

Daily global solar radiation modeling using data-driven techniques and empirical equations in a semi-arid climate

Saeed Samadianfard, Abolfazl Majnooni-Heris, Sultan Noman Qasem, Ozgur Kisi, Shahaboddin Shamshirband & Kwok-wing Chau

To cite this article: Saeed Samadianfard, Abolfazl Majnooni-Heris, Sultan Noman Qasem, Ozgur Kisi, Shahaboddin Shamshirband & Kwok-wing Chau (2019) Daily global solar radiation modeling using data-driven techniques and empirical equations in a semi-arid climate, *Engineering Applications of Computational Fluid Mechanics*, 13:1, 142-157, DOI: [10.1080/19942060.2018.1560364](https://doi.org/10.1080/19942060.2018.1560364)

To link to this article: <https://doi.org/10.1080/19942060.2018.1560364>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 07 Jan 2019.



Submit your article to this journal [↗](#)



Article views: 186



View Crossmark data [↗](#)

Daily global solar radiation modeling using data-driven techniques and empirical equations in a semi-arid climate

Saeed Samadianfard^a, Abolfazl Majnooni-Heris^a, Sultan Noman Qasem^{b,c}, Ozgur Kisi^d, Shahaboddin Shamshirband ^{e,f} and Kwok-wing Chau^g

^aDepartment of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran; ^bComputer Science Department, College of Computer and Information Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia; ^cComputer Science Department, Faculty of Applied Sciences, Taiz University, Taiz, Yemen; ^dFaculty of Natural Sciences and Engineering, Iliia State University, Tbilisi, Georgia; ^eDepartment for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam; ^fFaculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam; ^gDepartment of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hong Kong, People's Republic of China

ABSTRACT

Solar radiation, moisture and temperature are the most vital meteorological variables which affect plant growth. Due to the fact that the global solar radiation (GSR) is scarcely gauged at meteorological stations in developing countries, it is commonly estimated by data-driven techniques or by empirical equations. In this study, support vector regression (SVR), model trees (MT), gene expression programming (GEP) and adaptive neuro-fuzzy inference system (ANFIS) and several empirical equations were applied to assess the relations between GSR and several meteorological variables including minimum temperature (T_{min}), maximum temperature (T_{max}), relative humidity (RH), sunshine hours (n), maximum sunshine hours (N), corrected clear-sky solar irradiation (ICSKY), day of year (DOY) and extra-terrestrial radiation (R_g). For this purpose, the daily GSR measured from the beginning of 2011 to the end of 2013 at Tabriz synoptic station, which is located in semi-arid regions of Iran, were used. A direct strong relationship was observed to exist between the GSR and n. For evaluating the performances of studied techniques, three different statistical indicators were used namely root mean square error (RMSE), mean absolute error (MAE), and correlation coefficient (CC). Additionally, a Taylor diagram was utilized to test the similarity between the observed and predicted GSR values. Results indicated that the SVR-6 with input parameters of R_g , RH, T_{min} , T_{max} , n/N had better accuracy in predicting GSR with RMSE of 1.656, MAE of 0.990, CC of 0.980 and WI of 0.990 than the other models. Moreover, MT-6 ranked as the second best model in the prediction of GSR values. As an interesting point, studied empirical equations had lower accuracies comparing with the SVR, GEP, MT and ANFIS methods. For instance, GSR values were computed by Angstrom and Prescott equation, as the best empirical equation, with RMSE of 1.786, MAE of 1.156, CC of 0.977 and WI of 0.988. Conclusively, results from the current study proved that the SVR provided reasonable trends for GSR modeling at Tabriz synoptic station. Furthermore, MT models with linear equations can be implemented with a high degree of simplicity and acceptable precision in GSR estimation.

ARTICLE HISTORY

Received 25 September 2018
Accepted 14 December 2018

KEYWORDS

Data-driven technics;
empirical equations;
meteorological parameters;
global solar radiation

1. Introduction

Energy, which plays a vital role in the current societies, accelerates economic developments and has been thought to be as one of the global critical issues in the last few decades (Gairaa, Khellaf, Messlem, & Chellali, 2016). Furthermore, by sharp decreasing of the world reserves oil and also due to its high pollution, many believe that solar energy is one of the best substitutions of fossil fuels according to its unique characteristics such as worldwide accessibility and environmental friendly features (Shaddel, Javan, & Baghernia, 2016). Solar energy is mainly utilized to design solar systems

(Qazi et al., 2015), radiant floor cooling systems (Feng, Schiavon, & Bauman, 2016), environmental and agricultural studies (Kaufmann & Hagermann, 2015; Lamnatou & Chemisana, 2013) and managing the effects of global warming (Ming, De_Richter, Liu, & Caillol, 2014). However, despite the broad range of applications of solar energy, direct measuring of solar radiation is not available in most countries, especially developing ones. Furthermore, in some regions, sensors of solar radiation have not been installed in the meteorological stations. Even in some stations with these sensors, the measured data could be missing or inaccurate due to technical

CONTACT Shahaboddin Shamshirband  shahaboddin.shamshirband@tdtu.edu.vn

problems. So, the common practice in the mentioned situations is to use mathematical, empirical or recently data-driven techniques, which have been established on the basis of measured meteorological parameters, to have precise estimations of actual solar radiation (Ozgoren, Bilgili, & Sahin, 2012; Sun, Zhao, Zeng, & Yan, 2015). Data-driven techniques have the numerous applications in hydrological engineering and contemporary real-life problems (Chau, 2017; Fotovatikhah et al., 2018; Jekhoni, Delirhasannia, Alavipanah, Shahabi, & Samadianfard, 2015; Moazenzadeh, Mohammadi, Shamshirband, & Chau, 2018; Samadianfard, Delirhasannia, Kisi, & Agirre-Basurko, 2013; Samadianfard, Nazemi, & Sadradini, 2014; Samadianfard, Sattari, Kisi, & Kazemi, 2014; Taormina, Chau, & Sivakumar, 2015; Wu & Chau, 2011). So, the researchers attempt to develop accurate predictor models of solar radiation in various time scales that is a crucial issue of related solar energy sciences.

Data-driven techniques have been widely used for predicting solar radiation over the last decade. The key merit of them is that, unlike the empirical equations, they do not need a primary model of the associations among input and output data. Tymvios, Jacovides, Michaelides, and Scouteli (2005) analyzed the capabilities of artificial neural network (ANN) models comparing with Angstrom-Preccott linear equation for estimating daily global solar radiation (GSR) using input parameters in Athalassa, Cyprus. The obtained results proved that the ANN model, using climatological parameters of n , N and T_{\max} , provided more precise estimates of GSR than the studied empirical equations. Benghanem, Mellit, and Alamri (2009) developed different ANN models, using meteorological parameters of sunshine hours (n), temperature (T), relative humidity (RH) and day of year (DOY), for predicting GSR in Madinah, Saudi Arabia. They reported that the ANN model with input parameters of n and T produced accurate predictions with the correlation coefficient (CC) of 0.976. Koca, Oztop, Varol, and Koca (2011) tested the effects of different combinations of meteorological parameters in accurate estimation of GSR using ANN models in seven cities of Turkey. The results showed that the predictions of Antakya were more precise than those of the other studied cities. Additionally, they stated that the coordinates of the cities and their sunshine hours had undeniable effects on the accuracy of the GSR predictions. Wu and Liu (2012) estimated the GSR values using temperature parameters by developing a support vector machine (SVM) based models for 24 stations in China. The SVM model with input combinations of T and $T_{\max}-T_{\min}$ provided lower error meters. Mostafavi, Saeidi Ramiyani, Sarvar, Izadi Moud, and Mousavi (2013) proposed a new hybrid approach involving genetic programming and simulated annealing for

estimating GSR values. By performing sensitivity analysis, they found that T_{\min} and T_{\max} had vital effects on accurate estimation of solar radiation. In other research, Ramedani, Omid, Keyhani, Khoshnevisan, and Saboohi (2014) inspected the capabilities of several kernel functions of support vector regression (SVR) in estimating GSR values in Tehran. They stated that SVR with radial basis function had superior predictions than the adaptive neuro-fuzzy inference system (ANFIS) and ANN models. Kisi (2014) examined the capabilities of fuzzy genetic approach (FG) in GSR modeling of seven cities of Turkey. The coordinates of the studied locations and month number of the year were utilized as input parameters of FG and the results were compared with the results of ANN and ANFIS methods. He reported that the accuracy of FG outputs was superior to correspondent ANN and ANFIS predictions. In another study, SVR models with radial basis and polynomial kernel functions have been utilized by Piri, Shamshirband, Petkovic, Tong, and Rehman (2015) for estimating GSR using climatological parameters of n , minimum temperature (T_{\min}), maximum temperature (T_{\max}) and RH in two stations of Zahedan and Bojnurd, Iran. Results indicated that the precision of SVR estimates was higher than correspondent empirical results due to statistical analysis. Mohammadi, Shamshirband, Anisi, Alam, and Petkovic (2015) used SVR methodology for estimating the horizontal GSR using climatological parameters of n and maximum sunshine hours (N) as inputs and compared the obtained results with the correspondent estimates of empirical equations. The outcomes of the research proved the superior capabilities of SVR methodology comparing empirical equations. Citakoglu (2015) applied ANN and ANFIS methodologies for estimating GSR using meteorological parameters as inputs for various stations in Turkey. The obtained results showed that the ANN had superior accuracy comparing ANFIS and empirical equations. Mousavi, Mostafavi, Jaafari, Jaafari, and Hosseinpour (2015) developed genetic programming (GP)-based approaches for estimation of daily GSR using meteorological parameters. They stated that the presented models had higher performances in comparison with the regression-based models. Mehdizadeh, Behmanesh, and Khalili (2016) examined the capabilities of gene expression programming (GEP), ANN and ANFIS methods in addition to various empirical equations for estimation of daily GSR in Kerman City, Iran during the time period of 1992–2009. The findings of the study indicated that the ANN and ANFIS models with sunshine based parameters had prevailing performances in comparison with other studied methods. Sharifi, Reza-verdinejad, and Nourani (2016) performed a comparative research among ANN, GEP, Wavelet regression (WR)

