Integrating genetic algorithm and support vector machine for modeling daily reference evapotranspiration in a semi-arid mountain area

Zhenliang Yin, Xiaohu Wen, Qi Feng, Zhibin He, Songbing Zou and Linshan Yang

ABSTRACT

Accurate estimation of evapotranspiration is vitally important for management of water resources and environmental protection. This study investigated the accuracy of integrating genetic algorithm and support vector machine (GA-SVM) models using climatic variables for simulating daily reference evapotranspiration (ET_0). The developed GA-SVM models were tested using the ET_0 calculated by Penman–Monteith FAO-56 (PMF-56) equation in a semi-arid environment of Qilian Mountain, northwest China. Eight models were developed using different combinations of daily climatic data including maximum air temperature (T_{max}), minimum air temperature (T_{min}), wind speed (U_2), relative humidity (RH), and solar radiation (R_s). The accuracy of the models was evaluated using root mean square error (RMSE), mean absolute error (MAE), and correlation coefficient (r). The results indicated that the GA-SVM models successfully estimated ET_0 with those obtained by the PMF-56 equation in the semi-arid mountain environment. The model with input combinations of T_{min}, T_{max}, U_2, RH, and R_s had the smallest value of the RMSE and MAE as well as higher value of r (0.995) compared to other models. Relative to the performance of support vector machine (SVM) models and feed-forward artificial neural network models, it was found that the GA-SVM models proved superior for simulating ET_0.

Key words | climatic variables, genetic algorithm, reference evapotranspiration modeling, semi-arid mountain areas, support vector machine

Zhenliang Yin
Xiaohu Wen (corresponding author)
Qi Feng
Zhibin He
Songbing Zou
Linshan Yang
Key Laboratory of Ecohydrology of Inland River Basin,
Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences,
Lanzhou,
Gansu 730000,
China
E-mail: *xhwen@lzb.ac.cn*

INTRODUCTION

Evapotranspiration is the process of water transportation from the Earth's surface to the atmosphere including the evaporation process and transpiration process. It is a vital component of the hydrological cycle and water balance computing, which affects water resources management and planning (Traore *et al.* 2010; Shiri *et al.* 2013; Xing *et al.* 2016). Accurate estimation of evapotranspiration is crucial for water resource management, weather and climate studies, hydrology, reliable irrigation design, and determination of the water budget, especially in water shortage regions, for example, arid and semi-arid areas where water resources are under severe threat by overexploitation (Huo *et al.* 2013; Chatzithomas & Alexandris 2015; Shiri *et al.* 2015).

Evapotranspiration can be quantified either by the experimental method or mathematical method. It can be measured immediately by some instruments and equipment. However, this method is difficult, expensive and time-consuming (Pour Ali Baba *et al.* 2013; Ma *et al.* 2015; Zhang *et al.* 2015). The mathematical approach has to be preceded by the estimation of reference evapotranspiration (ET_0) (Xu *et al.* 2015). Many empirical and semi-empirical models have been developed to estimate ET_0 by applying meteorological data. Among the several methods for

doi: 10.2166/nh.2016.205

estimating ET_0, the Penman–Monteith FAO-56 (PMF-56) equation has been recommended as the normal equation to estimate ET_0 by the Food and Agriculture Organization of the United Nations (FAO). The PMF-56 equation has become the benchmark against all other ET_0 models, based on physical methods and requiring a large number of climatic variables, such as daily minimum temperature and maximum temperature, relative humidity, solar radiation, and wind speed. However, data such as weather variables are usually incomplete or not always acquirable for many regions and thus limit the effective use of the PMF-56 model (Kumar *et al.* 2002; Cobaner 2011).

ET_0 is an open, nonlinear, complex, and dynamic phenomenon due to its dependence on the interaction of several climatic elements (Luo *et al.* 2014). Thus, it is difficult to derive a definite equation to express all the related physical processes. As an alternative to conventional approaches, artificial neural network (ANN) is highly suitable for modeling the nonlinear processes (He *et al.* 2014; Deo & Şahin 2015; Si *et al.* 2015; Wen *et al.* 2015a). Many researchers have applied ANN for estimating ET_0 (Traore *et al.* 2010; Marti *et al.* 2011; Laaboudi *et al.* 2012; Yassin *et al.* 2016). These studies indicated that the ANN models were more superior in modeling ET_0 compared to the conventional methods such as Hargreaves, Priestley–Taylor, and some empirical and semi-empirical equations (Kumar *et al.* 2002; Landeras *et al.* 2008; Huo *et al.* 2012; Wen *et al.* 2015b). However, neural networks have some disadvantages such as training slowly, requiring a large amount of training data, and easily getting stuck in a local minimum (Principe *et al.* 2000). Support vector machine (SVM) is a novel learning machine based on a statistical learning theory and a structural risk minimization principle, which has been successfully applied for modeling the nonlinear system (Shiri *et al.* 2014; Feng *et al.* 2015). Given the same training conditions, SVM provides more dependable and better performance when compared to ANN (Gill *et al.* 2006; Çimen & Kisi 2009; Yoon *et al.* 2011). Over the last decade, SVM models have been used in a very wide range of applications to solve hydrological problems (Chou *et al.* 2010; Tan *et al.* 2012; Kalra *et al.* 2013; Wen *et al.* 2015b). Recently, researchers began to employ SVM for ET_0 modeling. Kişi & Cimen (2009) discussed the potential of SVM models in estimating ET_0 in central California, USA. The results showed that the SVM model could be built in as a module for estimating ET_0 values in a hydrological model. Kişi (2013) tested the capacities of a least square support vector machine (LSSVM) for modeling ET_0. It was found that LSSVM performed better than the ANN models and the empirical models in simulating ET_0 processes. Lin *et al.* (2012) established a SVM-based model for daily pan evaporation estimating and compared it with an ANN-based model. They found that the SVM was superior to ANN in modeling evaporation. Tabari *et al.* (2012, 2013) studied the potential of SVM for estimating ET_0 in a highland semi-arid environment in Iran. The results demonstrated that the SVM models achieved better ET_0 estimations than the regression- and climate-based models. Wen *et al.* (2015b) evaluated the application of SVM to model daily ET_0 by limited climatic data in an extremely arid region in northwest China. They drew the conclusion that the SVM could provide better performance when compared to ANN, Priestley–Taylor, Hargreaves, and Ritchie. These studies indicated that SVM could be applied to estimate ET_0, with relatively better performance than ANN in simulating the ET_0 process. Although possessing excellent features, SVM is limited in modeling ET_0 research because the users must define a great many parameters appropriately (Liu *et al.* 2011). The estimation accuracy and efficiency of the SVM depends on the hyper-parameters being set correctly. However, selecting the most appropriate training parameter value is a critical problem for application of SVM, which can affect the model performance of the SVM. Genetic algorithm (GA), a general suited optimization search approach based on a direct analogy to Darwinian natural genetics and selection in biological systems, can be used to generate appropriate solutions to optimize and search problems. Thus, the GA could be used to select appropriate SVM parameters (Abdullah *et al.* 2015). In recent research, GA has been applied to optimize SVM parameters in different fields (Pourbasheer *et al.* 2009; Liu & Jiao 2011; Chen *et al.* 2016); however, applications of GA-support vector machine for modeling ET_0 are limited (Shiri *et al.* 2011; Tao *et al.* 2015; Liu *et al.* 2016). Therefore, the proposed integration of the genetic algorithm and support vector machine (GA-SVM) model was applied to modeling the daily ET_0 in this paper, in which GA was used to optimize the parameters of the SVM. Furthermore, there have been few

studies conducted under the climatic conditions of the semi-arid environment of Qilian Mountain.

The main purposes of this study were to investigate the accuracy of GA-SVM models for estimating daily ET_0 with various combinations of daily meteorological data including: minimum air temperature (T_{min}), maximum air temperature (T_{max}), wind speed (U_2), relative humidity (RH), and solar radiation (R_s) in the semi-arid environment of Qilian Mountain, northwest China. Furthermore, the conventional grid algorithm-based SVM model and ANN model were also investigated for comparison.

