\mathrm{ET_{ref}}$ under hyper-arid environmental conditions. A combination of climatic parameters formed the eight GEP models. The input variables to the developed GEP models included maximum, minimum, and average air temperature; maximum, minimum, and average relative humidity; and wind speed, solar radiation, and plant height. The climatic data were collected from 19 meteorological stations in KSA covering the 1980 to 2010 period. Data obtained from the 13 meteorological stations were used to develop the GEP models, and data from the other six stations were used for spatial validation of the eight developed GEP models. In comparison with the PM model, the best-performing GEP model, GEP-MOD8, included all climatic parameters, whereas the worst-performing GEP model included temperature as the only climatic input parameter. Compared to the PM model, GEP-MOD8 with its algebraic equation can perfectly predict $\mathrm{ET_{ref}}$ for a wide range of h_c varying from 5 to 105 cm. Interestingly, GEP-MOD5, the GEP model without R_s, competed favorably with GEP-MOD8, the most accurate model, to predict the daily $\mathrm{ET_{ref}}$ in a hyper-arid environment. This interesting outcome leads to a possible conclusion that the climatic parameter U_2 is more important than R_s within the circumstances of this study. The importance analysis of the input parameters was inconclusive, indicating that the ultimate judgment of the effects of various input variables on $\mathrm{ET_{ref}}$ requires further investigation and in-depth analysis.

Acknowledgments The project was financially supported by King Saud University, Vice Deanship of Research Chairs. The climatic data used in this study were obtained from the Presidency of Meteorology and Environment (PME), Kingdom of Saudi Arabia. The authors highly appreciate this effort and would like thank the PME for its continued collaboration in providing the data.

References

Alazba AA (2004) Estimating palm water requirements using Penman-Monteith mathematical model. J King Saud Univ 16(2):137–152

Allen R. G, Pereira L. S, Raes D. and Smith M 1998. Crop evapotranspiration guidelines for computing crop water requirements, FAO Irrigation and Drainage, Paper No. 56, Food and Agriculture Organization of the United Nations, Rome

ASCE-EWRI (2005) The ASCE standardized reference evapotranspiration equation. Environmental and Water Resources Institute of the American Society of Civil Engineers Task Committee on Standardization of Reference Evapotranspiration Calculation, Washington DC, USA, p 70

Azamathulla HM (2013) Gene-expression programming to predict friction factor for southern Italian rivers. Neural Comput Applic 23: 1421–1426

Azamathulla HM, Ahmad Z (2012) Gene-expression programming for transverse mixing coefficient. J Hydrol 434–435:142–148

Azamathulla HM, Ghani AA, Leow CS, Chang KC, Zakaria NA (2011) Gene-expression programming for the development of a stage-discharge curve of the Pahang River. Water Resour Manag 25: 2901–2916

Chattopadhyay S, Jain R, Chattopadhyay G (2009) Estimating potential evapotranspiration from limited weather data over Gangetic West Bengal, India: a neurocomputing approach. Meteorol Appl 16: 403–411

Chiew FHS, Kamaladassa NN, Malano HM, McMahon TA (1995) Penman-Monteith, FAO-24 reference crop evapotranspiration and class-A pan data in Australia. Agric Water Manag 28:9–21

De Souza F and Yoder R. E 1994. ET estimation in the northeast of Brazil: Hargreaves or Penman-Monteith equation. Proceedings, Technical Paper, ASAE International Winter Meeting, American Society of Agricultural Engineers, St. Joseph, MI

Fernando AK, Shamseldin AY, Abrahart RJ (2012) Use of gene expression programming for multimodel combination of rainfall-runoff models. J Hydrol Eng 17(9):975–985

Ferreira C 2001a. Gene expression programming in problem solving. 6th Online World Conference on Soft Computing in Industrial Applications (Invited Tutorial)

Ferreira C (2001b) Gene expression programming: a new adaptive algorithm for solving problems. Complex Syst 13(2):87–129

Ferreira C (2006) Gene expression programming: mathematical modeling by an artificial intelligence, 2nd edn. Springer, Germany

Fooladmand HR, Zandilak H, Ravanan MH (2008) Comparison of different types of Hargreaves equation for estimating monthly evapotranspiration in the south of Iran. Arch Agron Soil Sci 54:321–330

Gavilan P, Lorite IJ, Tornero S, Berengena J (2006) Regional calibration of Hargreaves equation for estimating reference ET in a semiarid environment. Agric Water Manag 81:257–281

Gavilan P, Berengena J, Allen RG (2007) Measuring versus estimating net radiation and soil heat flux: impact on Penman-Monteith reference ET estimates in semiarid regions. Agric Water Manag 89:275–286

George BA, Reddy BRS, Raghuwanshi NS, Wallender WW (2002) Decision support system for estimating reference evapotranspiration. J Irrig Drain Eng 128(1):1–10

Guven A, Aytek A (2009) A new approach for stage-discharge relationship: gene-expression programming. J Hydrol Eng 14(8):812–820

Guven A, Aytek A, Yuce MI, Aksoy H (2007) Genetic programming-based empirical model for daily reference evapotranspiration estimation. CLEAN-Soil, Air, Water J 36:10–11

Hupet F, Vanclooster M (2001) Effect of the sampling frequency of meteorological variables on the estimation of the reference evapotranspiration. J Hydrol 243:192–204