and empirical equations for estimation of daily GSR. The obtained results proved the high ranking capabilities of ANN models in GSR estimation. Vakili, Sabbagh-Yazdi, Khosrojerdi, and Kalhor (2017) presented an ANN-based model for estimating GSR using particle matter, wind speed, RH and T in Tehran, Iran. They discussed that adding particle matter to climatological parameters increased the accuracy of GSR predictions. Rao, Premalatha, and Naveen (2018) examined the effects of different combinations of meteorological parameters in GSR estimation by ANN models. Obtained results proved that the ANN model with input parameters of theoretical sunshine hours and extraterrestrial radiation had the most accuracy in GSR estimation. Kaba, Sarıgöl, Avcı, and Kandirmaz (2018) implemented a deep learning theory for GSR estimation in 34 stations of Turkey. They concluded that the considered model had suitable accuracy in comparison to many previous researches in GSR estimation. Benali, Notton, Fouilloy, Voyant, and Dizene (2019) tried to predict hourly GSR values on the site of Odeillo, France using three soft computing techniques including ANN, smart persistence and random forest. They stated that the precision of random forest models was considerably higher than other studied models in GSR prediction.

Even though GSR is one of the most frequently measured parameters worldwide, the numbers of these stations are still scanty, mainly in developing countries. So, despite the great number of studies aimed to estimate GSR in different stations worldwide, evolving flexible and accurate techniques for valid predictions is highly required due to the importance of the availability of accurate GSR data. Furthermore, regarding the insufficient researches on applications of model tree (MT) and SVR in GSR estimation and the significance of different combinations of climatological parameters in increasing the estimation accuracies, the main goal of the current study is assessing the capabilities of MT and SVR methodologies and comparing with ANFIS, GEP and empirical equations in GSR estimation. For the comprehensive investigation of the aforementioned models, the comparisons are performed based on widely used statistical meters. Additionally, a Taylor diagram was used for selecting the relatively accurate model in GSR estimation.

2. Material and methods

2.1. Study area and data collection

The monthly climatic data of automated weather station (latitude 38°05' N, longitude 46°17'E) located in Tabriz, Iran, were used in the current study (Figure 1). Tabriz, which is situated at northwestern of Iran, is a

mountainous basin with a semi-arid climate and cold winters. Meteorological variables (Table 1) that were used for this study are: minimum temperature (T_{\min}), maximum temperature (T_{\max}), relative humidity (RH), sunshine hours (n), maximum sunshine hours (N), corrected clear-sky solar irradiation (ICSKY), day of year (DOY) and extra-terrestrial radiation (R_a) and global solar radiation (GSR) with the time period of 2011–2013. Table 2 represents the daily statistical parameters of the applied meteorological variables for Tabriz Station in both training and test phases. Most of the variables indicate normal distributions because they have low skewness values, except n/N and GSR which they show negative and positive skewed distributions, respectively. R_a shows higher correlation with GSR in both training and test data. Furthermore, DOY has the lowest correlation with GSR comparing with other meteorological parameters. Overall results reveal that the R_a , T_{\max} and T_{\min} have the first, second and third highest correlations with GSR values, respectively. Moreover, there is a high inverse correlation between the RH and GSR. The minimum GSR value of training data (1.97) is higher than the corresponding value of test data (1.09) and this may cause extrapolation difficulties for the applied data-driven models in estimation low GSR values in the test period. The observed meteorological data are shown in Figure 2.

2.2. Empirical equations

Empirical equations, which implement meteorological parameters, are beneficial for the estimation of solar radiation. Angstrom (1924) and Prescott (1940) proposed Angstrom-Prescott method for predicting GSR using n/N by implementing a linear relationship. Furthermore, Swartman and Ogunlade (1967) added RH parameter for accurate estimation of GSR. They also used a linear regression-based equation for GSR estimation. Bristow and Campbell (1984) and Allen (1997) estimated GSR utilizing minimum and maximum temperatures and applying nonlinear equations. Elagib and Mansell (2000) used an exponential function implementing n/N and Chen, Ersi, Yang, Lu, and Zhao (2004) presented accurate predictions of GSR values by adding minimum and maximum temperature difference. The selected empirical equations for comparing their accuracies with data-driven techniques are illustrated in Table 3.

2.3. Support vector regression

The support vector machine (SVM) is a broadly implemented estimator which was developed by Vapnik (1995) and uses the concepts of supervised learning. Later, a regression-based technique (SVR) was presented using



Figure 1. Studied synoptic station (URL1).

Table 1. Meteorological variables which are utilized for developing the studied models.

Variables	Notation
Minimum temperature (°C)	T_{\min}
Maximum temperature (°C)	T_{\max}
Relative humidity (%)	RH
Sunshine hours (hr)	N
Maximum sunshine hours (hr)	N
Corrected clear-sky solar irradiation (-)	ICSKY
Day of year (day)	DOY
Extra-terrestrial radiation ($\text{MJ m}^{-2} \text{d}^{-1}$)	R_0
Global solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$)	GSR

the theory of SVM and structural risk minimization for solving complicated problems. In this method, a kernel function is employed for converting a nonlinear problem to a linear one. Thus, ε – the insensitive loss function is recognized which indicated that the model permits tolerating errors up to ε in the training data sets. So, the SVR look for a linear function as follows:

$$P(x) = F^T x + L \quad (1)$$

where F and L represent the coefficients of the weight vector of the linear expression. This linear regression can be

defined as the following:

$$\min \frac{1}{2} \|F\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)$$

$$\text{subject to } \begin{cases} F^T x + L - y_i \leq \varepsilon + \xi_i^* \\ y_i - F^T x - L \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \quad i = 1, 2, \dots, N \end{cases} \quad (2)$$

C is a predefined constant trade-off factor for the grade of the experimental error. Additionally, kernel functions are one of the vital factors of SVR for accurate solving of complex issues (Smola & Schölkopf, 2004). In the current study, four different kernel functions, namely polynomial, normalized polynomial, Pearson VII function-based and radial basis function were utilized. Figure 3 specifies schematic configuration of SVR model.

2.4. Gene expression programming

Gene expression programming (GEP) implements computer programs for solving complicated problems (Ferreira, 2001a, 2001b). Also, the main distinction between GEP and its former version such as GP and genetic

Table 2. Statistical analysis of meteorological parameters.

	Variable	X_{mean}	X_{min}	X_{max}	S_x	C_V	C_{sx}	Correlation with GSR with GSR
Train	T_{min}	7.42	-13.20	28.20	9.94	1.34	-0.03	0.74
	T_{max}	19.46	-3.20	40.00	11.51	0.59	-0.08	0.81
	RH	50.10	15.00	93.75	17.73	0.35	0.27	-0.71
	n/N	0.65	0.00	0.96	0.28	0.43	-1.04	0.60
	ICSKY	4.92	0.00	9.00	2.50	0.00	-0.44	-0.49
	DOY	182.64	1.00	365.00	110	0.00	0.00	-0.01
	R_a	28.97	14.71	41.82	9.66	0.33	-0.11	0.85
	GSR	18.46	1.97	91.52	9.36	0.51	1.20	1.00
Test	T_{min}	8.90	-10.40	25.00	8.87	1.00	-0.08	0.70
	T_{max}	20.35	-2.20	38.80	10.48	0.52	-0.17	0.80
	RH	51.15	15.00	89.63	17.20	0.34	0.15	-0.76
	n/N	0.64	0.00	0.95	0.29	0.45	-1.07	0.59
	ICSKY	5.21	0.00	9.00	2.373	0.50	-0.58	-0.48
	DOY	183.00	1.00	365.00	105.51	0.58	0.00	-0.08
	R_a	28.93	14.71	41.82	9.67	0.33	-0.10	0.89
	GSR	18.30	1.09	86.59	9.08	0.50	1.09	1.00

Note: The terms X_{mean} , X_{min} , X_{max} , S_x , C_V and C_{sx} denote the mean, minimum, maximum, standard deviation, coefficient of variation and skewness, respectively.

algorithm (GA) is due to the differences in the nature of individuals (Mesbah, Soroush, & Rostampour Kakroudi, 2017). GEP algorithm begins by generating preliminary population through the random formation of chromosomes with the fixed length. Then, every chromosome is examined by a fitness function in order to modify the reproduction of the future generation. Various fitness functions can be contemplated for a GEP model. Owing to the fact the RRSE fitness function has been applied frequently in prior studies (e.g. Emamgolizadeh, Bateni, Shahsavani, Ashrafi, & Ghorbani, 2015), it was preferred herein. The mentioned process will be continued for a predefined number of generations or until an appropriate solution is established (Ferreira, 2006). Simple drawing of GEP algorithm can be comprehended from Figure 4.