MATERIALS AND METHODS

SVM

SVM is an innovative type of machine learning method developed to resolve both regression problems and classification. Support vector regression (SVR) is generally applied to describe regression, such as in the case of regression approximation, giving a set of data points $G = \{(x_i, y_i)\}n_i$ (x_i is the input vector; y_i the desired value; and n is the total number of data). To deal with nonlinear regression work, the inputs are foremost nonlinearly mapped into a high-dimension feature space where they are correlated linearly with the outputs. The following linear calculation function was applied to formalize SVR

$$f(x) = \omega \cdot \phi(x) + b \tag{1}$$

where ω is weight vector, b is a constant; $\Phi(x)$ denotes a mapping function in the feature space. The coefficients ω and b are estimated by minimizing

$$R_{reg}(f) = C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(f(x_i), y_i) + \frac{1}{2}\|w\|^2 \tag{2}$$

$$L_\varepsilon(f(x) - y) = \begin{cases} |f(x) - y| - \varepsilon & for |f(x) - y| \geq \varepsilon \\ 0 & otherwise \end{cases} \tag{3}$$

where both C and ε are regulation parameters. The first term $L_\varepsilon(f(x_i),y_i)$ is called the ε-intensive loss function. This function indicates that errors below ϵ are not penalized. The term $C\ 1/N\sum_{i=1}^{N} L_\varepsilon(f(x_i), y_i)$ is the empirical error. The term 1/2 measures the smoothness of the function. C evaluates the trade-off between the empirical risk and the smoothness of the model. The positive slack variables ξ and $\xi*$ represent the distance from the actual values to the corresponding boundary values of ε-tube. Equation (2) is transformed to the following constrained formation:

$$\text{minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i*) \tag{4}$$

$$\text{subject to } \begin{cases} y_i - (\langle w, x_i\rangle + b) \leq \varepsilon + \xi_i \\ \langle w, x_i\rangle + b - y_i \leq \varepsilon + \xi_i* \\ \xi_i, \xi_i* \geq 0 \end{cases} \tag{5}$$

After taking the Lagrangian and optimal conditions, a nonlinear regression function is obtained using the following expression:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i*)k(x_i, x) + b \tag{6}$$

where α_i and α_i* are the above-mentioned Lagrange multipliers. With the utilization of the Karush–Kuhn–Tucker conditions, only a limited number of coefficients will not be zero among α_i and α_i*. The related data points could be referred to the support vectors. $k(x_i, x)$ refers to kernel function describing the inner product in the D-dimension feature space.

$$k(x, y) = \sum_{i=1}^{D}\phi_j(x)\phi_i(y) \tag{7}$$

It shows that any symmetric kernel function k satisfying Mercer's condition corresponds to a dot product in some feature space (Boser *et al.* 1992). In this paper, radius basis function (RBF) is selected as the kernel function. The RBF is defined as follows:

$$k(x, y) = \exp\left(-\gamma\|x - y\|^2\right), \lambda > 0. \tag{8}$$

when the RBF kernel is used, three parameters including penalty parameter C, and kernel function's parameter γ are

considered. The general performance of SVM models depends on a proper setting of these parameters.

GA

GA is an adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics developed by Holland (1975). The procedure of GA simulates the processes of selection, crossover, and mutation to maintain superior solutions and to generate better and better offspring, making the solutions close to the objective function. Selection is performed to choose excellent chromosomes in the population for reproduction (Gao & Hou 2016). The better fit the chromosome, the more likely it will be selected. Crossover is performed randomly to choose a locus between two chromosomes to create two offspring. Mutation is performed randomly to flip some bits in a chromosome. In this paper, the parameters C and γ of SVM were optimized by GA.

The construction of SVM model based on GA optimization parameters (GA-SVM) is described below.

Step 1. Encode the SVM parameters:

The SVM parameters C and γ are directly coded to form chromosomes.

Step 2. Generate a random initial population:

Randomly generate an initial population of chromosomes which represent the SVM parameters C and γ.

Step 3. Evaluate fitness:

In this study, the five-fold cross-validation methodology was used to assess fitness (Fang *et al.* 2008). In the five-fold cross-validation methodology, the training data set is divided at random into five mutually exclusive subsets (folds) of approximately equal size. The performance of the parameters (C, γ) is measured by the root mean square error (RMSE) on the last subset. The above step is repeated five times to guarantee that each subset is used once for validation. Averaging the RMSE of the five trials obtains an estimation of the expected generalization error for training sets of size (4 * m/5), where m is the number of training data sets. As a consequence of this, the fitness function can be defined as the RMSE cross-validation on the training data set, and expressed as follows:

$$Fitness = \min\left(RMSE_{\text{cross-validation}}\right) = \min\left(\frac{1}{5}\sum_{j=1}^{5}\sqrt{\frac{5}{m}\sum_{i=1}^{m/5}(obs_i - pre_i)^2}\right) \quad (9)$$

where obs_i and pre_i are the observed value and predicted value, respectively.

Step 4. Genetic operators:

The real-valued GA uses selection, crossover, and mutation operators to generate the offspring of the existing population. Excellent chromosomes are selected from a population according to the fitness to yield offspring in the next generation. The roulette wheel selection principle is applied to choose chromosomes for reproduction. In crossovers, single point crossover is randomly adopted to exchange genes between two chromosomes. The mutation operation follows the crossover operation, and determines whether a chromosome should be mutated in the next generation.

Step 5. Stop condition:

If the stop condition is satisfied, the optimization will stop and return the best parameters C and γ. Otherwise, go back to step 3.

ANN

ANN is a massively parallel distributed information processing system that has certain performance characteristics resembling biological neural networks of the human brain. A neural network is characterized by its architecture that represents the pattern of connection between nodes, its method of determining the connection weights and the activation function. The most commonly used neural network structure is the feed forward hierarchical architecture. ANN customary architecture is composed of three layers. Many theoretical and experimental works have shown that a single hidden layer is sufficient for ANNs to approximate any complex nonlinear function. The Levenberg–Marquardt training algorithm was used to train the ANN model in our

research. The sigmoid and linear activation functions were used for the hidden and output node(s), respectively.

PMF-56 equation

In this paper, the PMF-56 was used to provide the GA-SVM targets to train and test the GA-SVM models. As the sole standard method for the computation of ET_0, the PMF-56 method is described by Allen *et al.* (1998):

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\ 900/T_{mean} + 273U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \tag{10}$$

where ET_0 is the reference evapotranspiration (mm day^{-1}); R_n is the net radiation (MJ m^{-2} day^{-1}); G is the soil heat flux (MJ m^2 day^{-1}); γ is the psychrometric constant (kPa $^\circ C^{-1}$); e_s is the saturation vapor pressure (kPa); e_a is the actual vapor pressure (kPa); Δ is the slope of the saturation vapor pressure–temperature curve (kPa $^\circ C^{-1}$); T_{mean} is the average daily air temperature ($^\circ C$); and U_2 is the mean daily wind speed at 2 m (m s^{-1}).

Model performance criteria

The performances of the models developed in this research were assessed using various standard statistical performance evaluation criteria. To evaluate the performance of SVM models, three statistical criteria were used. The considered statistical measures were: coefficient of correlation (r), RMSE, and mean absolute error (MAE). The r measures the degree to which two variables are linearly related. RMSE and MAE provide different types of information on the predictive capabilities of the model. The RMSE measures the goodness-of-fit relevant to high ET_0 values whereas the MAE yields a more balanced perspective of the goodness-of-fit at moderate value distribution of the estimation errors.

