Agricultural Water Management 163 (2016) 110–124

Contents lists available at ScienceDirect

Agricultural Water Management

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

Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate

Mohamed A. Yassin[a], A.A. Alazba[a,b], Mohamed A. Mattar[b,c,*]

[a] Alamoudi Water Chair, King Saud University, P.O. Box 2460, Riyadh 11451, Saudi Arabia
[b] Agricultural Engineering Department, King Saud University, P.O. Box 2460, Riyadh 11451, Saudi Arabia
[c] Agricultural Engineering Research Institute (AEnRI), Agricultural Research Center, P.O. Box 256, Giza, Egypt

ARTICLE INFO

Article history:
Received 7 March 2015
Received in revised form 2 August 2015
Accepted 11 September 2015
Available online 29 September 2015

Keywords:
Reference evapotranspiration
Penman–Monteith
Artificial intelligence
Arid environments

ABSTRACT

Artificial neural networks (ANNs) and gene expression programming (GEP) were compared to estimate daily reference evapotranspiration (ET_{ref}) under arid conditions. The daily climatic variables were collected by 13 meteorological stations from 1980 to 2010. The ANN and GEP models were trained on 65% of the climatic data and tested using the remaining 35%. The generalised Penman–Monteith (PMG) model was used as a reference target for evapotranspiration values, with h_c varies from 5 to 105 cm with increment of a centimetre. The developed models were spatially validated using climatic data from 1980 to 2010 taken from another six meteorological stations. The results showed that the eight ET_{ref} models developed using the ANN technique were slightly more accurate than those developed using the GEP technique. The ANN models' determination coefficients (R^2) ranged from 67.6% to 99.8% and root mean square error (RMSE) values ranged from 0.20 to 2.95 mm d^{-1}. The GEP models' R^2 values ranged from 64.4% to 95.5% and RMSE values ranged from 1.13 to 3.1 mm d^{-1}. Although the GEP models performed slightly worse than the ANN models, the GEP models used explicit equations.

1. Introduction

Evapotranspiration is the principal variable of the hydrological cycle affecting irrigation water requirements and the future planning and management of water resources. It can be determined either experimentally (directly) or mathematically (indirectly). It can be measured directly by using either a lysimeter or a water balance in a controlled crop area (Gavilan et al., 2007). However, this approach is difficult, time-consuming and expensive.

As the ET_{ref} depends on several interacting climatological factors, such as temperature, humidity, wind speed and radiation, it is difficult and complex to estimate it. Over the last 50 years, experts have developed many methods for estimating the ET_{ref}. Method selection essentially depends on the availability of measured climatic variables. The generalised Penman–Monteith (PMG) method is widely used in agricultural and environmental research to estimate the ET_{ref} and it coincides well with field observations. Many researchers acknowledge that the PMG model is the most promising standardised method for estimating the ET_{ref}. However, it requires a significant amount of climatic data, which may be unavailable or not be reliable in certain locations, especially when dealing with developing countries. In these cases, alternative methods that rely on fewer weather inputs are necessary.

Over the past decade, intelligent computational models have been developed as alternative methods for estimating the ET_{ref}, such as the artificial neural network (ANN) technique (Gorka et al., 2008). With the development of computer technology, ANNs have become increasingly important because of their wide application to different scientific areas. ANNs are defined as massive, parallel-distributed processors made of simple processing units, which have a natural propensity for storing experimental knowledge and making it available for use. ANNs are effective tools for modelling nonlinear processes, as they require few inputs and are able to map input-output relationships without any understanding of the physical process involved (Haykin, 1999).

Several studies have used ANN to estimate the ET_{ref} as a function of climatic variables. Bruton et al. (2000) first developed ANN models to estimate daily pan evaporation using weather data from Rome, Plains and Watkinsville, Georgia. Their ANN models estimated pan evaporation slightly better than multiple linear regression models and the Priestley–Taylor equation.

Kumar et al. (2002) developed an ANN model to estimate the ET_{ref} and evaluated appropriate combinations of various measured

* Corresponding author.
E-mail addresses: mmattar@ksu.edu.sa, dr.mohamedmattar@gmail.com (M.A. Mattar).

http://dx.doi.org/10.1016/j.agwat.2015.09.009
0378-3774/

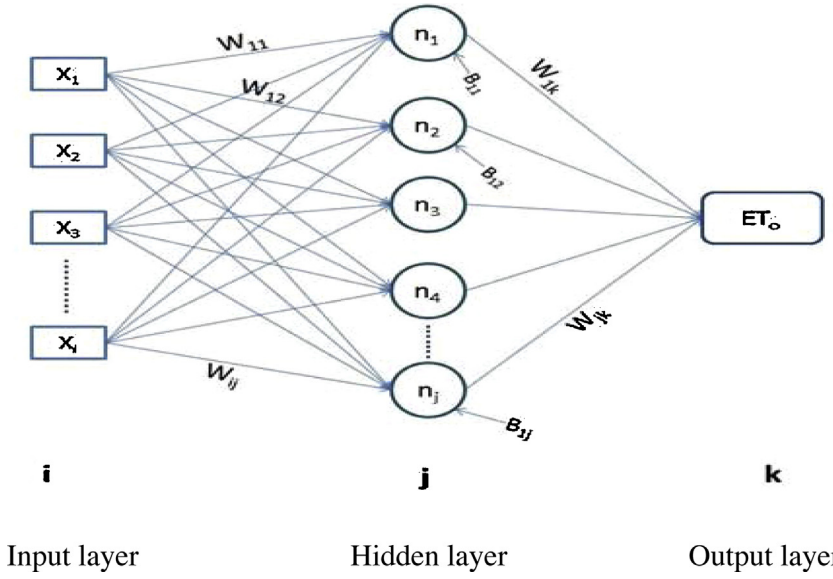

Fig. 1. Architecture of the ANN used to model the ET_{ref}.

weather data. The results indicated that their ANN model predicted the ET_{ref} better than the FAO-56 Penman–Monteith (PMFAO) method. Sudheer et al. (2003) and Trajkovic et al. (2003) reported the performance of radial basis function ANN models in ET_{ref} estimation. Arca et al. (2004) tested 11 combinations for estimating the ET_{ref} using ANN models. Under the most simplified combination, ET_{ref} was estimated as a function of two variables, the extra-terrestrial solar radiation and air temperature.

Landeras et al. (2008) used weather data collected from four weather stations of the Basque Meteorological Service from 1992 to 1996. They compared seven ANN models with different input combinations with ten locally calibrated empirical and semi-empirical ET_{ref} models, using PMFAO daily ET_{ref} values as a reference. The results showed the ANN models obtained better results than the locally calibrated ET_{ref} equations. Huo et al. (2012) trained and tested ANN models to forecast the ET_{ref} using 50 years of meteorological data from three stations in north-west China. They compared the ANN models' performances to multiple linear regressions, the Penman equation and two empirical equations. The results showed that the ANN models exhibited high precision compared to the other models and that ANN models with five inputs were more accurate than those with four or three inputs.

Gene expression programming (GEP) was invented by Ferreira (2001b) and is the natural development of genetic algorithms and genetic programming (GP). 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. It has been used to solve problems such as symbolic regression, multi-agent strategies, time series prediction, circuit design and evolutionary neural networks (Samadianfard, 2012).

GEP has been used in a number of hydrological and hydraulic modelling problems. Guven and Aytek (2009) used a GEP approach to model the stage–discharge relationship and compared the results with conventional methods. They found that the explicit algebraic formulations resulting from the GEP approach gave the best results. In a similar study, Azamathulla et al. (2011) developed mathematical models to estimate the stage–discharge relationship for the Pahang River based on GP and GEP techniques.

Ghani and Azamathulla (2011) used GEP to model the functional relationships of sediment transport in sewer pipe systems. More recent, Azamathulla and Ahmad (2012) used GEP to predict the transverse mixing coefficient in open channel flows. Zahiri and Eghbali (2012) used GEP to predict the flow discharge in compound channels.

Of the many published studies on the application of GEP in hydrological modelling. However, the use of GEP for modelling evapotranspiration has been recorded by only a few studies. Aytek and Kişi (2008) presented GP as a new tool for estimating the ET_{ref} using daily climatic variables obtained from the California Irrigation Management Information System database. The results obtained were compared to seven conventional ET_{ref} models. They found that the new model produced satisfactorily results and could be used as an alternative to the conventional models. However, Kişi and Guven (2010) investigated the accuracy of linear genetic programming, which is an extension of the GP technique, in modelling the daily ET_{ref} using the PMFAO equation. The linear genetic programming model was found to perform more accurately than the support vector regression model, artificial neural network and four empirical models. Terzi (2013) compared GEP, ANFIS as an alternative approach to estimate daily pan evaporation in Turkey. Traore and Guven (2013) used GEP for modelling the ET_0 using routing weather data from tropical seasonally dry regions of West Africa in Burkina Faso. This study investigates the application of the GEP and ANN for modelling daily ET_{ref}. Moreover, the performance of the GEP models is statistically compared with the ANN models developed.

2. Materials and methods

2.1. Artificial neural network

An artificial neural network (ANN) consists of a large number of interconnecting processing elements and is similar in structure to a biological neural network (Eslamian et al., 2012). ANN usually consists of layers of neurons, weights representing the connection strengths and a transfer or activation function.