2.5. Model trees

Recently, there has been an increasing trend in using decision trees in modeling complex nonlinear problems. M5 model tree (MT) as one of the commonly used conventional decision tree algorithms was introduced by Quinlan (1992) and has been created based on divide and conquer methodology and can be applied for prediction of several variables (Singh, Sachdeva, & Pal, 2016). MT has a high proficiency in handling a large number of attributes and is a strong technique in predictive purposes (Solomatine & Dulal, 2003). Moreover, MT is a mathematical predictive algorithm which its nodes of the trees are selected over the attributes that minimize the expected error as a function of the standard deviation of target parameter (Zhang & Tsai, 2007). MT implements three steps in its process including breaking the input space, establishing the tree and finally obtaining the

knowledge of it (Behnood, Olek, & Glinicki, 2015). The mentioned process is presented schematically in Figure 5.

Primarily, the input data is separated into different regions, as shown in Figure 5(a). The mentioned process is based on implementing linear regression and reducing the standard deviation of the errors among measured and predicted values. Then, using the splitted space from the first step, a decision tree, with leaf at the top and nodes at the bottom, is generated (typically presented in Figure 5(b)). The last step includes predicting target value using obtained the linear regression equations. At this step, input data is presented to the root and start passing its way to the nodes by evaluating the splitting parameters (X_i) (see Figure 5(c)). The mentioned criteria for generating linear regression equation, namely standard deviation reduction (SDR) can be computed as follows:

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

where T symbolizes a set of examples which reaches the nodes, T_i indicates the subset of examples and sd represents the standard deviation. Considering the splitting procedure, the data in child nodes have lower standard deviation in comparison with the parent node and are purer. After investigating the whole potential splits, MT selects the one that maximizes SDR. The mentioned procedure usually creates a large tree that may cause overfitting. For solving this issue, the tree must be pruned back, for instance by changing a sub tree with a leaf. Hence, the next stage contains pruning the overgrown tree and substituting the sub trees with linear regression functions. For additional detailed information, readers are recommended to study Quinlan (1992).

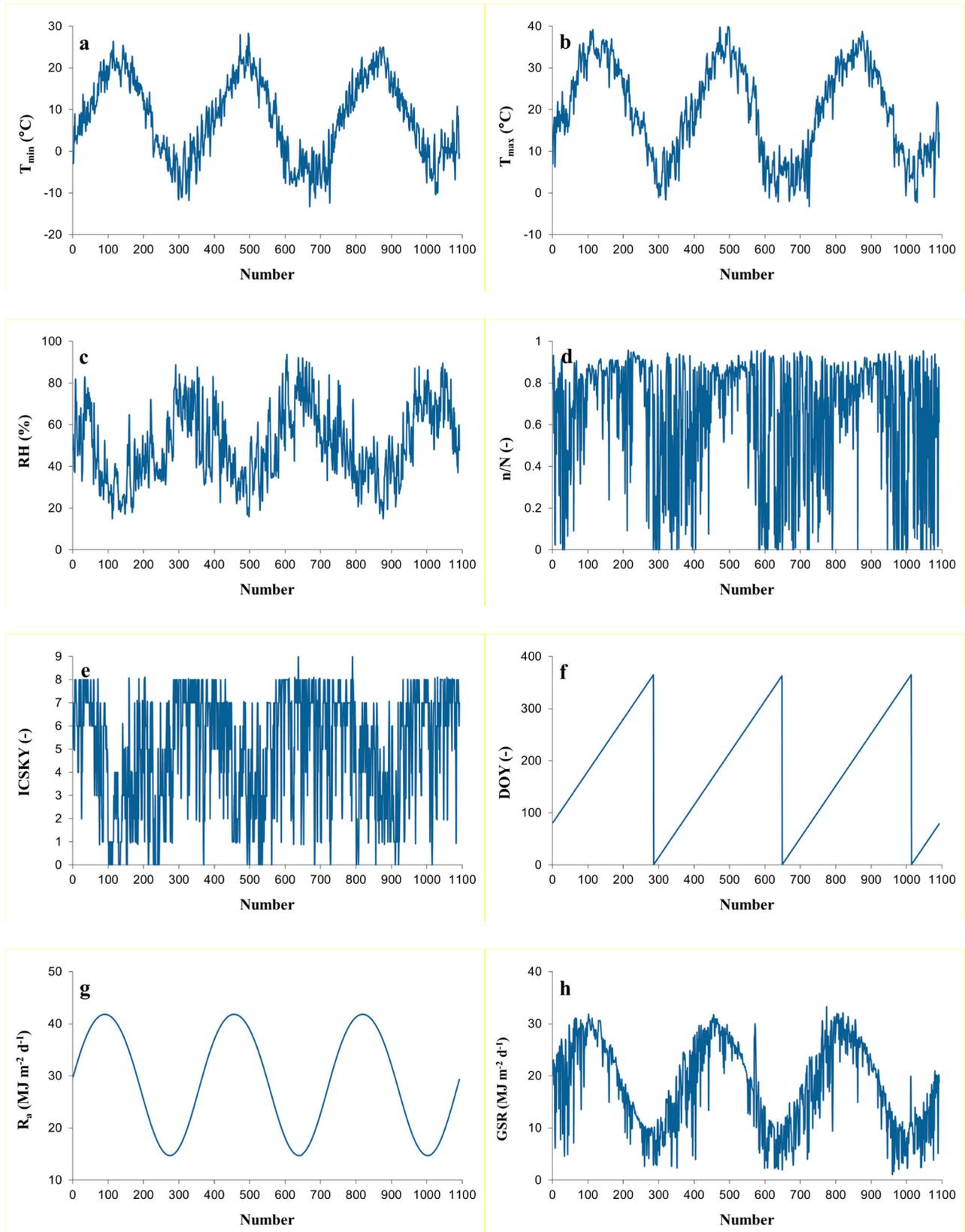


Figure 2. Time variation graphs of the meteorological parameters.

Table 3. Empirical equations employed in the current study.

Method	Notation	Equation	Source
Angstrom and Prescott	A-P	$R_s = R_a(a + b \times (n/N))$	Prescott (1940)
Swartman and Ogunlade	S-O	$R_s = a + b \times (n/N) + c \times RH$	Swartman and Ogunlade (1967)
Bristow and Campbell	B-C	$R_s = R_a \times a \times (1 - \exp(-b \times (T_{max} - T_{min})^c))$	Bristow and Campbell (1984)
Allen	A	$R_s = R_a \times a \times \sqrt{(T_{max} - T_{min})}$	Allen (1997)
Elagib and Mansell	E-M	$R_s = R_a \times a \times \exp(b \times (n/N))$	Elagib and Mansell (2000)
Chen, Ersi, Yang, Lu, and Zhao	C	$R_s = R_a \times (a \times \ln(T_{max} - T_{min}) + b \times (n/N)^c + d)$	Chen et al. (2004)

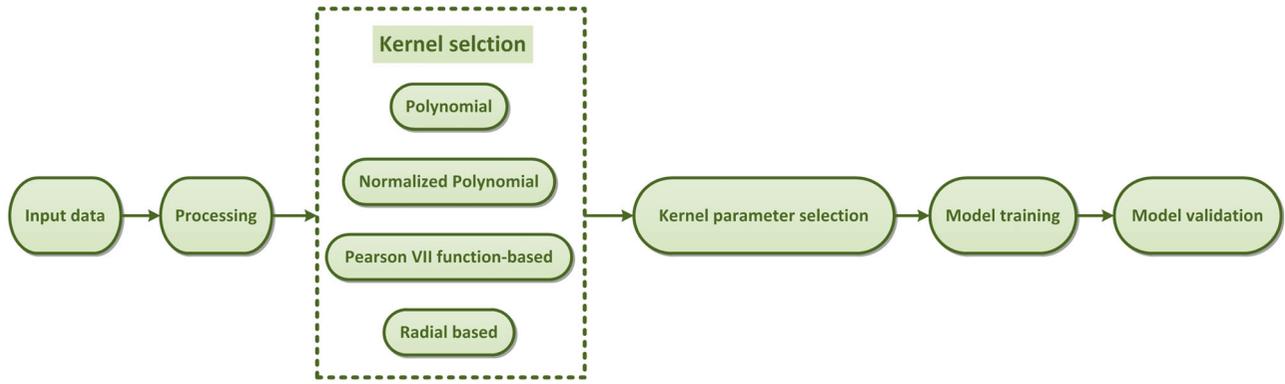


Figure 3. Schematic configuration of SVR model.

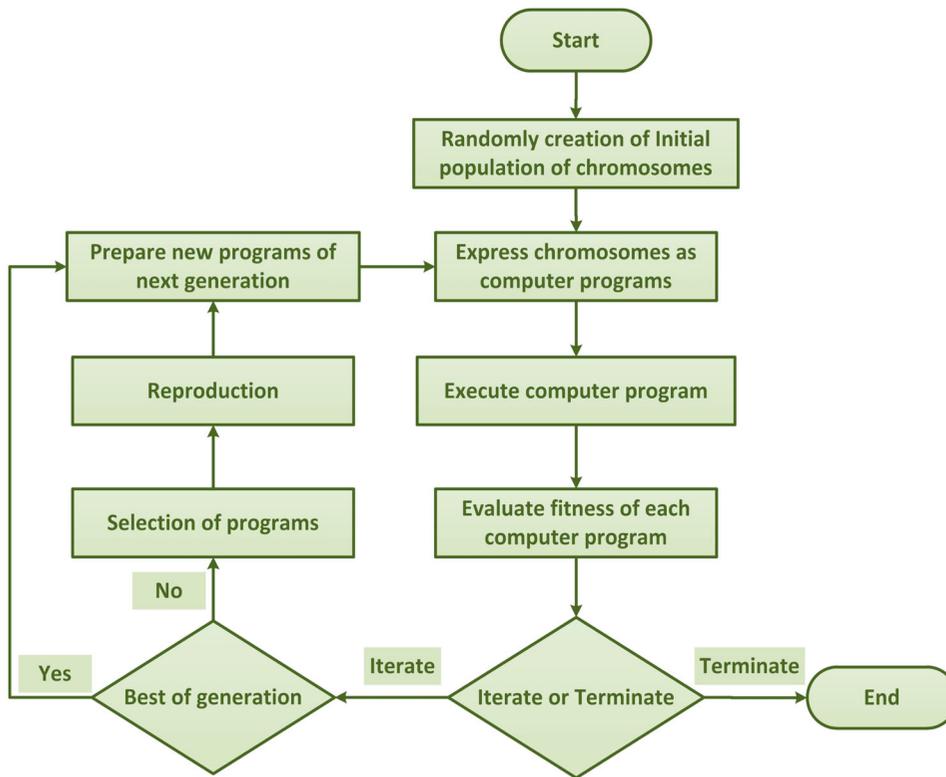


Figure 4. A GEP flowchart.