The following equations were used for the computation of the above parameters:

$$r = \frac{\sum_{i=1}^{n}\left(ET_{0i}^{p} - \overline{ET_0^p}\right)\left(ET_{0i}^{o} - \overline{ET_0^o}\right)}{\sqrt{\sum_{i=1}^{n}\left(ET_{0i}^{p} - \overline{ET_0^p}\right)^2 \sum_{i=1}^{n}\left(ET_{0i}^{o} - \overline{ET_0^o}\right)^2}} \tag{11}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(ET_{0i}{}^{p} - ET_{0i}{}^{0}\right)^2}{n}} \tag{12}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|ET_{0i}^{p} - ET_{0i}^{o}\right| \tag{13}$$

where ET_{0i}^{p} and ET_{0i}^{o} are the ith estimated and PMF-56 ET_0 values, respectively; $\overline{ET_0^p}$ and $\overline{ET_0^o}$ are the average of ET_{0i}^{p} and ET_{0i}^{o}; and n is the total number of data. The best fit between observed and calculated values would have $r = 1$, RMSE = 0, and MAE = 0, respectively.

In view of the requirements of the GA-SVM computation algorithm, the raw data of both the input and output variables were normalized to an interval by transformation. All the variables were normalized ranging from 0 to 1 as the following equation:

$$x_n = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{14}$$

where x_n and x_i represent the normalized and original training and testing data; x_{min} and x_{max} denote the minimum and maximum of the training and testing data.

CASE STUDY

Study area and data

The research was carried out in the Pailugou Watershed of Qilian Mountain, a small and typical mountainous valley in the Qilian mountains, northwest China. The area of the study watershed is 2.95 km^2, with a body length of 4.25 km. The elevation ranges from 2,600 m to 3,800 m, with a longitudinal slope of 1:4.19. The climate in this region is temperate, with a mean annual temperature of 0.5 $^\circ C$ and relative humidity of 60%. The mean annual precipitation is 368 mm, mainly occurring from May to October. The climatic data observation site located at 100 $^\circ$17′E, 38°24′N, with an elevation of 2,700 m above sea level, is typical of mountain grassland with moderate slope in the watershed of Qilian Mountain. The observation was implement by an IMKO ENVIS

environmental measurement system which was installed in the observation site to continually observe the primary climatic parameters from June 1, 2009 to December 31, 2011 (Figure 1). The climatic data, including air temperature, dew point temperature, wind velocity, net radiation, relative humidity, water vapor pressure, soil heat flux, soil temperature, and solar radiation were automatically recorded at 30 min intervals, and averaged into daily values. This resulted in a total of 944 daily averaged climatic parameters.

The daily climatic data employed in this study were composed of: minimum air temperature (T_{min}), maximum air temperature (T_{max}), wind speed (U_2), relative humidity (RH), and solar radiation (R_s). The data were divided into two sets: the data from June 1, 2009 to December 31, 2010, 549 records (about 60% of total data), were used for training the models, and the remaining 365 records from January 1, 2011 to December 31, 2011 (about 40%) were used for testing. A full year data set used in the identification period enabled inclusion of various hydrological conditions that are observable during different seasons of the year. In this way the model became robust for the different hydrological conditions that prevail in the total time series (Kişi 2006). The statistical parameters of daily climatic data are shown in Table 1. The minimum and maximum values of ET_0 used in the training periods ranged from 0.01 to 4.60 mm; however, the maximum value of the testing periods' ET_0 is 5.11 mm, which may cause difficulties in estimating the high ET_0 values.

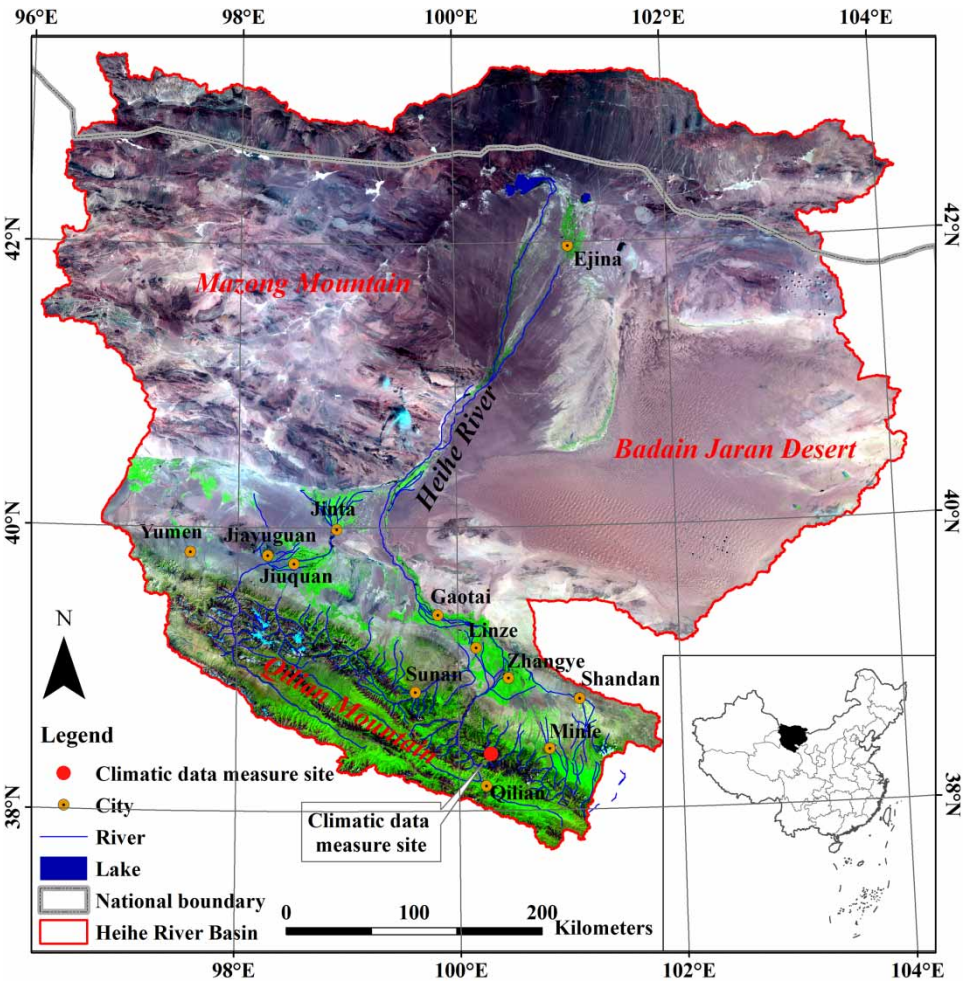

Figure 1 | Location study area and the climatic data measured site.

Table 1 | Statistical parameters of climatic data and the PMF-56 ET_0 in each data set

Climatic data and the PMF-56 ET_0	Minimum	Maximum	Mean	Std	SK
T_{min} (°C)					
All	−24.70	15.90	−2.89	9.63	−0.22
Training	−23.60	15.90	−2.53	9.45	−0.15
Testing	−24.70	13.10	−3.43	9.88	−0.30
T_{max} (°C)					
All	−21.00	29.00	7.98	10.60	−0.26
Training	−21.00	28.20	8.28	10.40	−0.23
Testing	−17.90	29.00	7.52	10.90	−0.29
RH (%)					
All	18.52	98.10	59.59	18.94	0.11
Training	19.75	98.10	61.30	18.49	0.06
Testing	18.52	97.49	57.01	19.33	0.22
U_2 (m s^{-1})					
All	0.46	2.08	1.16	0.26	0.07
Training	0.47	2.08	1.15	0.26	0.15
Testing	0.46	2.02	1.16	0.27	−0.05
R_s (MJ m^{-2} day^{-1})					
All	0.00	12.59	4.50	3.48	0.43
Training	0.00	12.59	4.39	3.50	0.47
Testing	0.00	12.06	4.65	3.45	0.37
PMF-56 ET_0 (mm day^{-1})					
All	0.01	5.11	1.44	1.25	0.73
Training	0.01	4.60	1.39	1.23	0.69
Testing	0.01	5.11	1.51	1.28	0.77

Std, standard deviation; SK, skewness.