In this study, an ANN model of multilayer perception with a universal function approximator is used. Fig. 1 depicts the model layers. The input layer (i) is connected to the hidden layer (j), which is in turn connected to the output layer (k) by means of the connection weights (W) and biases (B). The weights are used to change the throughput parameters and vary the connections to the neurons (n). The biases are used as additional elements inside the hidden

and output layer neurons. The neuron (processing element) in the hidden layer consists of aggregating weighted inputs, resulting in a quantity-weighted input (activation value). In the hidden layer, the neuron's activation value (h_j) is mathematically characterized using the following equation (Haykin, 1999):

$$n_{ij} = \sum_{i=1}^{N} W_{ij} X_i + B_j \quad (1)$$

$$h_j = f\left(n_{ij}\right) \quad (2)$$

The output layer neuron (Y_k) is given by the following equations:

$$n_{ik} = \sum_{j=1}^{N} W_{jk} h_j + B_k \quad (3)$$

$$Y_k = f\left(n_{jk}\right) \quad (4)$$

The hidden-layer neurons consist of activation functions (f) that help to translate the input variables (the activation values of the neurons) into the required output variable. The most common transfer functions in hydrological modelling are the sigmoid and hyperbolic tangent functions (Dawson and Wilby, 1998; Zanetti et al., 2007). The hyperbolic tangent is similar to the sigmoid but can exhibit different learning dynamics during training. The sigmoid function is used in this study. Its general functional form is:

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (5)$$

where x is either the value of n_{ij} or n_{jk}.

A feed-forward ANN that uses a back-propagation learning algorithm was employed in this study, as such ANNs are commonly used to estimate the ET_{ref}. The back-propagation learning algorithm optimises the error function to modify the link weight. More than 70% of the existing studies that applied ANN techniques to hydrological processes used the back-propagation learning algorithm because of its simplicity and robustness (Kumar et al., 2011). It controls the rate at which learning takes place using a momentum term and the learning rate. The momentum term is generally used to accelerate convergence and avoid local minima. A learning rate of 0.01 and a momentum factor of 0.8 are used.

2.2. Gene expression programming

Gene expression programming (GEP) is a new evolutionary artificial intelligence technique developed by Ferreira (2001a). According to Ferreira (2001a,b) the primary difference between GEP and its predecessors, genetic algorithms (GA) and genetic programming (GP), stems from the nature of the individuals: in GA, the individuals are linear strings of fixed length (chromosomes). In GP, the individuals are nonlinear entities of different sizes and shapes (parse trees). In GEP, the individuals are encoded as linear strings of fixed length (chromosomes) that are expressed as nonlinear entities of different sizes and shapes.

GEP uses chromosomes, which are usually composed of more than one gene of equal length, and expression trees or programmes, which are the expressions of the genetic information encoded in the chromosomes (Ferreira, 2006). The chromosomes are composed of multiple genes, each gene encoding a smaller sub-programme. In GEP, the linear chromosomes represent the genotype and the branched expression trees represent the phenotype (Ferreira, 2001b). Fig. 2 shows the organisation of a standard GEP model.

GEP is a complete genotype/phenotype system in which the genotype is totally separate from the phenotype. In contrast, in GP, the genotype and phenotype constitute one entangled mess, more formally referred to as a simple replicator system. As a result, GEP's genotype/phenotype system surpasses the GP system by a factor of 100–60,000 (Ferreira, 2001a,b).

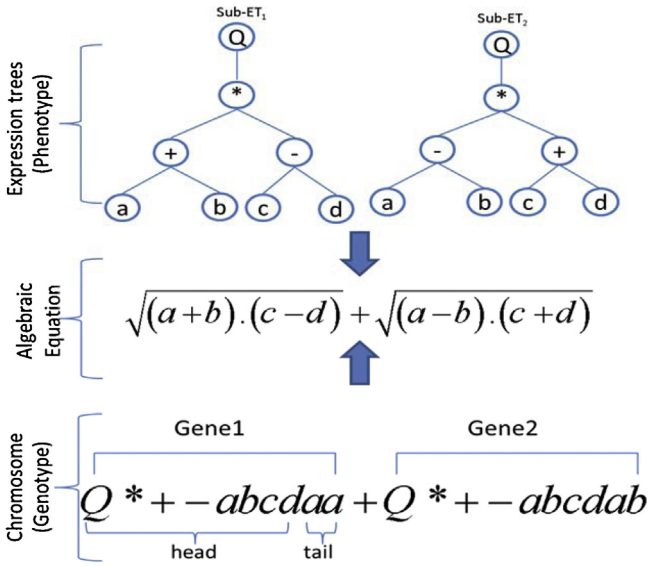

Fig. 2. GEP model of a chromosome with two genes and their phenotypes (Shiri et al., 2012).

GEP models encode their information in linear chromosomes, which are later translated or expressed in expression trees. These computer programmes are usually developed to solve a particular problem and are selected according to their ability to solve that problem (Guven and Aytek, 2009).

2.3. Study area and input data

Kingdom of Saudi Arabia is situated in the far southwest corner of Asia (Fig. 3), between latitudes 16°22′46″N and 32°14′00″N and longitudes 34°29′30″E and 55°40′00″E. It is the largest country in Arabia. The SA occupies about 70% of the area of the Arabian Peninsula with an approximate area of 1,950,000 km^2. It is divided into thirteen provinces, as shown in Fig. 3. This study considers all of the provinces. The provinces are arranged by area in descending order in Table 1.

For this study, climatic data was recorded at 19 meteorological stations selected from the 13 SA provinces. The spatial distribution of the selected stations within the provinces is shown in Fig. 3. Each province is represented by two stations, except for the provinces of Najran, Ha'il, Al-Jouf, Bisha, Al-Qasim, Jizan and Al-Baha, which are only represented by one station. The Presidency of Meteorology and Environment provided the data. The study's climatic data covers 31 years of daily meteorological information recorded from 1980 to 2010. The recorded data for all of the stations includes the maximum, minimum and mean air temperatures (T_x, T_n, and T_a) (°C); maximum, minimum and mean relative humidity (RH_x, RH_n and RH_a) (%); wind speed at a 2 m height (U_2) ($m\,s^{-1}$) and solar radiation (R_s) ($MJ\,m^{-2}\,d^{-1}$). Table 1 describes the meteorological stations and lists the annual averages of the climatic data from each station.

The ANN and GEP models take at most nine input variables, T_x, T_n, T_a, RH_x, RH_n, RH_a, U_2, R_s and the reference crop height (h_c) (m), which varies from 5 to 105 cm. This range is selected to cover both grass (10–15 cm) and alfalfa (30–80 cm). A random h_c value is chosen during training. The ET_{ref} is the output variable. The input variables are divided into three sets. The training set for the ANN and GEP models is composed of 65% of the daily data collected by 13 of the weather stations, Riyadh (North), Al-Qasim, Ha'il, Al-Jouf,

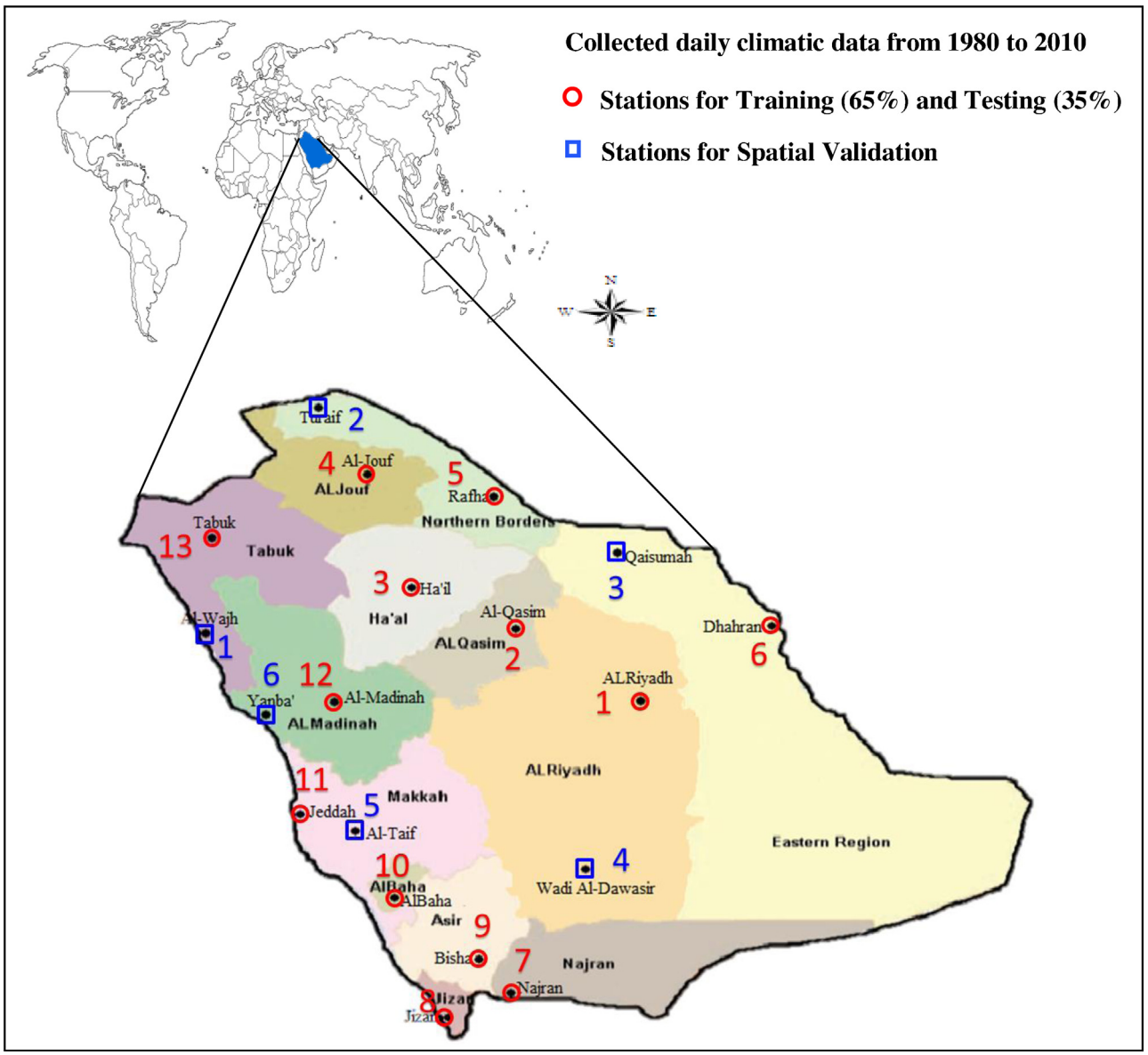

Fig. 3. Map of the KSA, showing its provinces and meteorological stations.