2.6. Adaptive neuro-fuzzy inference system

ANFIS, as one of the artificial intelligence methods, is based on Takagi-Sugeno fuzzy inference system (Saini & Kumar, 2016). This method has been broadly used in solving complicated engineering problems (for instance,

Olatomiwa, Mekhilef, Shamshirband, & Petković, 2015; Piri & Kisi, 2015). ANFIS includes fuzzy logic operations based on the theory of membership function and uses some if-then rules. ANFIS is consisted of three key parts including a fuzzy rule base, data and a suitable procedure.

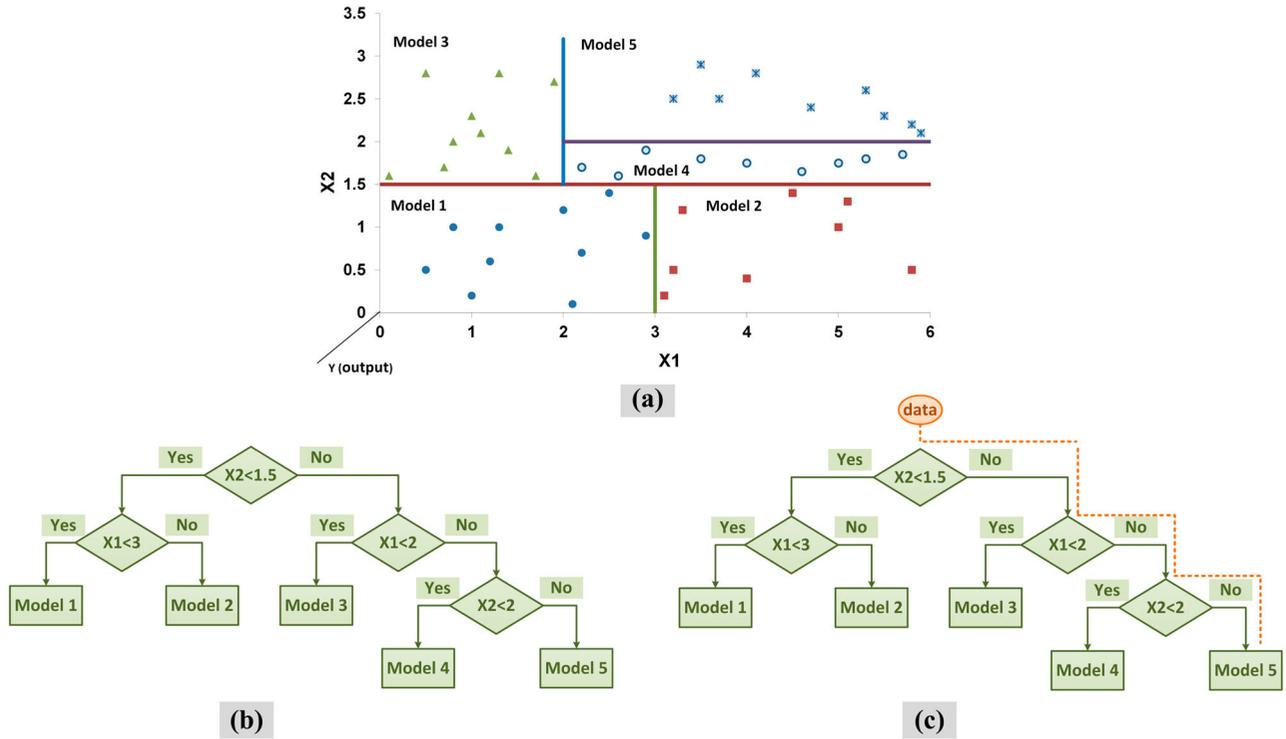


Figure 5. Breaking the input space by MT and predicting a target value for a new dataset. (a) Breaking the data (X1 and X2), (b) establishing the tree and (c) predicting new data by established MT.

The membership functions are defined by data and the inference process would be performed by the procedure (Akbarpour, Mohajeri, & Akbarpour, 2016). Moreover, ANFIS employs a first order Sugeno-type fuzzy inference system, which has two inputs and one output defined as follows:

Rule 1: if x is A_1 and y is B_1 , then $f_1 = p_1x + q_1y + r_1$

Rule 2: if x is A_2 and y is B_2 , then $f_2 = p_2x + q_2y + r_2$

Where x and y are input characteristics, A_1, A_2, B_1, B_2 are nonlinear functions' coefficients and p_1, q_1, p_2, q_2 are

linear functions' coefficients and $f(x,y)$ is first order polynomial (Ertunc, Ocak, & Aliustaoglu, 2013). A common ANFIS model is consisted of five different layers including fuzzification, multiplication, normalization, defuzzification and summation. A typical diagram of ANFIS is presented in Figure 6. For additional information about the details of mentioned layers, readers are referred to Akbarpour et al. (2016).

2.7. Evaluation parameters

For inclusive justification of the performance of the studied data-driven techniques and empirical equations,

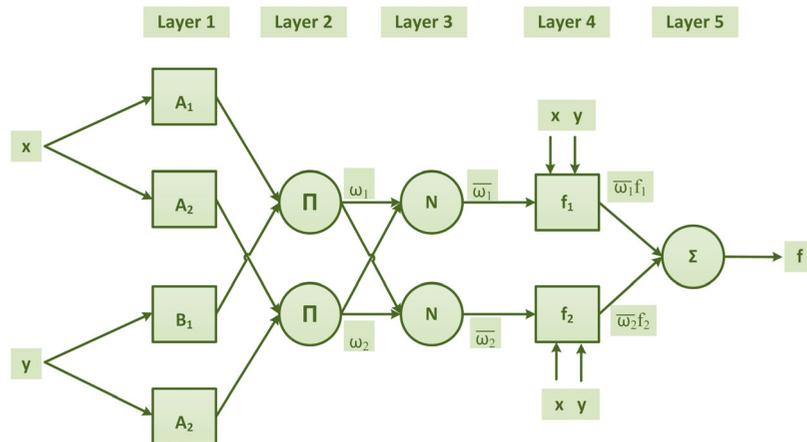


Figure 6. Typical structures of ANFIS for two inputs with four rules.

graphic plots and some statistical parameters were implemented as follows:

I: Correlation coefficient (CC) (Samadianfard et al., 2018), expressed as

$$CC = \frac{\left(\sum_{i=1}^n O_i P_i - \frac{1}{n} \sum_{i=1}^n O_i \sum_{i=1}^n P_i\right)}{\sqrt{\left(\sum_{i=1}^n O_i^2 - \frac{1}{n} \left(\sum_{i=1}^n O_i\right)^2\right) \left(\sum_{i=1}^n P_i^2 - \frac{1}{n} \left(\sum_{i=1}^n P_i\right)^2\right)}} \quad (4)$$

II: Root mean square error (RMSE) (Willmott & Matsuura, 2005) follows as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (5)$$

III: Mean absolute error (MAE) (Chai & Draxler, 2014) stated as

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (6)$$

IV: Willmott's Index of agreement (WI) (Willmott, Robeson, & Matsuura, 2012) expressed as

$$WI = 1 - \left[\frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}_{-i}| + |O_i - \bar{O}_{-i}|)^2} \right] \quad (7)$$

Table 4. Input parameters for implementation of studied models and empirical equations.

Number of input combinations	Input parameters
1	R_a
2	$R_a, n/N$
3	R_a, RH
4	R_a, T_{min}, T_{max}
5	$R_a, RH, T_{min}, T_{max}$
6	$R_a, RH, T_{min}, T_{max}, n/N$
7	$R_a, RH, T_{min}, T_{max}, n/N, DOY$
8	$R_a, RH, T_{min}, T_{max}, n/N, ICSKY$
9	$R_a, RH, T_{min}, T_{max}, n/N, DOY, ICSKY$

where O_i and P_i are the observed and predicted i^{th} value of the GSR.

V: Taylor Diagrams (Taylor, 2001).

Also, Taylor diagram has been exploited to establish its implication in hydrological modeling for assessing the precision of the predicted data. It is a solo diagram which simultaneously combines statistical parameters such as standard deviation, RMSE and CC values. Impressively, Taylor diagrams are capable of highlighting the preciseness of different models in comparison with observations with a series of points on a polar plot (IPCC, 2007; Taylor, 2001).