Model development

The selection of appropriate input variables is important for the SVM model development since it provides the basic information on the system being modeled. In the current study, different input combinations of various daily climatic data including T_{max}, T_{min}, U_2, RH, and R_s were used as inputs to estimate the ET_0 obtained using the PMF-56 equation. Input 1 was designed as temperature-based models; the other input structures were formed by inserting wind speed, solar radiation, and relative humidity into the input 1 combination, respectively. Finally, eight input combinations evaluated in the present study were: (1) T_{max} and T_{min}; (2) T_{max}, T_{min}, and U_2; (3) T_{max}, T_{min}, and R_s; (4) T_{min}, T_{max}, and RH; (5) T_{min}, T_{max}, U_2, and RH; (6) T_{min}, T_{max}, U_2, and R_s; (7) T_{min}, T_{max}, RH, and R_s; (8) T_{min}, T_{max}, U_2, RH, and R_s. The GA-SVM, SVM, and ANN models were trained and tested for each combination.

RESULTS AND DISCUSSION

The parameters C and γ of SVM were optimized by GA. Performance statistics of the GA-SVM models for PMF-56 ET_0 for the training and testing periods are given in Table 2. It was found that the difference between the values of the statistical indices of the training and validation set did not vary substantially.

Table 2 | Optimal SVM parameters with GA and the performance statistics of the GA-SVM models during the training and testing periods

		Parameter		Training periods			Testing periods		
Model	Input	*C*	γ	*r*	*RMSE* mm/day	*MAE* mm/day	*r*	*RMSE* mm/day	*MAE* mm/day
GA-SVM1	1	5	9.64	0.940	0.422	0.294	0.948	0.424	0.311
GA-SVM2	2	3.01	1.32	0.959	0.353	0.247	0.972	0.314	0.241
GA-SVM3	3	4.13	8.81	0.975	0.273	0.151	0.990	0.201	0.147
GA-SVM4	4	15.11	0.53	0.948	0.390	0.287	0.955	0.396	0.298
GA-SVM5	5	1.68	1.24	0.963	0.331	0.234	0.971	0.316	0.241
GA-SVM6	6	0.69	2.61	0.977	0.263	0.143	0.993	0.163	0.124
GA-SVM7	7	3.88	8.07	0.985	0.213	0.113	0.991	0.175	0.132
GA-SVM8	8	29.13	0.27	0.980	0.249	0.137	0.995	0.138	0.106

Considering all models, according to the results of the testing periods, the values of RMSE, MAE, and r ranged from 0.424 to 0.138 mm/day, 0.311 to 0.106 mm/day, and 0.948 to 0.995, respectively. We observed that the r values of all models were higher than 0.94, pointing to a strong relation between estimated and PMF-56 ET_0 values. The RMSE and MAE values less than 0.424 and 0.311 mm/day, respectively, indicate good and appropriate forecast. It appears that all the GA-SVM models demonstrated a high generalization capacity for the proposed model with relatively low error and high correlation, exhibiting a high accuracy for estimating PMF-56 ET_0. GA-SVM8, whose input combinations included T_{min}, T_{max}, U_2, RH, and R_s had the smallest value of the RMSE (0.138 mm/day) and MAE (0.106 mm/day) as well as a higher value of r (0.995) than other models in the testing periods therefore, it was selected as the best-fit GA-SVM model to estimate the PMF-56 ET_0. All the models, GA-SVM8, GA-SVM7, GA-SVM6, and GA-SVM3 performed similarly since the values of RMSE and MAE did not vary significantly, and all r values were also very close to unity. They were found to be better than the GA-SVM1, GA-SVM2, GA-SVM4, and GA-SVM5 models in the modeling of PMF-56 ET_0. For practical uses, the GA-SVM8, GA-SVM7, GA-SVM6, and GA-SVM3 models had good accuracy in PMF-56 ET_0 modeling and the selection of one model over the other should be dependent upon the available meteorological data. Although the GA-SVM1 model, with only minimum and maximum air temperature as inputs had the highest error rates, its performance was good. For practical uses, the GA-SVM1 model can be used where only air temperature data are available. This is especially true in the mountain areas where reliable weather data sets of solar radiation, relative humidity, and wind speed are limited.

Comparing the different GA-SVM models, we can find that the models significantly improved the accuracy of GA-SVM1 when either solar radiation or wind speed is integrated as additional input variables. GA-SVM3 and GA-SVM2 models, which insert solar radiation and wind speed into inputs, respectively, improved r, RMSE, and MAE by 4.4%, 52.6%, and 52.7% and 2.5%, 22.9%, and 22.5%, respectively, in comparison to the GA-SVM1 model. The reduction of RMSE and MAE and increase of r obtained with relative humidity were less than those obtained from solar radiation and wind speed. Although some studies reported that wind speed was more effective for estimating ET_0 (Popova *et al.* 2006; Traore *et al.* 2010; Cobaner 2011), in this study, the results show that solar radiation is the more effective and required climatic variable for modeling the ET_0 in this semi-arid mountain environment with high accuracy.

The hydrograph and scatter plots of ET_0 values computed by the PMF-56 equation and the values estimated by the GA-SVM models are shown in Figure 2. The ET_0 values estimated by the GA-SVM models closely followed the PMF-56 ET_0 values and followed the same trend, revealing that both the models showed good estimation accuracy of the PMF-56 ET_0 (Figure 2).

In order to assess the ability of the GA-SVM model relative to the grid algorithm-based SVM model and feed-forward ANN model, the eight SVM models and eight feed-forward ANN models were developed using the same variables' combinations as done for the GA-SVM input combinations of (1)–(8) for the PMF-56 ET_0 modeling. The appropriate model structures were determined for each input combination, then, the SVM and ANN models were tested, and the results were compared by means of performance statistics.

The parameters C and γ of SVM were optimized by grid algorithm and performance statistics of the GA-SVM models for PMF-56 ET_0 for the testing period are given in Table 3.

Considering the performance of the testing periods, the RMSE values ranged from 0.433 to 0.148 mm/day, MAE values ranged from 0.321 to 0.114 mm/day, and r values ranged from 0.947 to 0.994. The lower RMSE, MAE and higher r values implied the good performance of the SVM model for PMF-56 ET_0 modeling. Among the SVM models, SVM8, SVM7, SVM6, and SVM3 were found to be better than the SVM1, SVM2, SVM4, and SVM5 models in the modeling of PMF-56 ET_0. In the case of the GA-SVM models, SVM8, whose inputs combinations were T_{min}, T_{max}, U_2, RH, and R_s, had the smallest value of the RMSE (0.148 mm/day) and MAE (0.114 mm/day) and a higher value of r (0.994) than other models in the test periods; therefore, it was the best-fit SVM model for estimating the PMF-56 ET_0.