Table 1
Meteorological station sites and climatic parameters.

Provinces	Areas[a] (km^2)	Stations	Location			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 $m^{-2}\,d^{-1}$
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	19	25	56	29	44	4.2	26
		Yanba'	38.10	24.10	1	29	17	22	78	23	50	3.2	29
Makkah	137	Jeddah	39.17	21.40	12	34	22	28	81	37	60	2.6	23
		Al-Ta'if	40.50	21.50	1449	35	23	29	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	25	29	60	33	44	3.5	28
Ha'il	120	Ha'il	41.70	27.40	1013	34	22	28	81	37	60	2.3	14
Northern borders	104	Turaif	38.65	31.68	854	35	23	29	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	18	30	2.9	27
Jizan	13	Jizan	42.60	16.88	3	36	25	30	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). King Saudi Arabia: Facts and Numbers, edition 1.

Rafha, Dhahran, Najran, Jizan, Bisha, Al-Baha, Jeddah, Al-Madina and Tabuk, from 1980 to 2010. The training set is used to find the patterns present in the data. The testing set for the ANN and GEP models is composed of the remaining 35% of the data from the same weather stations and period as the training set. It is used to evaluate the generalisation abilities of the trained models. The ANN and GEP models' performances are checked once more with a validation data set. It is composed of the data collected by the remaining six weather stations, Turaif, Al-Wajh, Qaisumah, Yanba', Al-Ta'if and Wadi Al-Dawasir, from 1980 to 2010.

The statistical parameters of the daily climatic data sets (training, testing and validation) were given in Table 2. In Table 2, the X_m, X_n, X_a, S_x, K, and C_{sx} denote the maximum, minimum, mean, standard deviation and skewness, respectively.

2.4. Output/targeted data of the ANN and GEP models

The performances of the ANN and GEP models are compared to the PMG method. Many studies used the form of PMG method (FAO-56) which depend on the grass as a reference crop height (h_c). (Irmak et al., 2003; Gavilan et al., 2006). The PMG 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 advantages over many other mathematical equations. It can be used globally without any local calibrations due to its physical basis, is well-documented and has been validated with a significant amount of lysimeter data (Gocic and Trajkovic, 2010). Many researchers (Kumar et al., 2002; Trajkovic, 2005; Kişi and Özturk, 2007; Zanetti et al., 2007; Landeras et al., 2008; Jain et al., 2008; Dai et al., 2009; Traore et al., 2010) have used the PMG equation as a reference and standard equation to evaluate the results of their mathematical models. The daily ET_{ref} values from the PMG equation are used as the output/target variables in the ANN and GEP models. A form of the PMG model can be written as follows (Allen et al., 1998):

$$ET_{ref} = \lambda^{-1}\left(\frac{\Delta(R_n - G) + \rho_a c_p (e_s - e_a)/r_a}{\Delta + \gamma\left(1 + r_s/r_a\right)}\right) \quad (6)$$

where ET_{ref} is the reference evapotranspiration (mm d^{-1}), λ is the latent heat of vapourization (MJ kg^{-1}), Δ is the slope of the saturation vapour pressure–temperature curve at mean air temperature (kPa $^{\circ}C^{-1}$), R_n is the net radiation (MJ m^{-2} d^{-1}), G is the soil heat flux (MJ m^{-2} d^{-1}),ρ_a is the mean air density at constant pressure (kg m^{-3}), c_p is the specific heat of the air, 0.001013 (MJ kg^{-1} $^{\circ}C^{-1}$), e_s is the saturation vapour pressure at air temperature (kPa), e_a is the actual vapour pressure [kPa], r_a is the aerodynamic resistance (s m^{-1}), γ is the psychometric constant (kPa $^{\circ}C^{-1}$), r_s is the (bulk) surface resistance in (s m^{-1}).

An alternatively form of PMG based on the original PM equation (Eq. (6)) was used (Alazba, 2004):

$$ET_{ref} = \lambda^{-1}\left[\frac{\Delta}{\Delta + \gamma^*}(R_n - G) + \frac{\gamma}{\Delta + \gamma^*}K(e_s - e_a)\right] \quad (7)$$

where γ^* is the modified psychometric constant equal to $\gamma^* = \gamma(1 + \frac{r_s}{r_a})$ [kPa $^{\circ}C^{-1}$], and K is a parameter equal to $1.854 \times 10^5 \frac{\lambda/r_a}{T+273}$ [MJ m^{-2} d^{-1} kPa^{-1}].

All aforementioned parameters were calculated using equations provided by Allen et al. (1998). The soil heat flux (G) was assumed to be zero over the calculation time step period (24 h) (Allen et al., 2005). Aerodynamic resistance (r_a) is estimated simply by the following developed equation (Alazba, 2004):

$$r_a = \frac{1 - \ln(h_c)}{0.015U_2} \quad (8)$$

2.5. Models development using ANN and GEP

Several combinations of the input parameters were used as inputs to estimate the daily ET_{ref} using the ANN and GEP models. The input parameter combinations are listed in Table 3. Eight ANN and GEP models were developed to test the performance of different combinations of input parameters, including climatic parameters and a reference crop height chosen randomly during the training process.

Software Multiple Back-Propagation version 2.2.4 was used to develop the ANN model to estimate the ET_{ref} and the sigmoid transfer function. Nine input variables were used (the maximum input set of the ANN). The output as one neuron was in the output layer. The number of hidden neurons depended on several factors, such as the number of input and output neurons, the number of training cases, the amount of noise in the targets, the complexity of the function or classification to be learned, the architecture, the type of hidden unit activation function and the training algorithm (Kumar et al., 2011). The training data must be automatically normalized before they are exported to the ANN's feed-forward neural networks for training. Normalization is commonly between 0.15 and 0.85 in ANN modelling. The training characteristics were improved using:

$$X_n = (0.85 - 0.15)\left(\frac{X_0 - X_{min}}{X_{max} - X_{min}}\right) + 0.15 \quad (9)$$

where X_n is the normalized value, X_0 is the original value, X_{min} is the minimum value and X_{max} is the maximum value.

The input data can flow after it is normalized. They undergo unidirectional processing from the input layer, through the hidden layer, to the output layer. In the hidden layer, each neuron receives input signals from the input layer through the weights (Izadifar, 2010). The data are processed separately by each hidden layer neuron and the outputs are passed to the output layer neurons.

The network output and target outputs are computed at the end of each forward pass in the forward-propagation stage. If an error is higher than a selected value, a reverse pass is performed to modify the connection weights by minimizing the error between the target and computed outputs (back-propagation stage). Otherwise, the training stops. The best number of hidden neurons in the hidden layer is found by training many ANNs and repeating the trial and error procedure (Jain et al., 2008), taking into account the error values. The hidden layer initially has two nodes. The number of nodes increases in each trial by between one and four nodes, to a maximum of 20 nodes.

In the present work, the GeneXproTools 5.0 program is used to estimate the daily ET_{ref}. GEP model development consisted of five major steps (Ferreira, 2001a,b):

(1) Select the fitness function. The fitness (f_i) of an individual program (i) is measured by:

$$f_i = \sum_{j=1}^{C_t}\left(M - |C_{(i,j)} - T_{(j)}|\right) \quad (10)$$

where M is the selection range, $C_{(i,j)}$ is the value returned by the individual chromosome i for fitness case j (out of C_t fitness cases) and T_j is the target value for fitness case j. If $|C_{(i,j)} - T_j|$ (the precision) ≤ 0.01, then the precision is 0 and $f_i = f_{max} = C_t M$. The advantage of this fitness function is that the system can find the optimal solution by itself.

(2) Choose the set of terminals (T) and the set of functions (F) to create the chromosomes. For instance, the terminal set includes

Table 2
Daily statistical parameters of the climatic variables for the training, testing and validation processes.

Statistical parameters	Climatic variables							
	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} d^{-1})
Training and testing processes								
X_m	54.20	38.30	44	100	98	99	14.91	29.86
X_n	2.80	−9.40	−0.60	4	1	2	0.51	4.61
X_a	32.31	18.34	25.40	54.58	23.52	37.73	3.21	21.51
S_x	8.01	7.62	7.73	24.76	16.34	20.59	1.44	5.29
K	−0.48	−0.65	−0.66	−1.29	0.30	−0.98	1.82	−0.83
C_{sx}	−0.51	−0.40	−0.47	0.04	1.01	0.43	0.97	−0.25
Validation process								
X_m	51.70	36.80	42.50	100.00	95.00	99.00	14.40	30.01
X_n	2.00	−8.00	−3.00	6.00	0.00	2.00	0.51	4.92
X_a	31.41	17.65	24.59	62.68	26.72	43.96	3.87	21.65
S_x	8.11	7.31	7.65	24.96	16.38	20.75	1.55	5.56
K	−0.19	−0.47	−0.32	−1.10	0.16	−1.05	0.97	−0.92
C_{sx}	−0.37	−0.44	−0.42	−0.39	0.83	0.05	0.75	−0.29

Table 3
The input variables combinations used in the ANN and GEP models.

Model		Input parameters								
		Temperature (°C)			Relative humidity (%)			U_2 (m s^{-1})	R_s ()MJ m^{-2} d^{-1}	h_c (m)
		T_x	T_n	T_a	RH_x	RH_n	RH_a			
ANN-MOD1	GEP-MOD1	√	√	√						√
ANN-MOD2	GEP-MOD2	√	√	√	√	√	√			√
ANN-MOD3	GEP-MOD3	√	√	√				√		√
ANN-MOD4	GEP-MOD4	√	√	√					√	√
ANN-MOD5	GEP-MOD5	√	√	√	√	√	√	√		√
ANN-MOD6	GEP-MOD6	√	√	√	√	√	√		√	√
ANN-MOD7	GEP-MOD7	√	√	√				√	√	√
ANN-MOD8	GEP-MOD8	√	√	√	√	√	√	√	√	√

the following variables: T_x, T_n, T_a, RH_x, RH_n, RH_a, R_s, U_2 and h_c. The choice of functions depends on the user. In this study, different mathematical functions were used, such as +, −, ×, ÷, √, ∛, exp and sin. Eight input combinations were tested, as listed in Table 3.