3. Results and discussion

The capabilities of four data-driven techniques including SVR, MT, GEP and ANFIS in predicting daily GSR in Tabriz station, Iran were compared with the performances of six empirical equations, namely A-P, E-M, A, B-C, S-O and C. In the present study, 70% of the whole dataset was utilized to train and the 30% of was used to test the mentioned models. In other words, daily data from the beginning of 2011 to the end of 2012 were selected for training and the measured data of 2013 were implemented for testing.

Based on meteorological parameters defined in Table 1, nine various input combinations were contemplated in the current computations (Table 4). Furthermore, it is clear from Table 2 that R_a and temperature values have the most direct correlations with GSR, while RH has the higher inverse correlation with GSR. Hence, R_a, T_{min}, T_{max} and RH parameters were considered in most input combinations to increase the precision of the models. As shown in Table 4, nine different input combinations tried in the study are (1) R_a , (2) $R_a, n/N$, (3) R_a, RH , (4) R_a, T_{min}, T_{max} , (5) $R_a, RH, T_{min}, T_{max}$, (6) R_a, RH, T_{min} ,

Table 5. Statistical parameters of the various scenarios of data-driven techniques in the test period.

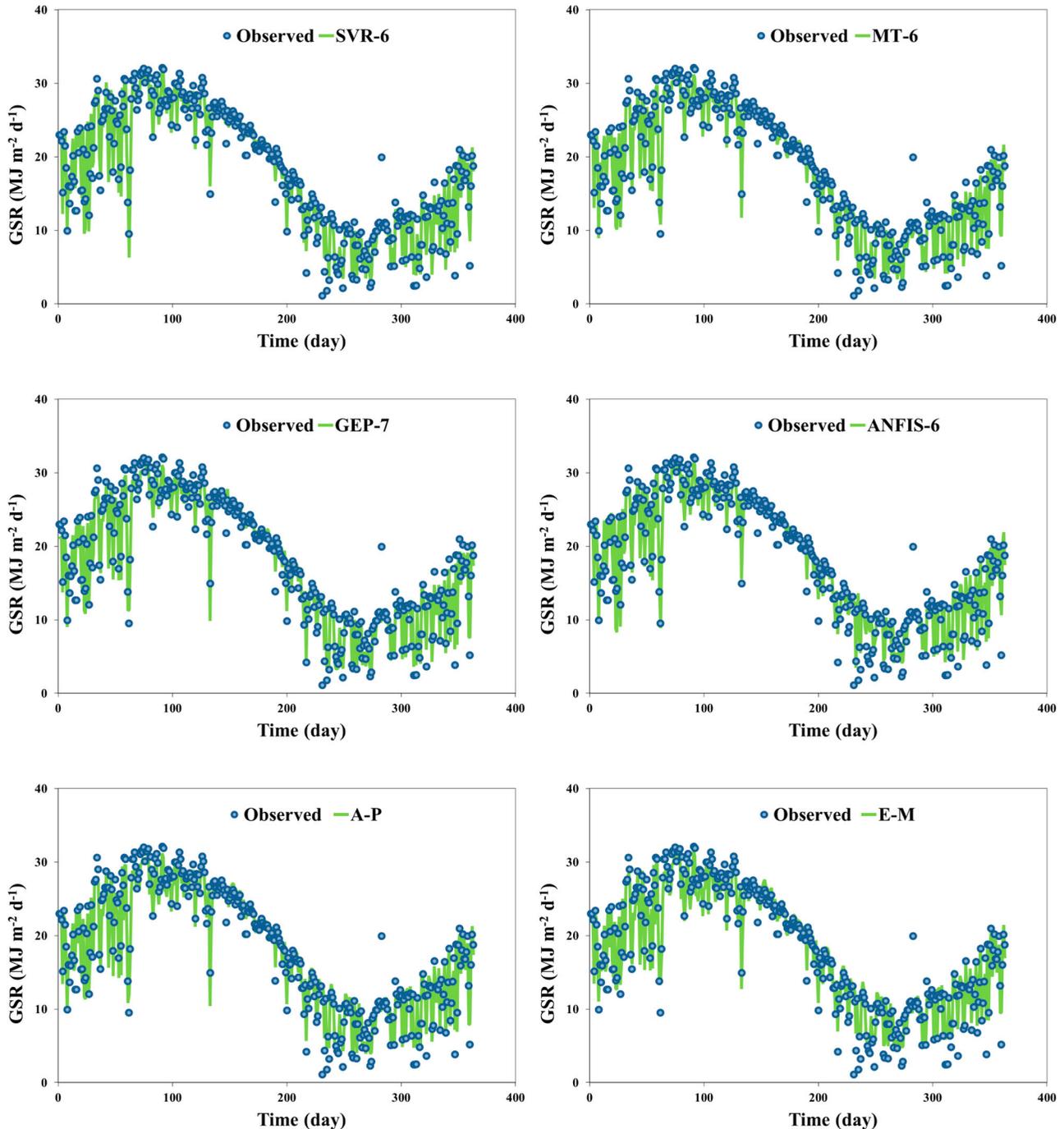
Method	Parameter	Sen 1	Sen 2	Sen 3	Sen 4	Sen 5	Sen 6	Sen 7	Sen 8	Sen 9
SVR	RMSE ($MJ m^{-2} d^{-1}$)	4.024	1.789	3.393	3.031	2.889	1.656	1.668	1.761	1.746
	MAE ($MJ m^{-2} d^{-1}$)	2.725	1.033	2.341	2.138	1.962	0.990	1.027	1.076	1.091
	CC	0.891	0.976	0.917	0.932	0.939	0.980	0.980	0.977	0.977
	WI	0.931	0.988	0.955	0.965	0.968	0.990	0.990	0.988	0.989
MT	RMSE ($MJ m^{-2} d^{-1}$)	3.744	1.743	3.232	3.016	2.929	1.672	1.691	1.674	1.693
	MAE ($MJ m^{-2} d^{-1}$)	2.905	1.070	2.388	2.206	2.166	1.008	1.017	1.008	1.017
	CC	0.891	0.977	0.920	0.933	0.938	0.979	0.979	0.979	0.979
	WI	0.938	0.988	0.958	0.965	0.968	0.989	0.989	0.989	0.989
GEP	RMSE ($MJ m^{-2} d^{-1}$)	3.789	1.851	3.355	3.050	3.093	1.822	1.681	1.756	1.871
	MAE ($MJ m^{-2} d^{-1}$)	2.945	1.095	2.594	2.313	2.346	1.064	1.064	0.125	1.099
	CC	0.889	0.975	0.914	0.932	0.928	0.975	0.979	0.977	0.974
	WI	0.937	0.987	0.954	0.964	0.962	0.988	0.989	0.988	0.987
ANFIS	RMSE ($MJ m^{-2} d^{-1}$)	3.730	1.906	3.373	3.232	2.823	1.937	2.073	1.986	2.331
	MAE ($MJ m^{-2} d^{-1}$)	2.894	1.250	2.512	2.383	2.019	1.171	1.435	1.447	1.618
	CC	0.892	0.973	0.914	0.920	0.940	0.972	0.969	0.971	0.960
	WI	0.939	0.986	0.955	0.957	0.969	0.985	0.983	0.985	0.979

Table 6. Statistical parameters of the considered empirical equations in the test period.

Parameter	A-P	E-M	A	B-C	S-O	C
RMSE ($\text{MJ m}^{-2} \text{d}^{-1}$)	1.786	1.869	3.008	2.999	6.620	2.481
MAE ($\text{MJ m}^{-2} \text{d}^{-1}$)	1.156	1.254	2.247	2.231	5.659	1.788
CC	0.977	0.975	0.932	0.932	0.597	0.955
WI	0.988	0.986	0.962	0.963	0.728	0.975

T_{\max} , n/N, (7) R_a , RH, T_{\min} , T_{\max} , n/N, DOY, (8) R_a , RH, T_{\min} , T_{\max} , n/N, ICSKY and (9) R_a , RH, T_{\min} , T_{\max} , n/N, DOY, ICSKY.

The results of statistical indicators such as CC, RMSE, MAE and WI for predicting daily GSR at Tabriz station for studied models and considered empirical equations are presented in Tables 5 and 6. The overall investigations proved that the precision of predictions increased from input combination 1–6 in all data-driven techniques. As mentioned earlier, four kernel functions including polynomial, normalized polynomial, Pearson VII function-based and radial basis function were examined for all SVR models. It was found that, except for SVR-1 which

**Figure 7.** Time series plot of observed and estimated GSR in the test period.

polynomial kernel function produced accurate results, Pearson VII function-based provided a precise prediction for all SVR models comparing other kernel functions. Among SVR models, SVR-6 which utilized input combinations of R_a , RH, T_{\min} , T_{\max} and n/N with RMSE of 1.656, MAE of 0.990, CC of 0.980 and WI of 0.990 had the best performances comparing with other SVR models and showed higher accuracy. Additionally, SVR-7 ranked as the second best SVR model. This indicates that DOY did not have a positive effect in decreasing the prediction accuracy. Similar to SVR models, the same trend was approximately seen between MT models. Comparing the results of various MT models revealed that the MT-6 which used input combinations of R_a , RH, T_{\min} , T_{\max} and n/N with RMSE of 1.672, MAE of 1.008, CC of 0.979 and WI of 0.989 had the highest accuracy among the MT models. However, MT increased the RMSE and MAE accuracy of the best SVR model (SVR-6) by 0.97% and 1.82%, respectively. Unlike SVR and MT models, input combination of seven resulted in accurate prediction comparing with other GEP models. Additionally, the analysis of GEP error meters proved that the GEP-7 by input combinations of R_a , RH, T_{\min} , T_{\max} , n/N and DOY with RMSE of 1.681, MAE of 1.064, CC of 0.979 and WI of 0.989 provided relatively more precise predictions in

comparison with other GEP models. Nevertheless, GEP increased the RMSE accuracy of SVR-6 and MT-6 by 1.51%, 0.54% and MAE by 7.47%, 5.56%, respectively. Finally, the results of the ANFIS models clearly showed that, similar to SRV and MT models, ANFIS-6 which implements input combinations of R_a , RH, T_{\min} , T_{\max} and n/N with RMSE of 1.937, MAE of 1.171, CC of 0.972 and WI of 0.985 provided the best prediction among the ANFIS models. However, the precision of all ANFIS models is lower than those of the SVR, MT and GEP models.