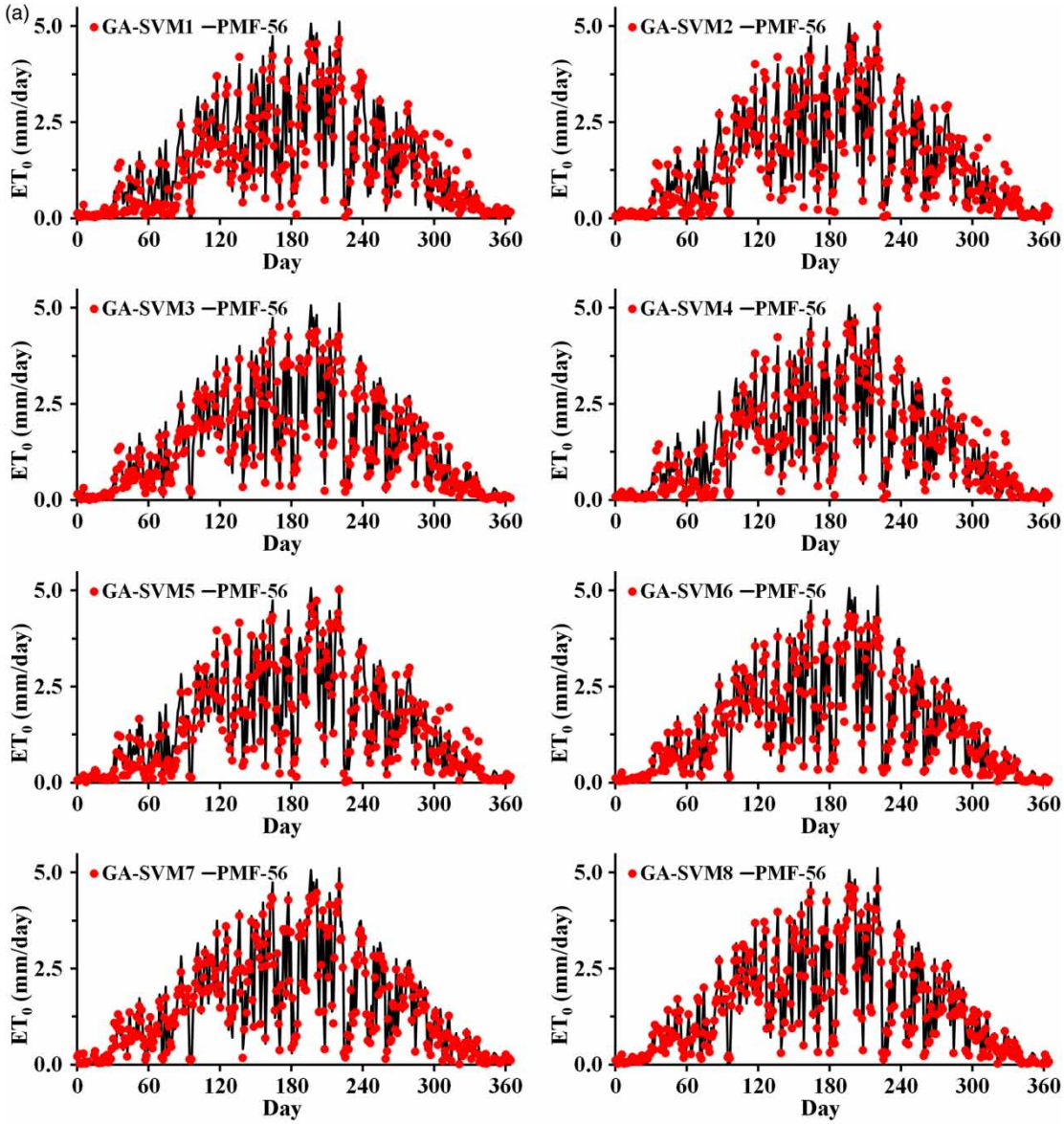

Figure 2 | Comparison of the ET_0 values estimated by the PMF-56 equation and the GA-SVM models during the testing periods. *(Continued.)*

The architectural identification of the ANN model is the primary important aspect of the modeling since inappropriate architecture may lead to under-fitting, over-fitting and computational overload. In the current research, the optimal number of neurons in the hidden layer was identified using a trial and error procedure by varying the number of hidden neurons from 2 to 20. Furthermore, the optimal network architecture was selected based on the one with minimum MSE. The final ANN architecture and the performance statistics of each model are shown in Table 4. It was observed that the ANN models produced slight variability in performance with the RMSE values varying from 0.473 to 0.176 mm/day, MAE values varying from 0.341 to 0.136 mm/day, and r values varying from 0.930 to 0.992 in the testing periods. ANN8, ANN7, ANN6, and ANN3 had similar performance that showed small differences between the RMSE, MAE, and r values. They were found to be better than the ANN1, ANN2, ANN4, and ANN5 models in the modeling of PMF-56 ET_0. Similar to GA-SVM and SVM models, the ANN8 model with inputs of T_{min}, T_{max}, U_2,

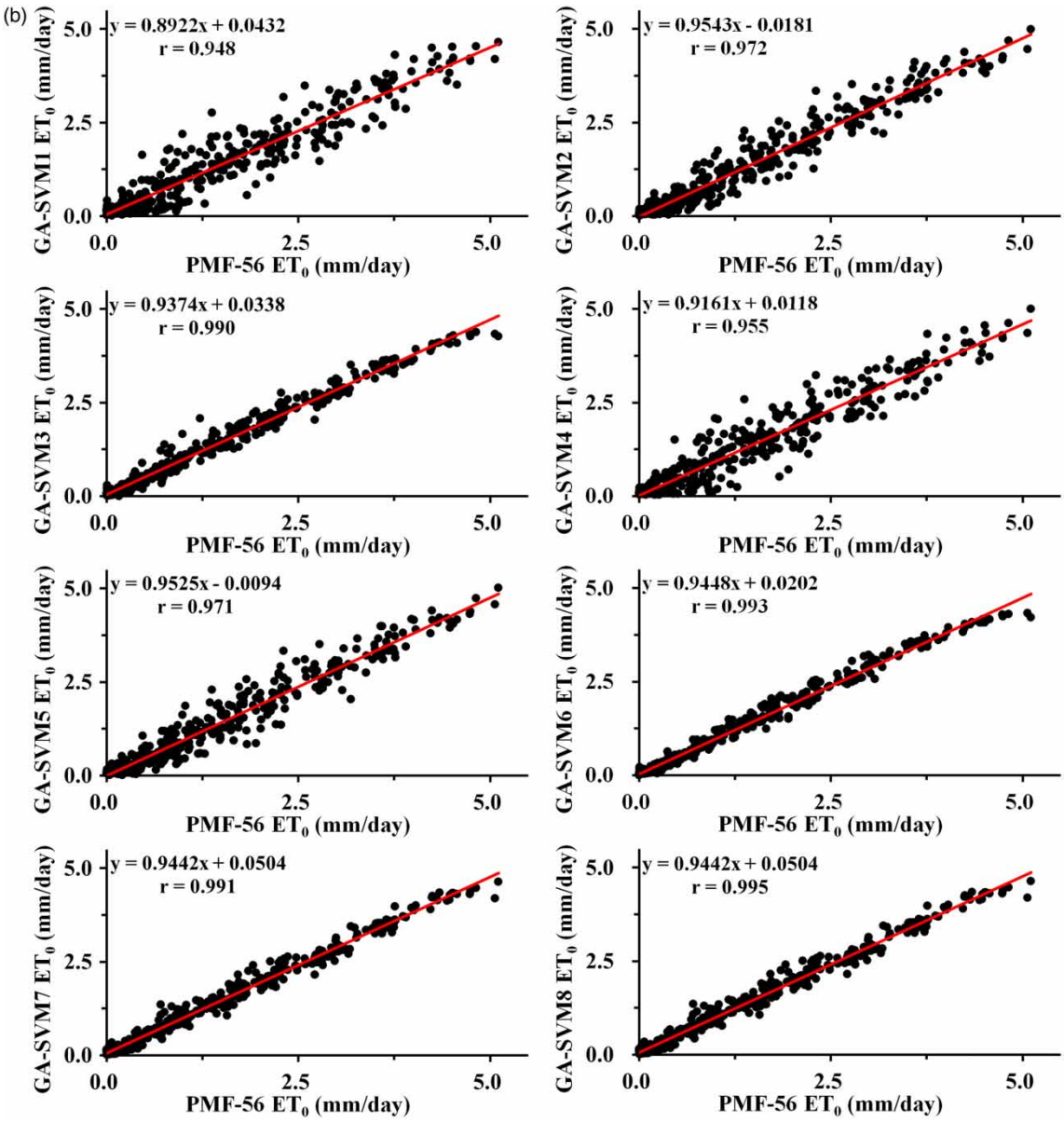

Figure 2 | Continued.

RH, and R_s had the best performance (RMSE = 0.176 mm/day, MAE = 0.136 mm/day, and $r = 0.992$) among the ANN models.

Comparing the performance criteria of the GA-SVM models (Table 2) with those of the SVM (Table 3) and ANN models (Table 4), it was observed that all three models generally gave low values of the RMSE and MAE as well as high *r*. The GA-SVM, SVM, and ANN models had good performance in PMF-56 ET_0 modeling. However, the comparison revealed that all the GA-SVM models performed little better than the corresponding SVM and ANN models in modeling the PMF-56 ET_0. The GA-SVM model produced a lower RMSE and MAE as well as a higher *r*, being the best according to the criteria. The SVM model had the second best performance and the ANN model was found to be the worst of all approaches investigated according to the criteria in this study.