(3) Choose the chromosomal architecture. A single gene and two head length was initially used. The number of genes and heads were increased one after another during each run and the training and testing performance of each model was monitored.
(4) Choose the linking function. Only addition or multiplication linking functions could be chosen for algebraic sub-trees.
(5) Select the set of GEP operators from mutation, transposition and recombination. This process was repeated for a pre-specified number of generations or until a solution was found.

2.6. Performance criteria

After training the ANN and GEP models and validating the data set, the ET_{ref} values were estimated and compared to the daily values from the PMG model. The comparisons were made using the following statistical parameters.

$$R^2 = \frac{\left(\sum_{i=1}^{n}\left(E_i-\bar{E}\right)\left(C_i-\bar{C}\right)\right)^2}{\sum_{i=1}^{n}\left(E_i-\bar{E}\right)^2 \times \sum_{i=1}^{n}\left(C_i-\bar{C}\right)^2} \quad (11)$$

$$\mathrm{OI} = \frac{1}{2}\left(1 - \frac{\mathrm{RMSE}}{E_{max}-E_{min}} + \mathrm{ME}\right) \quad (12)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(E_i-C_i)^2}{n}} \quad (13)$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|E_i-C_i|}{n} \quad (14)$$

where

E_i = value of ET_{ref} estimated by the PMG;
C_i = corresponding value calculated by mathematical ET_{ref} models;
n = number of observations;
$\bar{E}$ = average of the estimated values; and
$\bar{C}$ = average of the calculated values.

The coefficient of determination (R^2) measures the degree of correlation between the estimated and calculated values, where values approaching 1.0 indicate a good correlation. The root mean square error (RMSE) expresses the error in the same units that describe the variable (Legates and McCabe, 1999). The lower the RMSE, the better the matching. The overall index of the model performance (OI) combines the normalised RMSE and the model efficiency value. An OI value of 1.0 indicates a perfect fit between a model's estimated and calculated values (Alazba et al., 2012; Mattar et al., 2015; Mattar and Alamoud, 2015). The mean absolute error (MAE) is the average value of the absolute differences between the estimated and calculated values. A low MAE implies good model performance.

3. Results and discussion

3.1. Performances of ANN models

Fig. 1 represents the final architecture of ANN models. The second column of Table 4 refers to the number of input, hidden and output nodes of each ANN model. Furthermore, Table 4 presents the statistical results of the optimum ANN models using different input combinations to estimate the ET_{ref}. In training process, the ANN models' R^2 values ranged from 67.9 to 99.8%, OI values from 80.6 to 99.6%, RMSE values from 0.20 to 2.95 mm d^{-1} and MAE values from 0.15 to 2.12 mm d^{-1}.

Table 4
Statistical performance of the optimized ANN models during training and testing.

Model	Structure	Training				Testing			
		R^2 (%)	OI (%)	RMSE (mm d^{-1})	MAE (mm d^{-1})	R^2 (%)	OI (%)	RMSE (mm d^{-1})	MAE (mm d^{-1})
ANN-MOD1	4-20-1	67.9	80.6	2.95	2.12	67.6	80.3	3.00	2.20
ANN-MOD2	7-16-1	80.2	87.4	2.32	1.59	80.4	87.5	2.33	1.61
ANN-MOD3	5-20-1	87.1	91.4	1.87	1.35	87.1	91.4	1.89	1.33
ANN-MOD4	5-13-1	72.2	83.0	2.74	1.96	72.1	82.7	2.78	2.04
ANN-MOD5	8-20-1	99.1	98.9	0.502	0.403	99.1	98.9	0.51	0.41
ANN-MOD6	8-20-1	82.6	88.8	2.17	1.45	82.3	88.5	2.22	1.53
ANN-MOD7	6-16-1	88.6	92.3	1.75	1.23	88.8	92.4	1.76	1.19
ANN-MOD8	9-20-1	99.8	99.6	0.207	0.151	99.8	99.7	0.19	0.14

It can be observed that the absence or presence of some of the input variables in the input sets significantly affects the models' performances. The ANN-MOD1 temperature-based model only took the maximum, mean and minimum air temperatures. ANN-MOD1 performed worst, with R^2 = 67.9%, OI = 80.6%, RMSE = 2.95 mm d^{-1} and MAE = 2.12 mm d^{-1} (Table 4). ANN-MOD2 performed better than ANN-MOD1, due to the presence of the humidity variables. ANN-MOD4, which added solar radiation to the ANN-MOD1 combination, did not perform better than ANN-MOD2.

Additionally, the performance of ANN-MOD3 (R^2 = 87.1%, OI = 91.4%, RMSE = 1.87 mm d^{-1} and MAE = 1.35 mm d^{-1}) performed better than ANN-MOD1, ANN-MOD2 and ANN-MOD4. Comparing ANN-MOD7's results with those of the other ANN models shows that the accuracy of the ANN-MOD1 temperature-based model was significantly improved by the inclusion of both solar radiation and wind speed, as ANN-MOD7 had a 30.49% increase in the R^2 over ANN-MOD1. This is in agreement with Hupet and Vanclooster (2001).

ANN-MOD6 had higher RMSE (2.17 mm d^{-1}) and MAE (1.45 mm d^{-1}) values than ANN-MOD7. ANN-MOD6 replaces ANN-MOD7's wind speed with the humidity variables. However, switching ANN-MOD6's solar radiaton for wind speed, as in ANN-MOD5, resulted in a dramatic increase in R^2 from 82.6% to 99.1%, i.e. a 19.98% increase. This is in accordance with Kişi and Özturk (2007). Wind speed is likely to be an effective, powerful variable for accurately modelling the nonlinear complex process of ET_{ref} (Fisher et al., 2005; Xiaoying and Erda, 2005). ANN-MOD8, which has the full input set similar to the PMG model, performs better than the rest of the ANN models.

In testing, ANN-MOD1 statistics were R^2 = 67.6%, OI = 80.3%, RMSE = 3.0 mm d^{-1} and MAE = 2.20 mm d^{-1}. Fig. 4 and Table 4 show the results of adding either the humidity factors (ANN-MOD2), wind speed (ANN-MOD3) or solar radiation (ANN-MOD4) to ANN-MOD1. ANNMOD2 (R^2 = 80.4%, OI = 87.5%, RMSE = 2.33 mm d^{-1} and MAE = 1.61 mm d^{-1}) and ANN-MOD3 (R^2 = 87.1%, OI = 91.4%, RMSE = 1.89 mm d^{-1} and MAE = 1.33 mm d^{-1}) performed better than ANN-MOD1. A slightly worse performance (R^2 = 72.1%, OI = 82.7%, RMSE = 2.78 mm d^{-1} and MAE = 2.04 mm d^{-1}) was obtained for ANN-MOD4. This result indicates that solar radiation had a slight effect on modelling the ET_{ref}, as the R^2 value increased by 6.65% when solar radiation was added to ANN-MOD1. Fig. 4 and Table 4 also show the results of different input combinations (ANN-MOD5, ANN-MOD6, ANN-MOD7) and the full input set (ANN-MOD8).

The relative humidity factors seemed to be more effective than solar radiation in the modelling of the ET_{ref}, as the R^2 increased by 18.93% when the humidity factors were added to ANN-MOD1. Adding wind speed to the input combination improved the estimation accuracy significantly, due to its advection effects on the ET_{ref} (Kişi, 2007), as the R^2 increased by 28.84% when wind speed was added to ANN-MOD1.

3.2. Performances of GEP models

The algebraic equations that best estimate the ET_{ref} are given in Table 5. Moreover, the R^2, RMSE, OI and MAE statistics of each GEP model during training and testing are given in Table 6. GEP-MOD1 (whose inputs were the three air temperature variables and crop height) had the smallest R^2 (64.4%) and OI (92.2%) values and the highest RMSE (3.10 mm d^{-1}) and MAE (2.29 mm d^{-1}) values in training. Thus, GEP-MOD1 gave poor estimates. The relative humidity variables seem to have been the most effective in estimating the $ET_{ref,}$ as adding relative humidity to GEP-MOD1 (GEP-MOD2) significantly increased the performance, giving the largest R^2 increase (18%) and RMSE decrease (8%) in the training process.

GEP-MOD3 added wind speed and performed better than GEP-MOD1. In contrast, GEP-MOD4, which added solar radiation to the GEP-MOD1 combination, did not perform better than GEP-MOD2, with R^2 = 68.3% and RMSE = 2.92 mm d^{-1}. GEP-MOD7 added solar radiation to GEP-MOD3 and performed better than GEP-MOD3, increasing the R^2 from 76.1 to 82.2% and the OI from 93.9 to 95.4% and decreasing the RMSE from 2.64 to 2.2 mm d^{-1} and the MAE from 1.98 to 1.64 mm d^{-1}. Replacing relative humidity with wind speed resulted in a worse performance by GEP-MOD6 than GEP-MOD7. Conversely, replacing relative humidity with solar radiation resulted in a better performance by GEP-MOD5 than GEP-MOD7.

Furthermore, it can be seen from Table 6 that GEP-MOD8 outperformed the other models by all of the performance criteria. GEP-MOD8 ranked best in the training process. This was expected, as GEP-MOD8 considered all of the variables that have an influence on the ET_{ref}.