Irmak S, Irmak A, Allen RG, Jones JW (2003) Solar and net radiation based equations to estimate reference evapotranspiration in humid climates. J Irrig Drain Eng ASCE 129(5):336–347

Jensen ME, Burman RD, Allen RG (1990) Evapotranspiration and irrigation water requirements. ASCE Manuals and Reports on Engineering Practice No. 70. American Society of Civil Engineers, New York, pp 10017–2398

Kisi O (2007) Evapotranspiration modelling from climatic data using a neural computing technique. Hydrol Process 21:1925–1934

Kisi O, Guven A (2010) Evapotranspiration modeling using linear genetic programming technique. J Irrig Drain Eng 136(10):715–723

Kisi O, Ozturk O (2007) Adaptive neuro-fuzzy computing technique for evapotranspiration estimation. J Irrig Drain Eng ASCE 133(4):368–379

Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection. The MIT Press, Cambridge, MA

Kumar M, Raghuwanshi NS, Singh R, Wallender WW, Pruitt WO (2002) Estimating evapotranspiration using artificial neural network. J Irrig Drain Eng ASCE 128(4):224–233

Kumar M, Raghuwanshi NS, Singh R (2011) Artificial neural networks approach in evapotranspiration modeling: a review. Irrig Sci 29:11–25

Monteith J. L 1965. The state and movement of water in living organisms. In Proceedings, Evaporation and Environment, XIXth Symp., Soc. For Exp. Biol., Swansea. Cambridge Univ. Press: New York, 205–234

Monteith JL (1973) Principles of environmental physics. Edward Arnold, London

Naoum S, Tsanis KI (2003) Hydroinformatics in evapotranspiration estimation. Environ Model Softw 18:261–271

Parasuraman K, Elshorbagy A, Carey S (2007) Modeling the dynamics of the evapotranspiration process using genetic programming. Hydrol Sci J 52(3):563–578

Penman HL (1948) Natural evaporation from open water, bare soil and grass. Proc R Soc Lon 19:120–146

Sabziparvar AA, Tabari H (2010) Regional estimation of reference evapotranspiration in arid and semiarid regions. J Irrig Drain Eng 136(10):724–731

Sakthivel NR, Nair BB, Sugumaran V (2012) Soft computing approach to fault diagnosis of centrifugal pump. Appl Soft Comput 12:1574–1581

Samadianfard S (2012) Gene expression programming analysis of implicit Colebrook-White equation in turbulent flow friction factor calculation. J Pet Sci Eng 92–93:48–55

Saudi Geological Survey. (2012). Kingdom of Saudi Arabia: facts and numbers, edition 1

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 Agric Water Manag 97:635–644

Shiri J, Kisi O, Landeras G, Lopez JJ, Nazemi AH, Stuyt L (2012) Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (northern Spain). J Hydrol 414–415:302–316

Smith M, Allen RG, Monteith JL, Pereira LS, Perrier A, Pruitt WO (1991) Rep. on the expert consultation on revision of FAO Guidelines for Crop Water Requirements, Land and Water Dev. Division. FAO, Rome

Suleiman A, Hoogenboom G (2009) A comparison of ACSE and FAO-56 reference evapotranspiration for a 15-minute time step in humid climate conditions. J Hydrol 375(3–4):326–333

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

Tabari H, Hosseinzadeh Talaee P (2011) Local calibration of the Hargreaves and Priestley-Taylor equations for estimating reference evapotranspiration in arid and cold climates of Iran based on the Penman-Monteith model. J Hydrol Eng ASCE 16(10):837–845

Temesgen B, Eching S, Davidoff B, Frame K (2005) Comparison of some reference evapotranspiration equations for California. J Irrig Drain Eng 1311:73–84

Trajkovic S (2010) Testing hourly reference evapotranspiration approaches using lysimeter measurements in a semiarid climate. Hydrol Res 41(1):38–49

Trajkovic S, Kolakovic S (2010) Comparison of simplified pan-based equations for estimating reference evapotranspiration. J Irrig Drain Eng 136(2):137–140

Traore S, Guven A (2012) Regional-specific numerical models of evapotranspiration using gene-expression programming interface in Sahel. Water Resour Manag 26:4367–4380

Walter I. A, Allen R. G, Elliott R, Itenfisu D, Brown, P, Jensen M. E, Mecham B., Howell T. A, Snyder R, Echings S, Spofford T, Hattendorf M, Martin D, Cuenca R. H. and Wright J. L 2001. The ASCE standardized reference evapotranspiration equation. Environmental and Water Resources Institute (EWRI) of the ASCE

Wang Y, Namaona W, Traore S, Zhang Z (2009) Seasonal temperature-based models for reference evapotranspiration estimation under semi-arid condition of Malawi. Afr J Agric Res 4(9):878–886

Xu CY, Singh VP (2002) Cross comparison of empirical equations for calculating potential evapotranspiration with data from Switzerland. Water Resour Manag 16:197–219

Yin Y, Wu S, Zheng D, Yang Q (2008) Radiation calibration of FAO56 Penman–Monteith model to estimate reference crop evapotranspiration in China. Agric Water Manag 95:77–84

Zanetti SS, Sousa EF, Oliveira VPS, Almeida FT, Bernardo S (2007) Estimating evapotranspiration using artificial neural network and minimum climatological data. J Irrig Drain Eng ASCE 133(2):83–89