Moreover, investigating the statistical parameters of empirical equations from Table 6 shows that the commonly used A-P model which employs R_a and n/N with RMSE of 1.786, MAE of 1.156, CC of 0.977 and WI of 0.988 presents more accurate predictions in comparison with other empirical equations. Although A-P was selected as the best empirical equation in the current study, but its RMSE and MAE accuracies are increased by 16.97% and 18.28% using the best data-driven model (SVR-6), respectively. Also, E-M with the same input parameters of A-P, was selected as the second best empirical equation with RMSE of 1.869, MAE of 1.254, CC of 0.975 and WI of 0.986. So, the investigating the error meters of A-P and E-M clearly proves that the linear

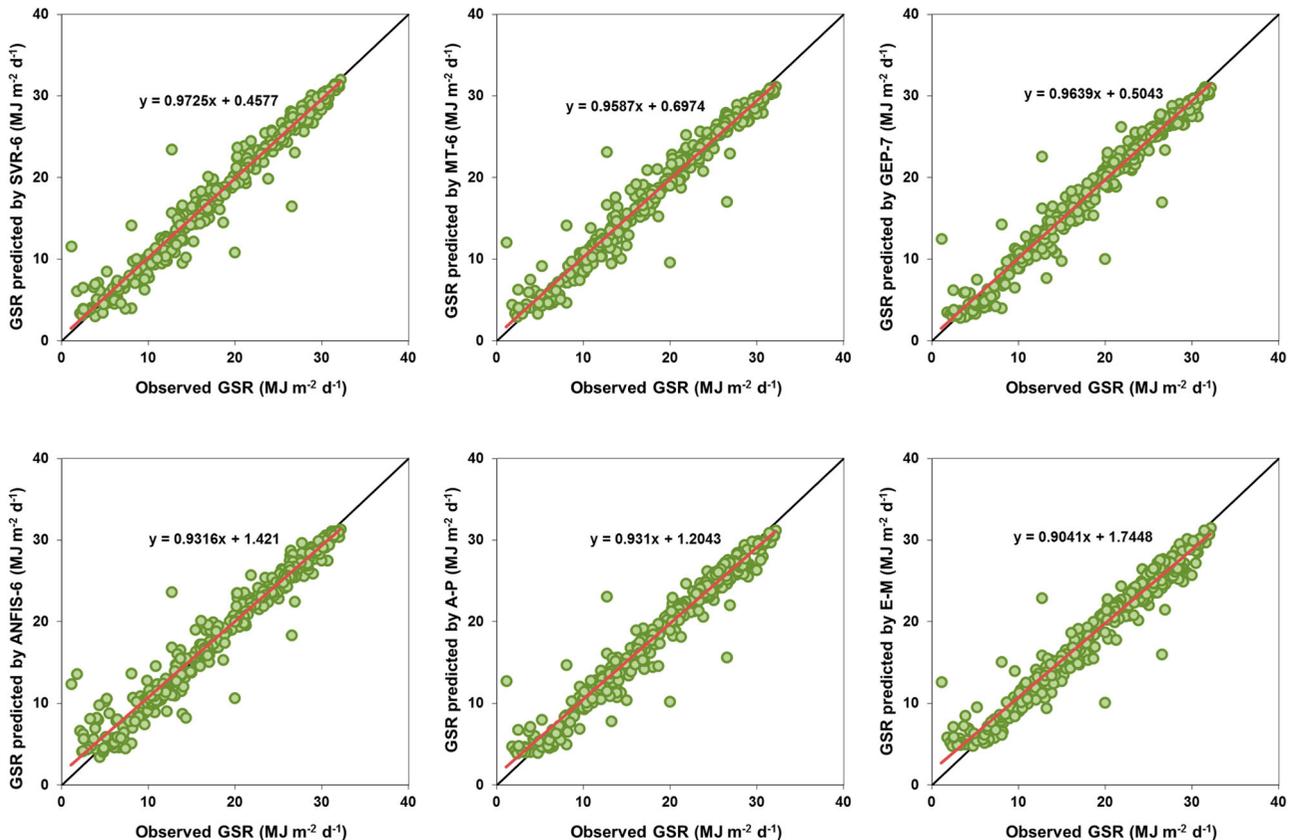


Figure 8. Scatterplots of observed and predicted GSR in the test period.

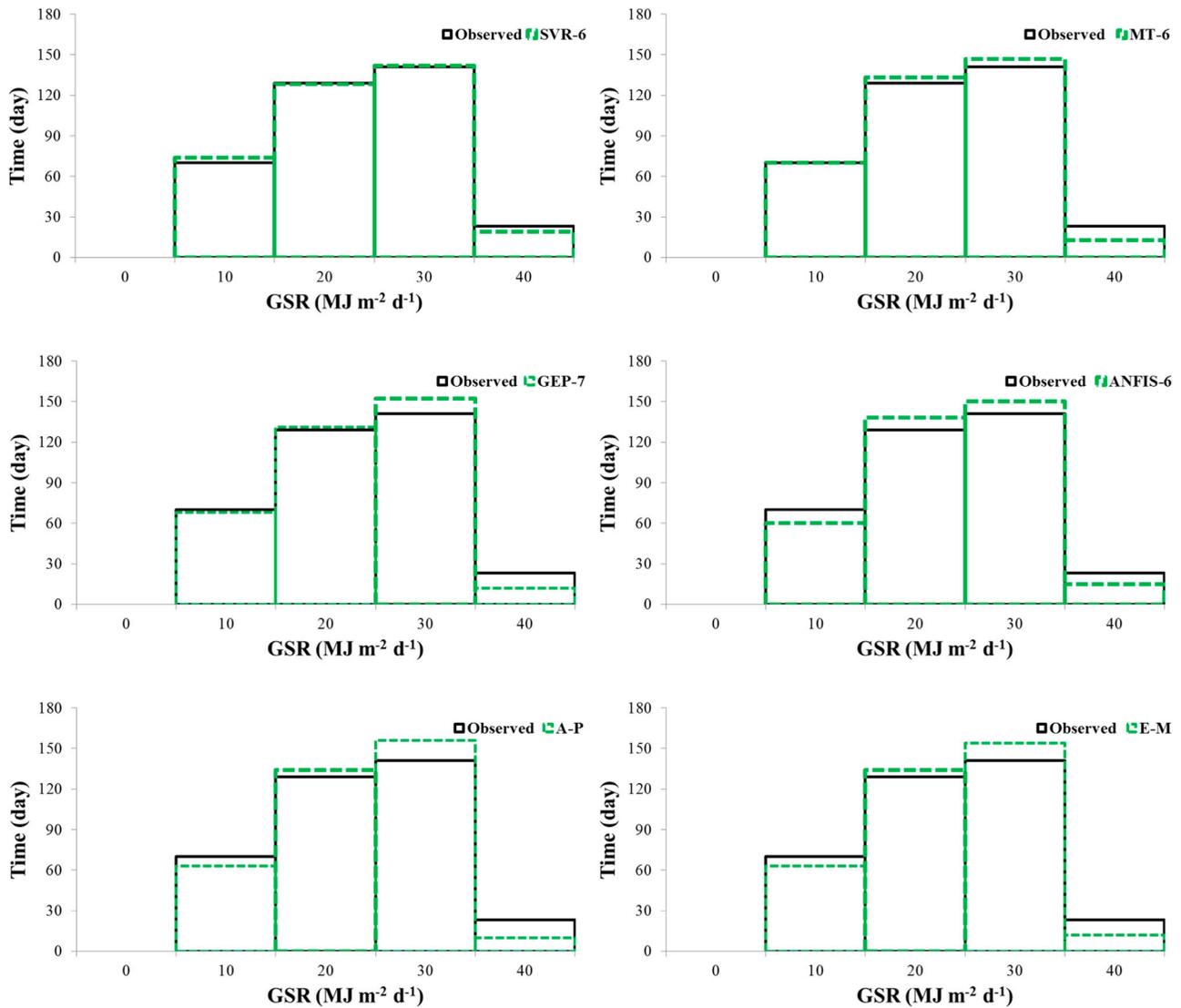


Figure 9. Histograms of GSR in the test period.

relationship between GSR and input parameters of R_a and n/N (existed in A-P equation) provides more accurate predictions in comparison with exponential relationship (existed in E-M equation).

Figure 7 illustrates the observed and predicted values of GSR in the test period using the best models of each data-driven techniques including SVR-6, MT-6, GEP-7 and ANFIS-6 and two best empirical equations including A-P and E-M. It can be comprehended that the predictions of SVR-6 are in better agreement with observed GSR values comparing other models. Furthermore, Figure 8 presents the observed and predicted GSR for the selected best models in the test period in the form of scatterplot. Also, it can be seen from Figure 8 that the estimates of SVR-6 and MT-6 are less scattered through the exact lines. In other words, the estimates of SVR-6 and MT-6 are closer to the exact lines than other studied models.