During testing process, the GEP models had R^2 values ranging from 63.2 to 95.4%, OI values from 77.3 to 96.1%, RMSE values from 1.14 to 3.2 mm d^{-1} and MAE values from 0.83 to 2.42 mm d^{-1}. It can be observed from Table 5 that the GEP models with high R^2 and OI values and low RMSE and MAE values were able to predict the target values with an acceptable degree of accuracy. Furthermore, GEP-MOD1 statistics were R^2 = 63.6%, OI = 77.9%, RMSE = 3.19 mm d^{-1} and MAE = 2.40 mm d^{-1}. Fig. 5 and Table 6 give the results of adding the relative humidity variables (GEP-MOD2), wind speed (GEP-MOD3) or solar radiation (GEP-MOD4). GEP-MOD2 (R^2 = 71.0%, OI = 81.0%, RMSE = 2.94 mm d^{-1} and MAE = 2.17 mm d^{-1}) and GEP-MOD3 (R^2 = 76.8%, OI = 84.2%, RMSE = 2.65 mm d^{-1} and MAE = 1.98 mm d^{-1}) produced better results, whereas GEP-MOD4 (R^2 = 67.9%, OI = 80.4%, RMSE = 2.99 mm d^{-1} and MAE = 2.21 mm d^{-1}) performed slightly worse. This result indicates the slight effect of solar radiation on modelling the ET_{ref}, as the R^2 only increased by 6.76% when solar radiation was added to GEP-MOD1. The relative humidity seemed to be more effective than solar radiation in modelling the ET_{ref}, as the R^2 increased by 11.63% when relative humidity was added to GEP-MOD1. Adding wind speed into the input combination improved the estimation accuracy significantly, due to its advection effects on evapotranspiration.

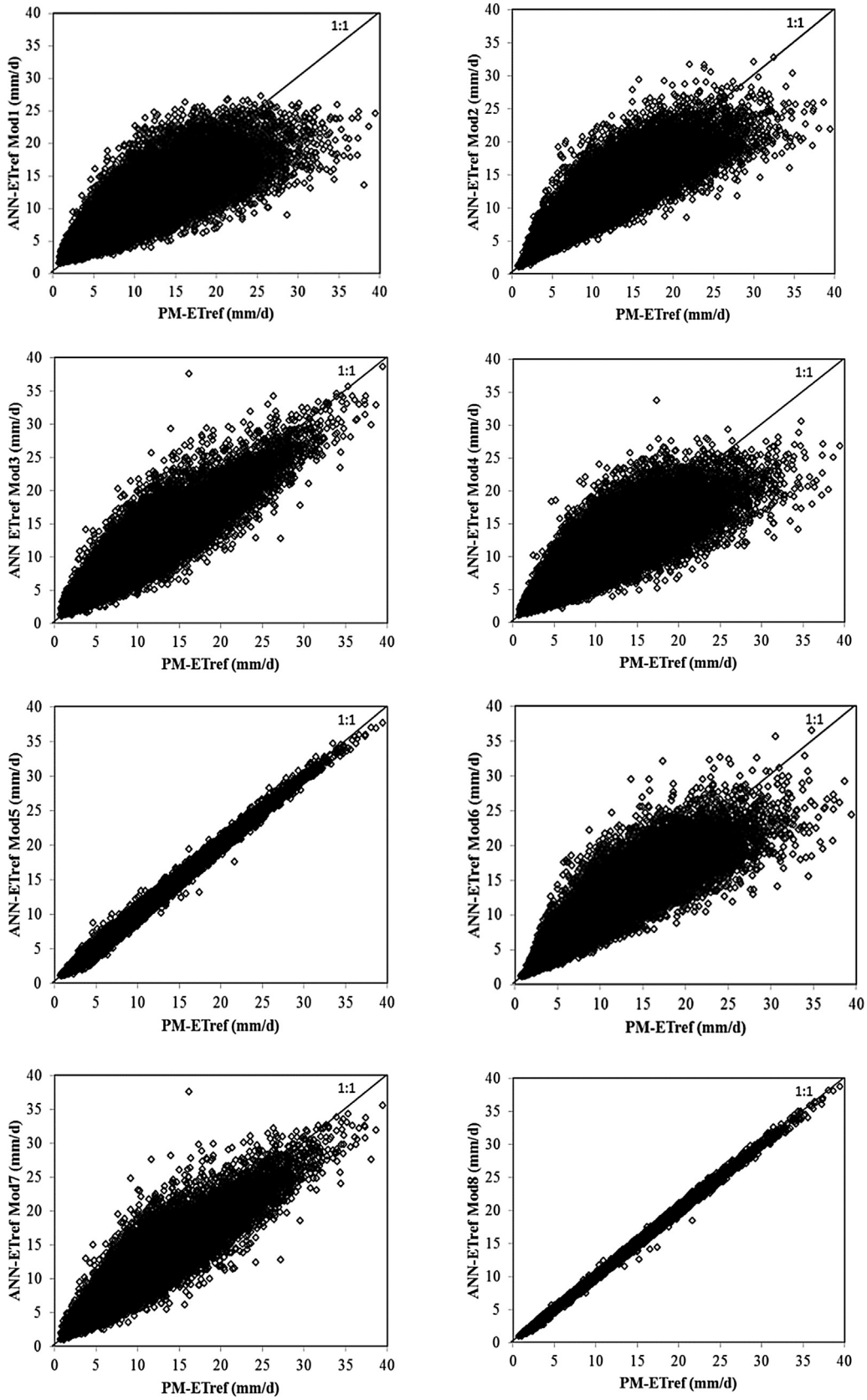

Fig. 4. Performance of ANN during testing process, using 35% of the data collected from 1980 to 2010 by 13 weather stations.

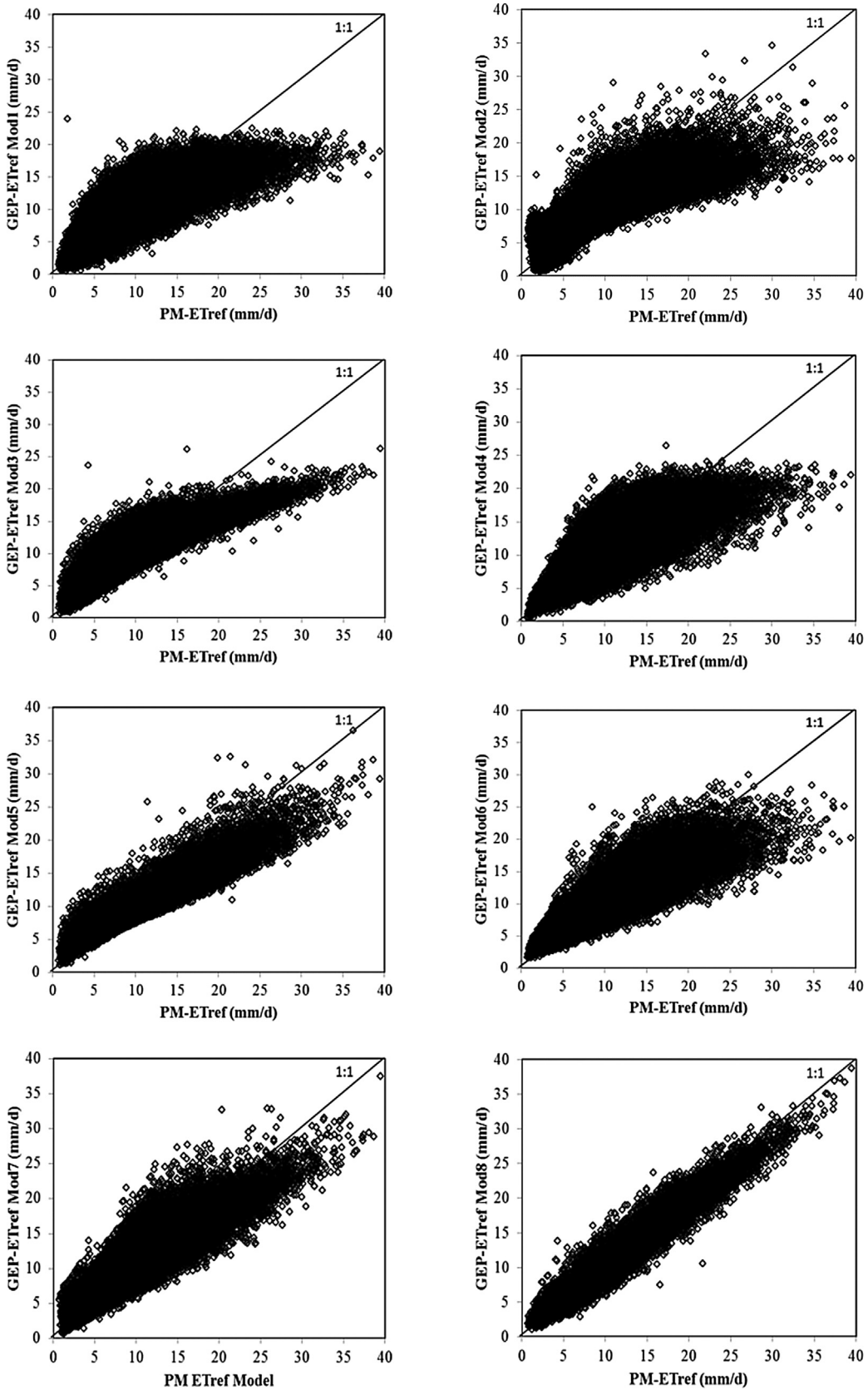

Fig. 5. Performance of GEP during testing process, using 35% of the data collected from 1980 to 2010 by 13 weather stations.

Table 5
Developed algebraic equations by gene expression programming for estimating ET_{ref}.