Table 3 | Optimal SVM parameters with grid algorithm and the performance statistics of the SVM models during the training and testing periods

		Parameter		Training periods			Testing periods		
Model	**Input**	***C***	γ	***r***	***RMSE* mm/day**	***MAE* mm/day**	***r***	***RMSE* mm/day**	***MAE* mm/day**
SVM1	1	0.57	48.50	0.943	0.412	0.282	0.947	0.433	0.321
SVM2	2	0.57	5.28	0.961	0.342	0.237	0.969	0.329	0.251
SVM3	3	0.33	16	0.975	0.275	0.153	0.989	0.206	0.150
SVM4	4	1	5.73	0.953	0.374	0.268	0.951	0.404	0.308
SVM5	5	0.57	5.29	0.967	0.313	0.216	0.965	0.347	0.262
SVM6	6	1	9.19	0.980	0.245	0.124	0.992	0.179	0.129
SVM7	7	1	16	0.987	0.201	0.106	0.988	0.211	0.153
SVM8	8	0.58	9.19	0.986	0.205	0.104	0.994	0.148	0.114

Table 4 | The structure and the performance statistics of the ANN models during the training and testing periods

			Training periods			Testing periods		
Model	**Input**	**Structure**	***r***	***RMSE* mm/day**	***MAE* mm/day**	***r***	***RMSE* mm/day**	***MAE* mm/day**
ANN1	1	2-12-1	0.939	0.439	0.315	0.941	0.473	0.341
ANN2	2	3-4-1	0.959	0.352	0.259	0.967	0.361	0.280
ANN3	3	3-11-1	0.971	0.301	0.174	0.986	0.245	0.174
ANN4	4	3-13-1	0.952	0.380	0.289	0.930	0.482	0.349
ANN5	5	4-6-1	0.964	0.330	0.243	0.968	0.347	0.266
ANN6	6	4-4-1	0.974	0.280	0.176	0.990	0.224	0.169
ANN7	7	4-6-1	0.982	0.232	0.156	0.985	0.232	0.155
ANN8	8	5-11-1	0.984	0.223	0.141	0.992	0.176	0.136

Figures 3 and 4 show the hydrograph and scatter plots of both the PMF-56 ET_0 and the values obtained using the best SVM (SVM8) and ANN (ANN8) models of the testing period. It revealed that both models showed good prediction accuracy of the PMF-56 ET_0. Compared to Figure 2 for the GA-SVM8 model, the GA-SVM8 model showed better accuracy than SVM8 and ANN8 models. The performance of the SVM8 was better than that of the ANN8 model, further confirming that although the GA-SVM, SVM, and feed-forward ANN had comparable performance during testing period, the performance of the GA-SVM was superior to that of the SVM and feed-forward ANN model

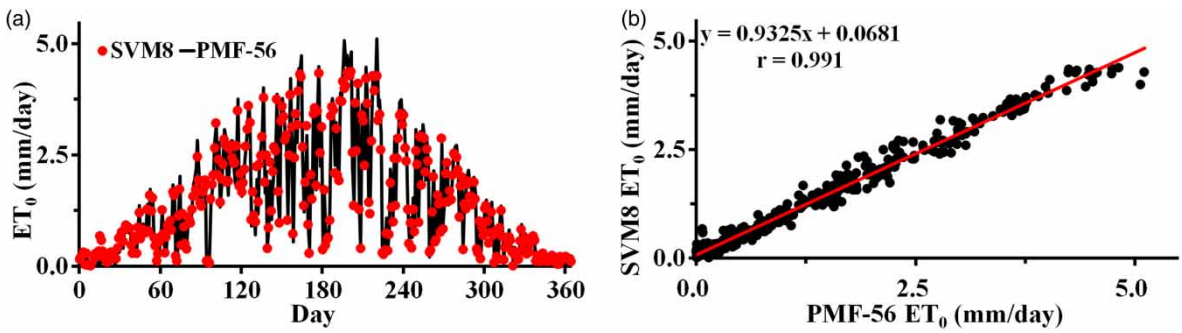

Figure 3 | Comparison of the ET_0 values estimated by the PMF-56 equation and the SVM8 model during the testing periods.

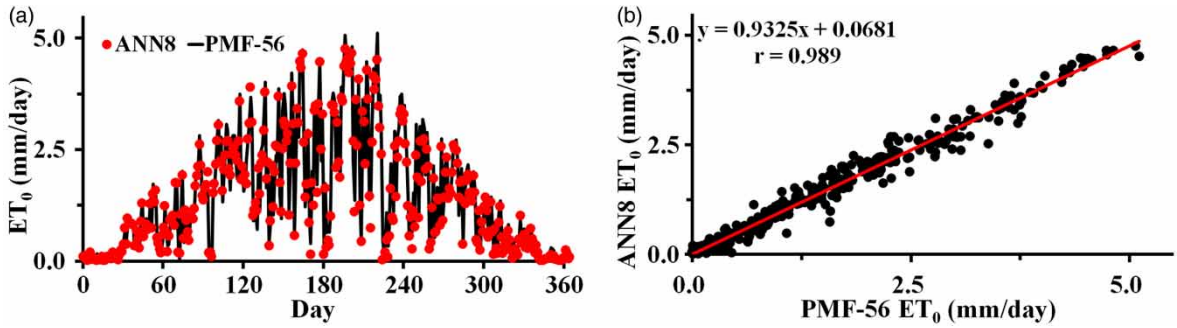

Figure 4 | Comparison of the ET_0 values estimated by the PMF-56 equation and the ANN8 model during the testing periods.

during the more important independent testing stage. All comparative analysis of the GA-SVM, SVM, and ANN models showed that the GA-SVM models performed a little better than the SVM and ANN models.

The estimation of total PMF-56 ET_0 obtained from the estimated ET_0 values was also considered for comparison due to its importance in water balance calculation, water resources planning and management. The total estimated ET_0 amounts in testing periods are given in Table 5, showing that most models underestimated total PMF-56 ET_0 value. The GA-SVM8, SVM8, and ANN8 models, whose input parameters were T_{min}, T_{max}, U_2, RH, and R_s, estimated the total PMF-56 ET_0 value of 551.19 mm as 538.96 mm, 537.94 mm, and 536.54 mm, with an underestimation of 2.22%, 2.40%, and 2.66%, respectively. The total PMF-56 ET_0 amount estimates of the GA-SVM8 and SVM8 were closer to the PMF-56 ET_0 values. Among the models, the GA-SVM8 model had the best estimate (−2.22%) and the SVM8 had the second best estimate (−2.40%).

Table 5 | Total ET_0 values and relative error calculated by various models during the testing periods

	GA-SVM		SVM		ANN	
Input	Total ET_0. (mm)	Relative error (%)	Total ET_0. (mm)	Relative error (%)	Total ET_0. (mm)	Relative error (%)
1	507.64	−7.90	502.39	−8.85	486.14	−11.80
2	519.51	−5.75	517.90	−6.04	495.39	−10.12
3	529.13	−4.00	528.85	−4.05	524.70	−4.81
4	509.35	−7.59	520.98	−5.48	516.94	−6.21
5	521.64	−5.36	520.88	−5.50	504.69	−8.44
6	528.21	−4.17	528.35	−4.14	513.22	−6.89
7	533.74	−3.17	535.93	−2.77	525.42	−4.68
8	538.96	−2.22	537.94	−2.40	536.54	−2.66

It is important to evaluate not only the average estimation error but also the distribution of estimation errors when assessing the performance of any model for its applicability in modeling ET_0. Comparing the monthly ET_0 estimates of the best GA-SVM, SVM, ANN model with the PMF-56 ET_0 estimates, Figure 5 contains biases, i.e., model estimate monthly ET_0 minus the corresponding PMF-56 ET_0, for the testing data set. These confirmed the relative superiority of the GA-SVM models in comparison with either the SVM and ANN models for modeling the ET_0.

Overall, the GA-SVM, SVM, and ANN models can give good prediction performance and can be successfully applied to establish models that could provide accurate and reliable PMF-56 ET_0 modeling. The results suggested that the GA-SVM models were superior to the SVM and ANN in the PMF-56 ET_0 modeling.