As mentioned in the literature review, Ramedani et al. (2014) concluded that SVR with radial basis function with RMSE of 3.3 predicted GSR with acceptable accuracy in Tehran station, Iran. So, comparing the obtained results from the current study with the findings of Ramedani et al. (2014) indicates that the accuracy of SVR-6 is significantly higher than those implemented by Ramedani et al. (2014).

An extra investigation of observed and predicted values by SVR-6, MT-6, GEP-7, ANFIS-6, A-P and E-M models is also performed. Figure 9 illustrates the probability distribution of the testing data. These figures are beneficial for understanding the capabilities of different studied models in GSR estimation. It is apparent from Figure 9 that probability distributions of the predicted GSR by SVR-6 are closer to the observed one for most intervals presented in the test period.

For exhaustive examination of attained results using the best models of different data-driven techniques and two best empirical equations, including SVR-6, MT-5, GEP-7, ANFIS-6, A-P and E-M, a Taylor diagram has been applied for evaluation of the statistical parameters among observed and predicted GSR using mentioned models. Figure 10 presents the declared diagram for the best models by which the distance from the reference point is a measurement of the centered RMSE (Taylor, 2001). Therefore, the best model is identified by the point with higher CC and the lower RMSE (Heo et al., 2014). As it is obviously clear in Figure 10, the SVR-6 generated accurate results which were much closer to

the observation in comparison with other studied models. MT-6 and GEP-7 are also better than the empirical models according to the Taylor diagram.

The founded results of Taylor diagram verified that the SVR-6 which utilized input combinations of R_a , RH, T_{min} , T_{max} and n/N with RMSE of 1.656, MAE of 0.990, CC of 0.980 and WI of 0.990 had the best performances comparing other studied models.

Moreover, one of the key features of the GEP is its proficiency for providing mathematical expression among input and output parameters. In other words, GEP model is capable of finding an explicit mathematical function relating the input and output parameters. The resulted

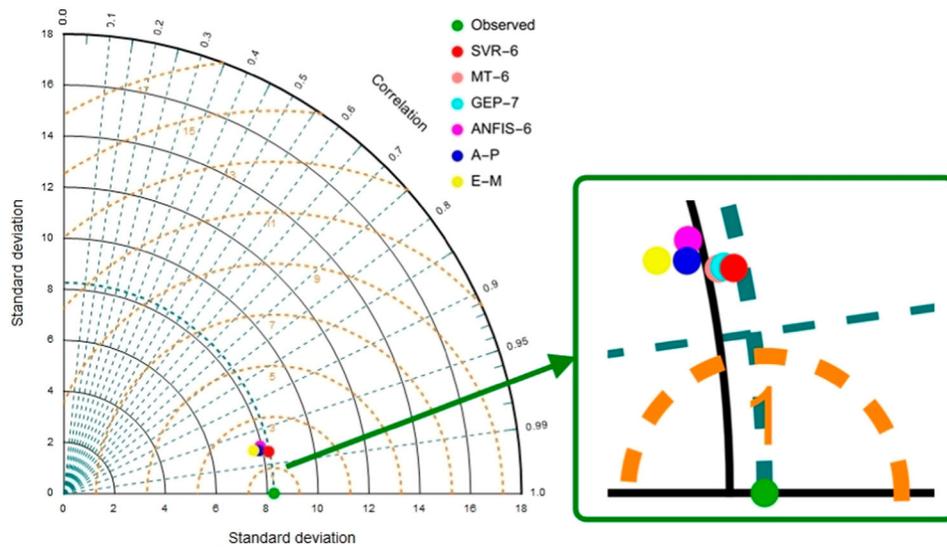


Figure 10. Taylor diagram of the predicted GSR in test period.

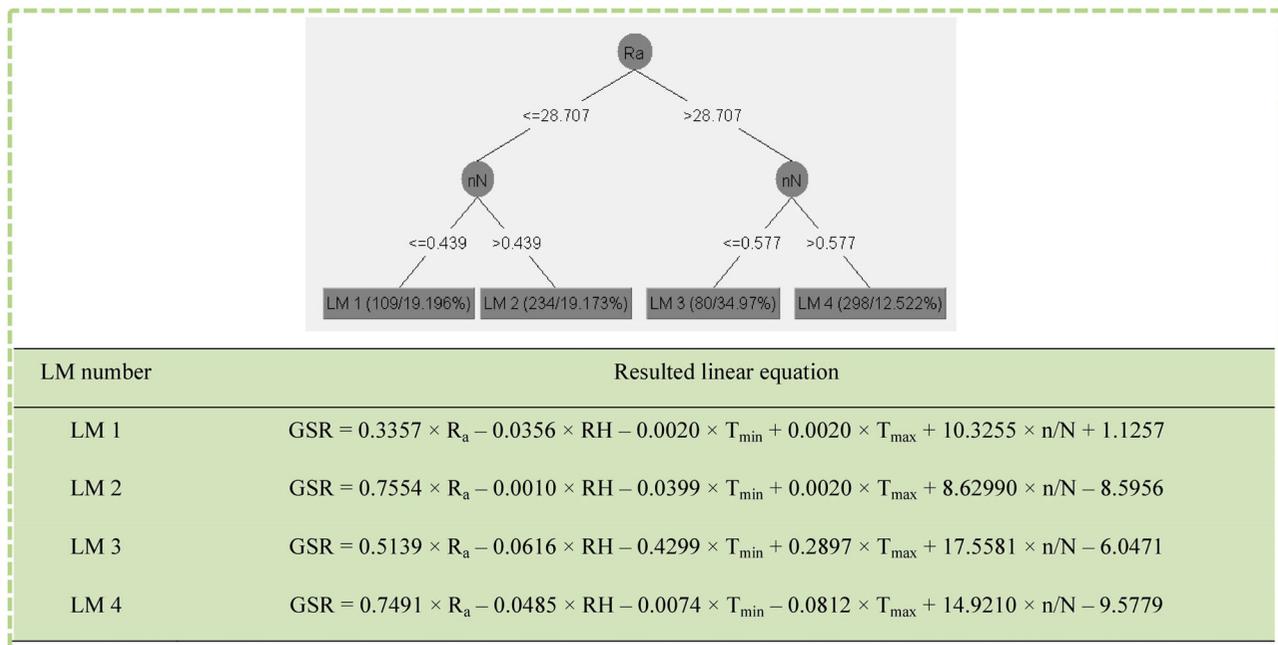


Figure 11. MT-6 resulted tree, which has 4 different linear rules.

equation of GEP-7, as the best GEP model in predicting daily GSR, is as follows:

$$\begin{aligned} \text{GSR} = & -1.00888 - 0.763834 \times \exp(-n/N) \\ & \times (R_a - 1.00888) + R_a - \frac{1.408 + RH - R_a}{RH} \\ & + \arcsin(n/N) + \tan(\cos(\sin(0.123332 \times T_{\max}))) \end{aligned} \quad (8)$$

It should be noted that Equation (8) is dimensionless. However, the dimensions of used parameters are as mentioned in Table 1. In the final equation of GEP-7 (Equation (8)), not all parameters of input combination 7 are certainly presented. This can be the result of the automatic procedure of GEP for selecting more effective input parameters in modeling GSR values.

Also, the trees structure acquired by MT-6, as an accurate model of MT, is illustrated in Figure 11. It should be noted that the most optimum and also simple trees, which gave acceptable and accurate results, were selected as the best MT models. This pictured tree model is only based on the values of R_a and n/N . As a result of MT classifier, 4 rules were produced which are shown in Figure 10. These simple linear equations can straightforwardly be implemented to predict daily GSR at Tabriz station.

4. Conclusion

As mentioned before, direct measurement of solar radiation is considered a difficult task, especially in developing countries. So, in the current research, the performances of four data-driven techniques including SVR, MT, GEP and ANFIS and six empirical equations in predicting GSR at Tabriz station, Iran have been examined. For that purpose, time series of daily GSR and several meteorological parameters such as R_a , RH, T_{\min} , T_{\max} , n/N , DOY, ICSKY in the time period of 2011–2013 were gathered and 70% of them was selected as train data and 30% for test data. Then, nine various combinations of input parameters have been considered for training data-driven techniques and finding the best combination of them to precise predications of GSR. Furthermore, the capabilities of the studied models and empirical equations were extensively inspected using three statistical parameters including RMSE, MAE and CC. Also, Taylor diagram was utilized for fully assessment of the studied models. The obtained results revealed that the SVR-6, which implements input combination of R_a , RH, T_{\min} , T_{\max} and n/N with RMSE of 1.656, MAE of 0.990, CC of 0.980 and WI of 0.990 had the best performances comparing with other data-driven techniques and empirical equations. Besides, graphical Taylor diagram proved the

effectiveness of SVR-6 in GSR modeling. Additionally, it was found that only A-P and E-M empirical equations had the acceptable accuracy and other empirical equations did not provide accurate predictions of GSR. Furthermore, it was suggested that MT-6 with four simple and explicit linear equations can also be applied beneficially for precise predictions of GSR. As a conclusion, SVR and MT models proved to be accurate models in predicting GSR in Tabriz station and could be used in practice for solar radiation estimation. Future works can focus on examining the applicability of different data-driven techniques on increasing the accuracy of solar radiation prediction.