Model	Algebraic equation
GEP-MOD1	$ET_{ref1} = \frac{(((8.98-T_x)+2.177)-(T_a+h_c))}{T_a-h_c+1.48} + \frac{((T_x/-5.24)-T_n-(T_n+T_a))}{(T_x/h_c)-0.446} + \frac{h_c-2T_x}{-6.135/h_c} + \frac{T_a}{3.578} - h_c + 0.0687$
GEP-MOD2	$ET_{ref2} = \sqrt{\sqrt{(((RH_a - T_n) - 38.8399) - (T_n - 3.0405))^2}} + \left(\left(\frac{(T_n-6.9194)^2}{T_x\times(-1.2424-RH_x)}\right)^2 \times RH_a\right) + \left(\sqrt{(((RH_n + T_a) - RH_n) \times T_a)^{0.5}} \times h_c\right)$
GEP-MOD3	$ET_{ref3} = \left(T_a \times (\exp(-4.9524) \times (R_s - 5.2876)) - \exp\left(\frac{h_c}{T_x}\right) + \left(T_a^2 \times \left(\frac{h_c}{(T_n/1.8349)+103.2835}\right)\right) + \left(R_s \times \exp\left(-0.07307 \times \left(\exp\left(\frac{(h_c)}{h_c}\right)\right)\right)\right)\right)$
GEP-MOD4	$ET_{ref4} = \frac{T_a}{(9.8405-U_n)/(T_a\times h_c))-h_c+3.066} + U_2 \times \left(h_c - \exp\left(\sqrt{h_c} - U_2\right)\right) - h_c \times \left(\frac{2.404T_n/T_x}{0.6563}\right) + 0.8438$
GEP-MOD5	$ET_{ref5} = \frac{(R_s-T_n-5.6758)/\exp(h_c)}{h_c+15.9313} + \left(\frac{0.058657_x^2}{2U_2-49.6336}\right)^2 + (0.01142 \times R_s \times T_a \times h_c) + U_2$
GEP-MOD6	$ET_{ref6} = \frac{R_s}{\left(\sqrt{(7.34/h_c)-6.21+\sqrt{RH_x}}\right)} + \exp\left(\frac{T_n}{\ln(-8h_c+T_a+RH_n)^2}\right) + \sqrt{T_x \times \left(\left(\frac{R_s}{RH_a}\right)^{0.5} \times h_c^{0.5}\right)}$
GEP-MOD7	$ET_{ref7} = \frac{(T_x\times(U_2-RH_a))/(RH_x-0.8)}{(RH_a\times(-4.7/U_2))} + \sqrt{T_x - \frac{(-8.45\times U_2\times T_a\times h_c)}{RH_n}} + \left(\frac{2T_n}{RH_x} + \frac{h_c}{U_2} + \frac{0.38}{U_2} - 1.63\right)$
GEP-MOD8	$ET_{ref8} = \frac{(10.4+(h_c\times T_a)-(U_2\times R_s))}{(-6.35-(2\times U_2))} + \frac{((5.3\times T_a)-1.6)}{((7.4+RH_n)\times(1.8-RH_x))} - \frac{U_2}{(75/((RH_x-T_x)-R_s))} + \frac{((T_a\times h_c)\times T_n)}{(((RH_a-U_2)+14)+(T_a-U_2))} + U_2 + 1.8$

Table 6
Statistical performance of the optimized GEP models during training and testing.

Model	Training				Testing			
	R^2 (%)	OI (%)	RMSE (mm d^{-1})	MAE (mm d^{-1})	R^2 (%)	OI (%)	RMSE (mm d^{-1})	MAE (mm d^{-1})
GEP-MOD1	64.4	92.2	3.10	2.29	63.6	77.9	3.19	2.40
GEP-MOD2	72.2	93.1	2.85	2.07	71.0	81.0	2.94	2.17
GEP-MOD3	76.1	93.9	2.64	1.98	76.8	84.2	2.65	1.98
GEP-MOD4	68.3	92.9	2.92	2.11	67.9	80.4	2.99	2.21
GEP-MOD5	97.8	96.1	1.98	1.46	89.3	89.6	2.09	1.53
GEP-MOD6	77.5	94.5	2.48	1.73	77.6	85.6	2.52	1.81
GEP-MOD7	82.2	95.4	2.20	1.64	82.6	88.6	2.21	1.63
GEP-MOD8	95.5	98.1	1.12	0.83	95.4	96.3	1.14	0.83

A similar procedure was applied to add either wind speed or solar radiation to GEP-MOD2. The R^2 increased drastically by 25.77%, from 71.0 to 89.3%, when wind speed was added to GEP-MOD2. However, the addition of solar radiation to GEP-MOD2 did not result in a significant increase in R^2 (9.29% increase). Furthermore, solar radiation slightly increased the R^2 by 7.55% when it was added to GEP-MOD3. This result indicates that solar radiation had an insignificant effect on the modelling of the ET_{ref}. GEP-MOD8 outperformed the other models by all of the performance criteria.

The developed GEP models were compared with the results obtained from PMG model. Fig. 5 compares the results on the testing data set, using a scatter plot of the estimated ET_{ref} values with the 45° exact model line. It is obvious from Fig. 5 that the GEP-MOD8 estimates were closer to the corresponding ET_{ref} values estimated by the PMG model than those of the other GEP models. Most of the GEP models underestimated the PMG ET_{ref} values when the values are greater than approximately 20 mm d^{-1}.

3.3. Comparison of the ANN and GEP models

In validation process, the performance of the models was further evaluated in other six stations located in the Kingdom of Saudi Arabia during the same period for training and testing processes. The performances of the ANN and GEP models are provided in Table 7, Figs. 6 and 7. Furthermore, Fig. 8 showed that the relationship between ANN models versus GEP models during validation process. As the results indicate the trend of the results at the 6 stations was the same as that of the 13 stations, which is taken in training and testing processes. The best performance criteria at the six stations were obtained by the input combinations containing U_2.

Table 7 showed that the ANN-MOD8 model with a R^2 and RMSE of 99.8% and 0.2 mm d^{-1} can be selected as the best model for ET_{ref} estimation to containing all the climatic items. The ANN-MOD5 model ranked second with a R^2 and RMSE of 98.9% and 0.56 mm d^{-1} when removing R_s from eighth input combi-

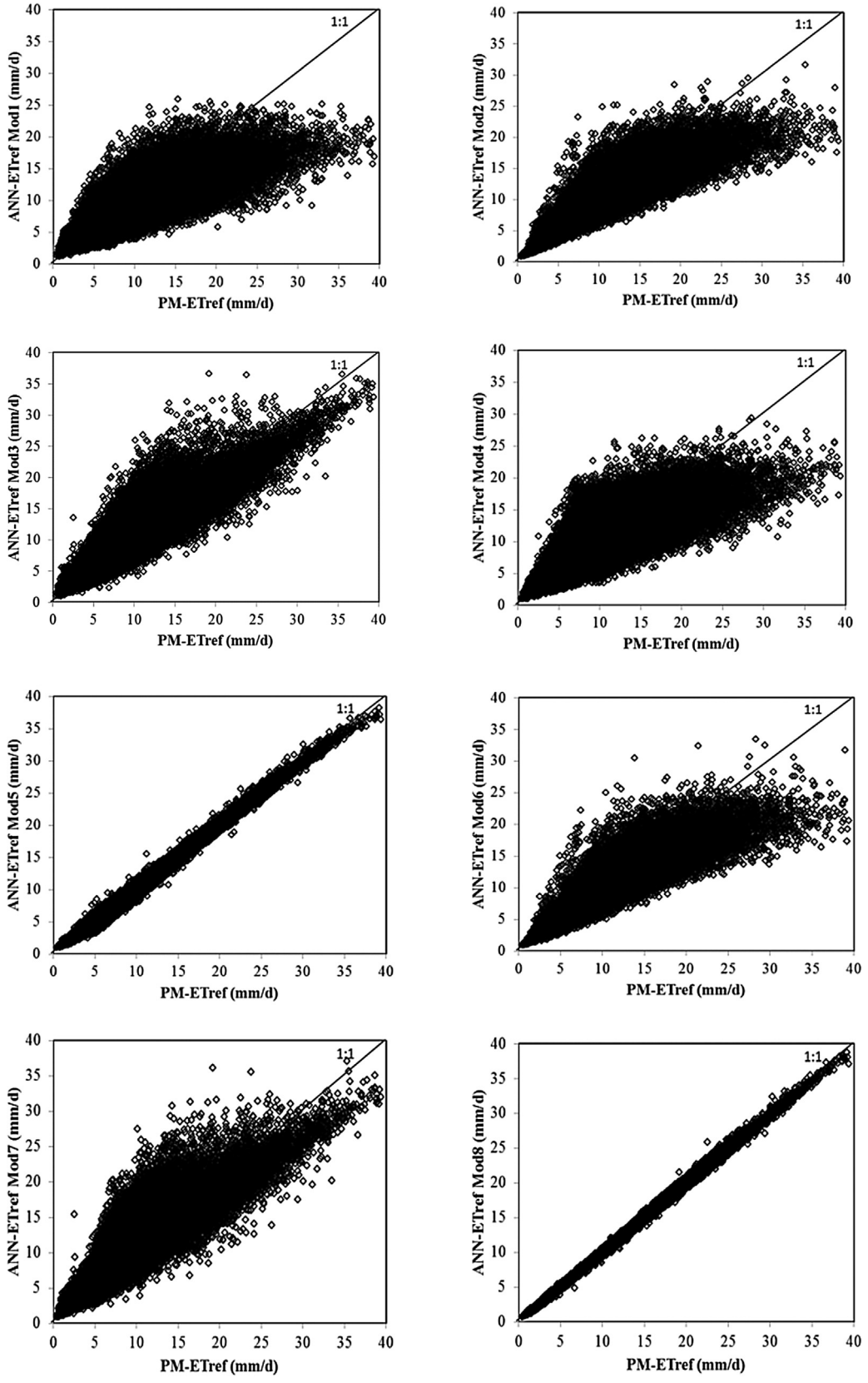

Fig. 6. Performance of ANN models during validation process, using the data collected from 1980 to 2010 by six weather stations.