Although this study tried to seek the best models in GA-SVM, SVM, and ANN models, the truth is that the more climate variables used to train the models, the more accurate the results in estimating PMF-56 ET_0. In ungauged regions, the climate data are limited; however, the temperature data are easy to access and process, and in this situation, the GA-SVM, SVM, and ANN models with only minimum and maximum air temperature as inputs can all be used to estimate PMF-56 ET_0.

CONCLUSIONS

The objective of this paper was to investigate the accuracy of integrating GA-SVM model using climatic variables for

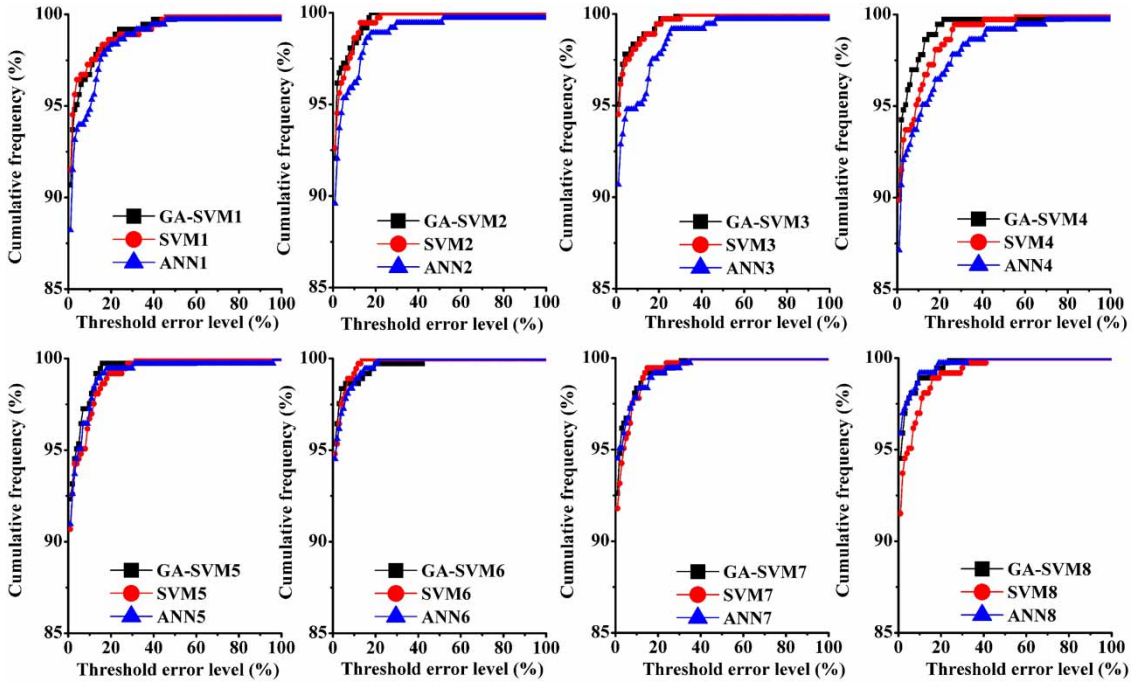

Figure 5 | Error (mm) in estimated monthly ET_0 by different model assumptions during the testing periods.

modeling of daily reference evapotranspiration (ET_0), in which GA was used to optimize the parameters of the SVM. The developed GA-SVM models were tested using the ET_0 calculated by PMF-56 equation of a semi-arid environment of Qilian Mountain, northwest China. Eight models were developed using different combinations of daily climatic data including maximum air temperature (T_{max}), minimum air temperature (T_{min}), wind speed (U_2), relative humidity (RH), and solar radiation (R_s). The results showed that the GA-SVM models exhibited a successful performance for estimating ET_0. The GA-SVM model whose combinations inputs included T_{min}, T_{max}, U_2, RH, and R_s had the best accuracy for the estimation of PMF-56 ET_0.

In order to assess the ability of the GA-SVM model relative to the conventional grid algorithm-based SVM model and feed-forward ANN model, the eight SVM models and eight feed-forward ANN models were developed using the same variables combinations as done for the GA-SVM input combinations for comparison. The results showed that the proposed GA-SVM model had better accuracy performance than the grid algorithm-based SVM model and ANN model in modeling the PMF-56 ET_0. This study suggests that the GA-SVM model provides accurate PMF-56 ET_0 modeling and can be successfully applied to PMF-56 ET_0 modeling in a semi-arid mountain area where evapotranspiration measurements or the complete climatic data for applying the PMF-56 method are often not available.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31370466, 41522102, 41571031), and the technology innovation team of China Academy of Science (Cross and Cooperation), and the Foundation for Excellent Youth Scholars of CAREERI, CAS, and the China Postdoctoral Science Foundation (Grant No. 2015M572620). The authors wish to thank the anonymous reviewers for reading the manuscript, and for their suggestions and critical comments.

REFERENCES

Abdullah, S. S., Malek, M. A., Abdullah, N. S., Kisi, O. & Yap, K. S. 2015 Extreme learning machines: a new approach for

prediction of reference evapotranspiration. *J. Hydrol.* **527**, 184–195.

Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. 1998 *Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements,* FAO Irrigation and Drainage Paper 56. FAO, Rome, Italy.

Boser, B. E., Guyon, I. M. & Vapnik, V. N. 1992 A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, ACM Press, New York, USA, pp. 144–152.

Chatzithomas, C. D. & Alexandris, S. G. 2015 Solar radiation and relative humidity based, empirical method, to estimate hourly reference evapotranspiration. *Agr. Water Manage.* **152**, 188–197.

Chen, P., Yuan, L. F., He, Y. G. & Luo, S. 2016 An improved SVM classifier based on double chains quantum genetic algorithm and its application in analogue circuit diagnosis. *Neurocomputing* (in press). doi: 10.1016/j.neucom.2015.12.131.

Chou, P. H., Wu, M. J. & Chen, K. K. 2010 Integrating support vector machine and genetic algorithm to implement dynamic wafer quality prediction system. *Expert Syst. Appl.* **37** (6), 4413–4424.

Çimen, M. & Kisi, O. 2009 Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. *J. Hydrol.* **378** (3–4), 253–262.

Cobaner, M. 2011 Evapotranspiration estimation by two different neuro-fuzzy inference systems. *J. Hydrol.* **398** (3–4), 292–302.

Deo, R. C. & Şahin, M. 2015 Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia. *Atmos. Res.* **161–162**, 65–81.

Fang, S. F., Wang, M. P., Qi, W. H. & Zheng, F. 2008 Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials. *Comp. Mater. Sci.* **44** (2), 647–655.

Feng, Q., Wen, X. H. & Li, J. G. 2015 Wavelet analysis–support vector machine coupled models for monthly rainfall forecasting in arid regions. *Water Resour. Manage.* **29** (4), 1049–1065.

Gao, X. & Hou, J. 2016 An improved SVM integrated GS–PCA fault diagnosis approach of Tennessee Eastman process. *Neurocomputing* **174**, 906–911.

Gill, M. K., Asefa, T., Kemblowski, M. W. & McKee, M. 2006 Soil moisture prediction using support vector machines. *J. Am. Water. Resour. Assoc.* **42** (4), 1033–1046.

He, Z. B., Wen, X. H., Liu, H. & Du, J. 2014 A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *J. Hydrol.* **509**, 379–386.

Holland, J. H. 1975 *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor, MI, USA.

Huo, Z., Feng, S., Kang, S. & Dai, X. 2012 Artificial neural network model for reference evapotranspiration in an arid area of northwest China. *J. Arid Environ.* **82**, 81–90.

Huo, Z., Dai, X., Feng, S., Kang, S. & Huang, G. 2013 Effect of climate change on reference evapotranspiration and aridity index in arid region of China. *J. Hydrol.* **492**, 24–34.

Kalra, A., Ahmad, S. & Nayak, A. 2013 Increasing streamflow forecast lead time for snowmelt-driven catchment based on large-scale climate patterns. *Adv. Water Resour.* **53**, 150–162.