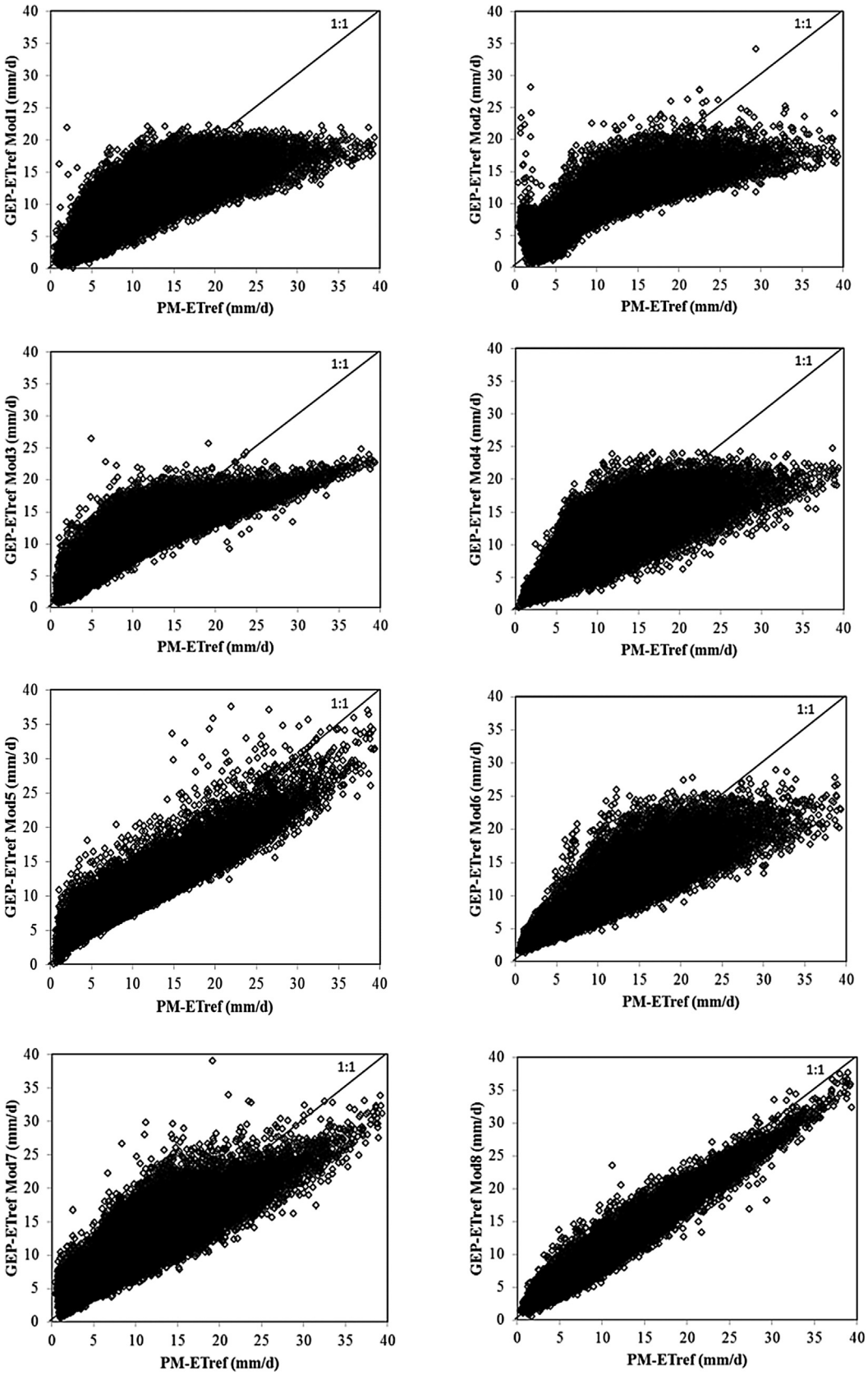

Fig. 7. Performance of GEP models during validation process, using the data collected from 1980 to 2010 by six weather stations.

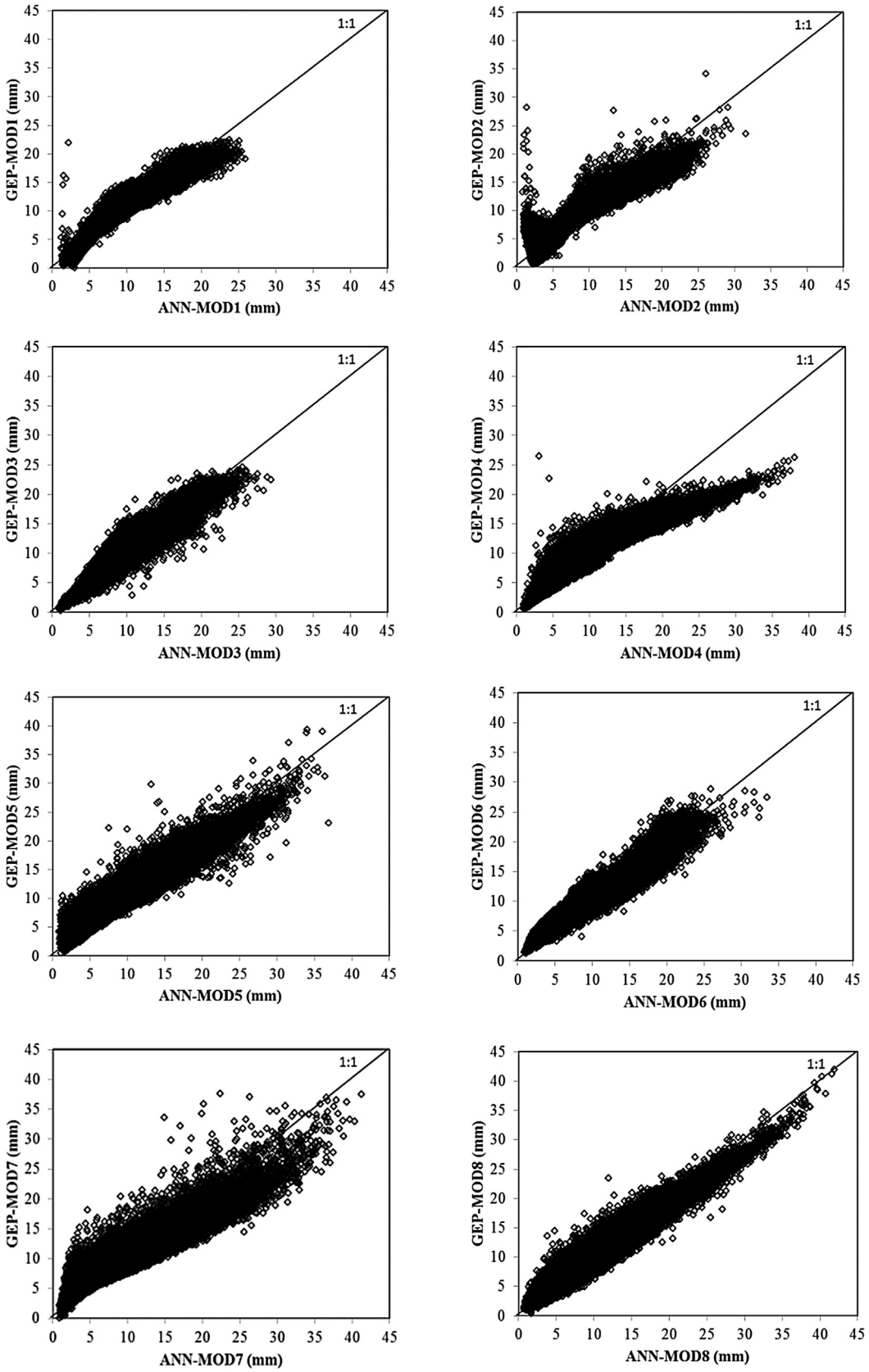

Fig. 8. Performance of ANN versus GEP models during validation process, using the data collected from 1980 to 2010 by six weather stations.

Table 7
Statistical performance of the optimized ANN and GEP models during validation, using data collected from 1980 to 2010 by six weather stations.

ANN models	R^2 (%)	OI (%)	RMSE (mm d^{-1})	MAE (mm d^{-1})	GEP models	R^2 (%)	OI (%)	RMSE (mm d^{-1})	MAE (mm d^{-1})
ANN-MOD1	66.6	79.2	3.19	2.23	GEP-MOD1	64.0	78.3	3.27	2.35
ANN-MOD2	83.1	87.5	2.41	1.61	GEP-MOD2	73.6	81.2	3.02	2.04
ANN-MOD3	83.9	89.2	2.22	1.59	GEP-MOD3	71.3	81.0	3.04	2.31
ANN-MOD4	66.5	79.9	3.13	2.19	GEP-MOD4	65.4	79.4	3.17	2.26
ANN-MOD5	98.9	98.8	0.56	0.44	GEP-MOD5	87.3	89.8	2.15	1.62
ANN-MOD6	85.1	88.8	2.27	1.45	GEP-MOD6	81.0	86.3	2.54	1.74
ANN-MOD7	83.1	88.3	2.32	1.63	GEP-MOD7	78.4	85.4	2.63	1.97
ANN-MOD8	99.8	99.7	0.21	0.15	GEP-MOD8	95.6	96.2	1.21	0.89

nation. The GEP-MOD8 model with R^2 and RMSE of 95.6% and 1.21 mm d^{-1} can be considered as the third best model. The GEP-MOD5, ANN-MOD6, ANN-MOD3, ANN-MOD2, ANN-MOD7 and GEP-MOD3 models ranked 4th place to 9th, respectively. The R^2 values of 9 models were the more than 80%. By looking at the Figs. 6 and 7 noted that the best model is achieved when there are all the climatic variables as in the eighth combination which is significantly close results with the PMG model. It is also noted the superiority in ANN models than GEP models. It is found the models containing U_2 are the best in performance.

Generally, the use of ANN models give high accuracy in the estimate but the use of GEP models is easier in the application to give it an explicit algebraic equation in calculating ET_{ref}. The main advantages of using ANN are their flexibility and ability to model non-linear relationships. Many studies have proven the superiority of ANN and GEP approaches for estimating ET_{ref} with minimal climatic parameters. Hence, based on the current results, the approaches presented here would allow more accurate estimations without the need for the availability of all data.

4. Conclusions

The main objective of this paper was to assess the performance of ANN and GEP models for estimating daily ET_{ref} at the KSA conditions. Eight combinations of the daily climate variables, maximum, mean, and minimum air temperature; maximum, mean, and minimum relative humidity; wind speed; solar radiation; and crop height, were used as inputs for the ANN and GEP techniques. The ET_{ref} was estimated from the PMG equation and used as a target variable. Nineteen meteorological stations were chosen from all regions of the KSA, representing all of the climatic conditions, Al-Qasim, Ha'il, Al-Jouf, Rafha, Dhahran, Najran, Jizan, Bisha, Al-Baha, Jeddah, Al-Madina, Tabuk, Turaif, Al-Wajh, Qaisumah, Yanba', Al-Ta'if and Wadi Al-Dawasir. Their daily climatic data collected from 1980 to 2010. Our results suggested that the derived ANN and GEP techniques should be used if meteorological stations supply an incomplete data set, through the lack or loss of some climatic variables, as the models gave estimated ET_{ref} values that were very close to the standard ET_{ref} values for the Saudi Arabian climatic conditions. The ANN models gave the most accurate estimates, but the GEP models are easier to use, as they calculate the ET_{ref} using explicit algebraic equations.

Acknowledgement

With sincere respect and gratitude, we would like to express deep thanks to Deanship of Scientific Research, King Saud University and Agriculture Research Center, College of Food and Agriculture Sciences for the financial support, sponsoring and encouragement.

References

Alazba, A.A., 2004. Estimating palm water requirements using Penman–Monteith mathematical model. J. King Saud Univ. 16 (2), 137–152.
Alazba, A., Mattar, M.A., El-Nesr, M.N., Amin, M.T., 2012. Field assessment of friction head loss and friction correction factor equations. J. Irrig. Drain. Eng. – ASCE 138 (2), 166–176.
Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration guidelines for computing crop water requirements. In: FAO Irrigation and Drainage, Paper No. 56, Food and Agriculture Organization of the United Nations, Rome.
Allen, R.G., Walter, I.A., Elliott, R.L., Howell, T.A., Itenfisu, D., Jensen, M.E., Snyder, R.L., 2005. The ASCE standardized reference evapotranspiration equation. In: Task Committee on Standardization of Reference Evapotranspiration of the EWRI of the ASCE, USA.
Arca, B., Beniscasa, F., Vincenzi, M., 2004. Evaluation of neural network techniques for estimating evapotranspiration", National Research Council – Research Institute for the Monitoring of Agroecosystems (IMAes), via Funtana di Lu Colbu 4/A, 07100 Sassari, Italy.
Aytek, A., Kişi, Ö., 2008. Genetic programming approach to suspended sediment modelling. J. Hydrol. 351 (3–4), 288–298.
Md. Azamathulla, H., Ghani, A.A., Leow, C.S., Chang, K.C., Zakaria, N.A., 2011. Gene-expression programming for the development of a stage–discharge curve of the pahang river. Water. Resour Manage. 25, 2901–2916.
Md. Azamathulla, H., Ahmad, Z., 2012. Gene expression programming for transverse mixing coefficient. J. Hydrol. 434–435, 142–148.
Bruton, J.M., McClendon, R.W., Hoogenboom, G., 2000. Estimating daily pan evaporation with artificial neural networks. Trans. ASAE 43 (2), 491–496.
Chiew, F.H.S., Kamaladassa, N.N., Malano, H.M., McMahon, T.A., 1995. Penman–Monteith, FAO-24 reference crop evapotranspiration and class-A pan data in Australia. Agric. Water Manage. 28, 9–21.
Dai, X., Shi, H., Li, Y., Ouyang, Z., Huo, Z., 2009. Artificial neural network models for estimating regional reference evapotranspiration based on climate factors. Hydrol. Process. 23, 442–450.
Dawson, W.C., Wilby, R., 1998. An artificial neural networks approach to rainfall-runoff modelling. Hydrol. Sci. J. 43 (1), 47–66.
De Souza, F., Yoder, R.E., 1994. ETestimation in the north east of Brazil: Hargreaves or Penman–Monteith equation. In: Proceedings, Technical PaperASAE International Winter Meeting, American Society of Agricultural Engineers, St. Joseph, Mich.
Eslamian, S.S., Gohari, S.A., Zareian, M.J., Firoozfar, A., 2012. Estimating Penman–Monteith reference evapotranspiration using artificial neural networks and genetic algorithm: a case study. Arab. J. Sci. Eng. 37, 935–944.
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, second ed. Springer-Verlag, Germany.
Fisher, J.B., Terry, A., DeBiase, A., Qi1, Y., Xu, M., Allen, H., 2005. Evapotranspiration models compared on a Sierra Nevada forest ecosystem. Environ. Modell. Softw. 20, 783–796.
Gavilan, P., Berengena, J., Allen, R.G., 2007. Measuring versus estimating net radiation and soil heat flux: impact on Penman–Monteith reference ET estimates in semiarid regions. Agric. Water Manage. 89, 275–286.
Gavilan, P., Lorite, I.J., Tornero, S., Berengena, J., 2006. Regional calibration of Hargreaves equation for estimating reference ET in a semiarid environment. Agric. Water Manage. 81, 257–281.
Ghani, A.A., Md. Azamathulla, H., 2011. Gene expression programming for sediment transport in sewer pipe systems. J. Pipeline Syst. Eng. Pract. 2 (3), 102–106.
Gocic, M., Trajkovic, S., 2010. Software for estimating reference evapotranspiration using limited weather data. Comp. Electron. Agric. 71, 158–162.
Gorka, L., Amais, O.B., Jose, J.L., 2008. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). Agric. Water Manage. 95, 553–565.
Guven, A., Aytek, A., 2009. A new approach for stage–discharge relationship: gene-expression programming. J. Hydrol. Eng. 14 (8), 812–820.
Haykin, S., 1999. Neural Networks. A Comprehensive Foundation. Prentice Hall International Inc., New Jersey.
Huo, Z., Feng, S., Kang, S., Dai, X., 2012. Artificial neural network models for reference evapotranspiration in an arid area of northwest China. J. Arid Environ. 82, 81–90.

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, R.G., Jones, J.W., 2003. Solar and net radiation based equations to estimate reference evapotranspiration in humid climates. J. Irrig. Drain. Eng. – ASCE 129 (5), 336–347.
Izadifar, Z., 2010. Modelling and Analysis of Actual Evapotranspiration Using Data Driven and Wavelet Techniques. Master Thesis. University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
Jain, S.K., Nayak, P.C., Sudhir, K.P., 2008. Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation. Hydrol. Process. 22 (13), 2225–2234.
Kişi, Ö., 2007. Evapotranspiration modelling from climatic data using a neural computing technique. Hydrol. Process. 21, 1925–1934.
Kişi, Ö., Guven, A., 2010. Evapotranspiration modelling using linear genetic programming technique. J. Irrig. Drain. Eng. 136 (10), 715–723.
Kişi, Ö., Özturk, Ö., 2007. Adaptive neuro-fuzzy computing technique for evapotranspiration estimation. J. Irrig. Drain. Eng. – ASCE 133 (4), 368–379.
Kumar, M., Raghuwanshi, N.S., Singh, R., 2011. Artificial neural networks approach in evapotranspiration modelling: a review. Irrig. Sci. 29, 11–25.
Kumar, M., Raghuwanshi, N.S., Singh, R., Wallender, W.W., Pruitt, W.O., 2002. Estimating evapotranspiration using artificial neural network. J. Irrig. Drain. Eng. – ASCE 128 (4), 224–233.
Landeras, G., Barredo, A.O., Lopez, J.J., 2008. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). Agric. Water Manage. 95, 553–565.
Legates, D.R., McCabe, J., 1999. Evaluating the use of "goodness-of fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35 (1), 233–241.
Mattar, M.A., Alamoud, A.I., 2015. Artificial neural networks for estimating the hydraulic performance of labyrinth-channel emitters. Comput. Electron. Agric. 114 (5), 189–201.
Mattar, M.A., Alazba, A.A., Zin El-Abedin, T.K., 2015. Forecasting furrow irrigation infiltration using artificial neural networks. Agric. Water Manage. 148, 63–71.
Naoum, S., Tsanis, K.I., 2003. Hydroinformatics in evapotranspiration estimation. Environ. Modell. Softw. 18, 261–271.
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. King Saudi Arabia: Facts and Numbers, edition1.
Shiri, J., Kişi, O., Landeras, G., Lopez, J.J., Nazemi, A.H., Stuyt, L., 2012. Daily reference evapotranspiration modelling by using genetic programming approach in the Basque Country (Northern Spain). J. Hydrol. 414–415, 302–316.
Sudheer, K.P., Gosain, A.K., Ramasastri, K.S., 2003. Estimating actual evapotranspiration from limited climatic data using neural computing technique. J. Irrig. Drain. Eng. – ASCE 129, 214–218.
Terzi, O., 2013. Daily pan evaporation estimation using gene expression programming and adaptive neural-based fuzzy inference system. Neural Comput. Appl. 23, 1035–1044.
Traore, S., Guven, A., 2013. New algebraic formulations of evapotranspiration extracted from gene-expression programming in the tropical seasonally dry regions of West Africa. Irrig. Sci. 31 (1), 1–10.
Trajkovic, S., Todorovic, B., Satanical, M., 2003. Forecasting reference evapotranspiration by artificial neural networks. J. Irrig. Drain. Eng. – ASCE 129, 454–457.
Trajkovic, S., 2005. Temperature-based approaches for estimating reference evapotranspiration. J. Irrig. Drain Eng. ASCE 131 (4), 316–323.
Traore, S., Wang, Y., Kerh, T., 2010. Artificial neural network for modelling reference evapotranspiration complex process in Sudano–Sahelian Zone. Agric. Water Manage. 97, 707–714.
Xiaoying, L., Erda, L., 2005. Performance of the Priestley–Taylor equation in the semi-arid climate of North China. Agric. Water Manage. 71, 1–17.
Zahiri, A., Eghbali, P., 2012. Gene expression programming for prediction of flow discharge in compound channels. J. Civil Eng. Urban. 2 (4), 164–169.
Zanetti, S.S., Sousa, E.F., Oliveira, V.P.S., Almeida, F.T., Bernardo, S., 2007. Estimating evapotranspiration using artificial neural network and minimum climatological data. J. Irrig. Drain. Eng. – ASCE 133 (2), 83–89.