Kişi, Ö. 2006 Daily pan evaporation modelling using a neuro-fuzzy computing technique. *J. Hydrol.* **329**, 636–646.

Kişi, O. 2013 Least squares support vector machine for modeling daily reference evapotranspiration. *Irrigation Sci.* **31** (4), 611–619.

Kişi, O. & Cimen, M. 2009 Evapotranspiration modelling using support vector machines. *Hydrolog. Sci. J.* **54** (5), 918–928.

Kumar, M., Raghuwanshi, N. S., Singh, R., Wallender, W. W. & Pruitt, W. O. 2002 Estimating evapotranspiration using artificial neural network. *J. Irrig. Drain. E.-ASCE* **128**, 224–233.

Laaboudi, A., Mouhouche, B. & Draoui, B. 2012 Neural network approach to reference evapotranspiration modeling from limited climatic data in arid regions. *Int. J. Biometeorol.* **56** (5), 831–841.

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). *Agr. Water Manage.* **95**, 553–565.

Lin, G. F., Lin, H. Y. & Wu, M. C. 2012 Development of a support-vector-machine-based model for daily pan evaporation estimation. *Hydrol. Process.* **27** (22), 3115–3127.

Liu, H. B. & Jiao, Y. B. 2011 Application of genetic algorithm-support vector machine (GA-SVM) for damage identification of bridge. *Int. J. Comput. Intell. Appl.* **10** (4), 383–397.

Liu, S., Tai, H., Ding, Q., Li, D., Xu, L. & Wei, Y. 2011 A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. *Math. Comput. Model.* **58** (s3–4), 458–465.

Liu, Y., Yu, M., Ma, X. Y. & Xing, X. G. 2016 Estimating models for reference evapotranspiration with core meteorological parameters via path analysis. *Hydrol. Res.* doi: 10.2166/nh.2016.240.

Luo, Y. F., Chang, X. M., Peng, S. Z., Khan, S., Wang, W. G., Zheng, Q. & Cai, X. L. 2014 Short-term forecasting of daily reference evapotranspiration using the Hargreaves–Samani model and temperature forecasts. *Agr. Water Manage.* **136**, 42–51.

Ma, N., Zhang, Y. S., Xu, C. Y. & Szilagyi, J. 2015 Modeling actual evapotranspiration with routine meteorological variables in the data-scarce region of the Tibetan Plateau: comparisons and implications. *J. Geophys. Res.* **120** (8), 1638–1657.

Marti, P., Manzano, J. & Royuela, A. 2011 Assessment of a 4-input artificial neural network for ET_0 estimation through data set scanning procedures. *Irrigation Sci.* **29** (3), 181–195.

Popova, Z., Kercheva, M. & Pereira, L. S. 2006 Validation of the FAO methodology for computing ET_0 with limited data, application to south Bulgaria. *Irrig. Drain.* **55** (2), 201–215.

Pour Ali Baba, A., Shiri, J., Kisi, O., Fakheri Fard, A., Kim, S. & Amini, R. 2013 Estimating daily reference evapotranspiration using available and estimated climatic data by adaptive

neuro-fuzzy inference system and artificial neural networks. *Hydrol. Res.* **44** (1), 131–146.

Pourbasheer, E., Riahi, S., Reza Ganjali, M. & Norouzi, P. 2009 Application of genetic algorithm-support vector machine (GA-SVM) for prediction of BK-channels activity. *Eur. J. Med. Chem.* **44**, 5023–5028.

Principe, J. C., Euliano, N. R. & Lefebvre, C. W. 2000 *Neural and Adaptive Systems: Fundamentals Through Simulations*. Wiley, New York, USA.

Shiri, J., Dierickx, W., Pour Ali Baba, A., Nemati, S. & Ghorbani, M. A. 2011 Estimating daily pan evaporation from climatic data of the state of Illinois, USA using adaptive neuro-fuzzy inference system and artificial neural network. *Hydrol. Res.* **42** (6), 491–502.

Shiri, J., Sadraddini, A. A., Nazemi, A. H., Kisi, O., Marti, P., Fard, A. F. & Landeras, G. 2013 Evaluation of different data management scenarios for estimating daily reference evapotranspiration. *Hydrol. Res.* **44** (6), 1058–1070.

Shiri, J., Nazemi, A. H., Sadraddini, A. A., Landeras, G., Kisi, O., Fard, A. F. & Marti, P. 2014 Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. *Comput. Electron. Agr.* **108**, 230–241.

Shiri, J., Marti, P., Nazemi, A. H., Sadraddini, A. A., Kisi, O., Landeras, G. & Fard, A. F. 2015 Local vs. external training of neuro-fuzzy and neural networks models for estimating reference evapotranspiration assessed through k-fold testing. *Hydrol. Res.* **46** (1), 72–88.

Si, J. H., Feng, Q., Wen, X. H., Xi, H. Y., Yu, T. F., Li, W. & Zhao, C. Y. 2015 Modeling soil water content in extreme arid area using an adaptive neuro-fuzzy inference system. *J. Hydrol.* **527**, 679–687.

Tabari, H., Kisi, O., Ezani, A. & Hosseinzadeh Talaee, P. 2012 SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *J. Hydrol.* **444–445**, 78–89.

Tabari, H., Martinez, C., Ezani, A. & Hosseinzadeh Talaee, P. 2013 Applicability of support vector machines and adaptive neurofuzzy inference system for modeling potato crop evapotranspiration. *Irrigation Sci.* **31** (4), 575–588.

Tan, G., Yan, J., Gao, C. & Yang, S. 2012 Prediction of water quality time series data based on least squares support vector machine. *Procedia Engineering* **31**, 1194–1199.

Tao, X. E., Chen, H., Xu, C. Y., Hou, Y. K. & Jie, M. X. 2015 Analysis and prediction of reference evapotranspiration with climate change in Xiangjiang River Basin, China. *Water Sci. Eng.* **8** (4), 273–281.

Traore, S., Wang, Y. M. & Kerh, T. 2010 Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. *Agr. Water Manage.* **97** (5), 707–714.

Wen, X. H., Feng, Q., Yu, H. J., Wu, J., Si, J. H., Chang, Z. Q. & Xi, H. Y. 2015a Wavelet and adaptive neuro-fuzzy inference system conjunction model for groundwater level predicting in a coastal aquifer. *Neural Comput. Appl.* **26** (5), 1203–1215.

Wen, X. H., Si, J. H., He, Z. B., Wu, J., Shao, H. B. & Yu, H. J. 2015b Support-vector-machine-based models for modeling daily reference evapotranspiration with limited climatic data in extreme arid regions. *Water Resour. Manage.* **29**, 3195–3209.

Xing, X. G., Liu, Y., Zhao, W. G., Kang, D. G., Yu, M. & Ma, X. Y. 2016 Determination of dominant weather parameters on reference evapotranspiration by path analysis theory. *Comput. Electron. Agr.* **120**, 10–16.

Xu, L. H., Shi, Z. J., Wang, Y. H., Zhang, S. L., Chu, X. Z., Yu, P. T., Xiong, W., Zuo, H. J. & Wang, Y. N. 2015 Spatiotemporal variation and driving forces of reference evapotranspiration in Jing River Basin, northwest China. *Hydrol. Process.* **29** (23), 4846–4862.

Yassin, M. A., Alazba, A. A. & Mattar, M. A. 2016 Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate. *Agr. Water Manage.* **163**, 110–124.

Yoon, H., Jun, S. C., Hyun, Y., Bae, G. O. & Lee, K. K. 2011 A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* **396** (1–2), 128–138.

Zhang, K. X., Pan, S. M., Zhang, W., Xu, Y. H., Cao, L. G., Hao, Y. P. & Wang, Y. 2015 Influence of climate change on reference evapotranspiration and aridity index and their temporal-spatial variations in the Yellow River Basin, China, from 1961 to 2012. *Quatern. Int.* **380–381**, 75–82.

First received 7 April 2016; accepted in revised form 14 July 2016. Available online 9 September 2016