Acknowledgments

This paper is published as part of a research project supported by the University of Tabriz Research Affairs Office.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Shahaboddin Shamshirband  <http://orcid.org/0000-0002-6605-498X>

References

- Akbarpour, H., Mohajeri, M., & Akbarpour, M. (2016). Pore diameter of nanoporous anodic alumina: Experimental study and application of ANFIS and MLR. *Chemometrics and Intelligent Laboratory Systems*, 153, 82–91.
- Allen, R. (1997). Self calibrating method for estimating solar radiation from air temperature. *Journal of Hydrologic Engineering*, 2, 56–67.
- Angstrom, A. (1924). Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Quarterly Journal of the Royal Meteorological Society*, 50, 121–126.
- Behnood, A., Olek, J., & Glinicki, M. A. (2015). Predicting modulus elasticity of recycled aggregate concrete using M5' model tree algorithm. *Construction and Building Materials*, 94, 137–147.
- Benali, L., Notton, G., Fouilloy, A., Voyant, C., & Dizene, R. (2019). Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renewable Energy*, 132, 871–884.
- Benghanem, M., Mellit, A., & Alamri, S. N. (2009). ANN-based modelling and estimation of daily global solar radiation data: A case study. *Energy Conversion and Management*, 50, 1644–1655.
- Bristow, K. L., & Campbell, G. S. (1984). On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology*, 31, 159–166.

- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7, 1247–1250.
- Chau, K. W. (2017). Use of meta-heuristic techniques in rainfall-runoff modelling. *Water*, 9(3), 1–6.
- Chen, R. S., Ersi, K., Yang, J. P., Lu, S. H., & Zhao, W. Z. (2004). Validation of five global radiation models with measured daily data in China. *Energy Conversion and Management*, 45, 1759–1769.
- Citakoglu, H. (2015). Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation. *Computers and Electronics in Agriculture*, 118, 28–37.
- Elagib, N., & Mansell, M. G. (2000). New approaches for estimating global solar radiation across Sudan. *Energy Conversion and Management*, 41, 419–434.
- Emamgolizadeh, S., Bateni, S. M., Shahsavani, D., Ashrafi, T., & Ghorbani, H. (2015). Estimation of soil cation exchange capacity using Genetic Expression Programming (GEP) and Multivariate Adaptive Regression Splines (MARS). *Journal of Hydrology*, 529, 1590–1600.
- Ertunc, H. M., Ocak, H., & Aliustaoglu, C. (2013). ANN-and ANFIS-based multi-staged decision algorithm for the detection and diagnosis of bearing faults. *Neural Computing and Applications*, 22(1), 1–12.
- Feng, H., Schiavon, S., & Bauman, F. (2016). New method for the design of radiant floor cooling systems with solar radiation. *Energy and Buildings*, 125, 9–18.
- 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 Systems*, 13(2), 87.
- Ferreira, C. (2006). *Gene expression programming: Mathematical modeling by an artificial intelligence*. Berlin: Springer, p. 478.
- Fotovatikhah, F., Herrera, M., Shamshirband, S., Chau, K. W., Aizollahzadeh Ardabili, A., & Piran, J. (2018). Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work. *Engineering Applications of Computational Fluid Mechanics*, 12(1), 411–437.
- Gairaa, K., Khellaf, A., Messlem, Y., & Chellali, F. (2016). Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach. *Renewable and Sustainable Energy Reviews*, 57, 238–249.
- Heo, K. Y., Ha, K. J., Yun, K. S., Lee, S. S., Kim, H. J., & Wang, B. (2014). Methods for uncertainty assessment of climate models and model predictions over East Asia. *International Journal of Climatology*, 34, 377–390.
- IPCC. (2007). *Climate change 2007: The physical science basis*. Intergovernmental Panel on Climate Change, 446, 727–728.
- Jeiouni, M., Delirhasannia, R., Alavipanah, S. K., Shahabi, M., & Samadianfard, S. (2015). Spatial analysis of groundwater electrical conductivity using ordinary kriging and artificial intelligence methods (Case study: Tabriz plain, Iran). *GEOFIZIKA*, 32(2), 191–208.
- Kaba, K., Sarigül, M., Avci, M., & Kandirmaz, H. M. (2018). Estimation of daily global solar radiation using deep learning model. *Energy*, 162, 126–135.
- Kaufmann, E., & Hagermann, A. (2015). Penetration of solar radiation into pure and Mars-dust contaminated snow. *Icarus*, 252, 144–149.
- Kisi, O. (2014). Modeling solar radiation of mediterranean region in Turkey by using fuzzy genetic approach. *Energy*, 64, 429–436.
- Koca, A., Oztop, H. F., Varol, Y., & Koca, G. O. (2011). Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey. *Expert Systems with Applications*, 38, 8756–8762.
- Lamnathou, C., & Chemisana, D. (2013). Solar radiation manipulations and their role in greenhouse claddings: Fluorescent solar concentrators, photoselective and other materials. *Renewable and Sustainable Energy Reviews*, 27, 175–190.
- Mehdizadeh, S., Behmanesh, J., & Khalili, K. (2016). Comparison of artificial intelligence methods and empirical equations to estimate daily solar radiation. *Journal of Atmospheric and Solar-Terrestrial Physics*, 146, 215–227.
- Mesbah, M., Soroush, E., & Rostampour Kakroudi, M. (2017). Predicting physical properties (viscosity, density, and refractive index) of ternary systems containing 1-octyl-3-methylimidazolium bis (trifluoromethylsulfonyl) imide, esters and alcohols at 298.15 K and atmospheric pressure, using rigorous classification techniques. *Journal of Molecular Liquids*, 225, 778–787.
- Ming, T., De_Richter, R., Liu, W., & Caillol, S. (2014). Fighting global warming by climate engineering: Is the earth radiation management and the solar radiation management any option for fighting climate change? *Renewable and Sustainable Energy Reviews*, 31, 792–834.
- Moazenzadeh, R., Mohammadi, B., Shamshirband, S., & Chau, K. W. (2018). Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. *Engineering Applications of Computational Fluid Mechanics*, 12(1), 584–597.
- Mohammadi, K., Shamshirband, S., Anisi, M. H., Alam, K. A., & Petkovic, D. (2015). Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Conversion and Management*, 91, 433–441.
- Mostafavi, E. S., Saeidi Ramiyani, S., Sarvar, R., Izadi Moud, H., & Mousavi, S. M. (2013). A hybrid computational approach to estimate solar global radiation: An empirical evidence from Iran. *Energy*, 49, 204–210.
- Mousavi, S. M., Mostafavi, E. S., Jaafari, A., Jaafari, A., & Hosseinpour, F. (2015). Using measured daily meteorological parameters to predict daily solar radiation. *Measurement*, 76, 148–155.
- Olatomiwa, L., Mekhilef, S., Shamshirband, S., & Petković, D. (2015). Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria. *Renewable and Sustainable Energy Reviews*, 51, 1784–1791.
- Ozgoren, M., Bilgili, M., & Sahin, B. (2012). Estimation of global solar radiation using ANN over Turkey. *Expert Systems with Applications*, 39, 5043–5051.
- Piri, J., & Kisi, O. (2015). Modelling solar radiation reached to the earth using ANFIS, NNARX, and empirical models (Case studies: Zahedan and Bojnurd stations). *Journal of Atmospheric and Solar-Terrestrial Physics*, 123, 39–47.
- Piri, J., Shamshirband, S., Petkovic, D., Tong, C. W., & Rehman, M. H. (2015). Prediction of the solar radiation on the Earth

- using support vector regression technique. *Infrared Physics & Technology*, 68, 179–185.
- Prescott, J. A. (1940). Evaporation from water surface in relation to solar radiation. *Transactions of the Royal Society of South Australia*, 64, 114–125.
- Qazi, A., Fayaz, H., Wadi, A., Raj, R. G., Rahim, N. A., & Ahmed Khan, W. (2015). The artificial neural network for solar radiation prediction and designing solar systems: A systematic literature review. *Journal of Cleaner Production*, 104, 1–12.
- Quinlan, J. R. (1992). Learning with continuous classes. In A. Adams & L. Sterling (Eds.), *Proc. AI'92 (Fifth Australian Joint Conf. On artificial intelligence*, (pp. 343–348). Singapore: World Scientific.
- Ramedani, Z., Omid, M., Keyhani, A., Khoshnevisan, B., & Saboohi, H. (2014). A comparative study between fuzzy linear regression and support vector regression for global solar radiation prediction in Iran. *Solar Energy*, 109, 135–143.
- Rao, D. V. S., Premalatha, M., & Naveen, C. (2018). Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ANN approach: A case study. *Renewable and Sustainable Energy Reviews*, 91, 248–258.
- Saini, R., & Kumar, P. (2016). Optimization of chlorpyrifos degradation by Fenton oxidation using CCD and ANFIS computing technique. *Journal of Environmental Chemical Engineering*, 4, 2952–2963.
- Samadianfard, S., Asadi, E., Jarhan, S., Kazemi, H., Kheshtgar, S., Kisi, O., . . . Abdul Manaf, A. (2018). Wavelet neural networks and gene expression programming models to predict short-term soil temperature at different depths. *Soil and Tillage Research*, 175, 37–50.
- Samadianfard, S., Delirhasannia, R., Kisi, O., & Agirre-Basurko, E. (2013). Comparative analysis of ozone level prediction models using gene expression programming and multiple linear regression. *GEOFIZIKA*, 30, 43–74.
- Samadianfard, S., Nazemi, A. H., & Sadraddini, A. A. (2014a). M5 model tree and gene expression programming based modeling of sandy soil water movement under surface drip irrigation. *Agriculture Science Developments*, 3(5), 178–190.
- Samadianfard, S., Sattari, M. T., Kisi, O., & Kazemi, H. (2014b). Determining flow friction factor in irrigation pipes using data mining and artificial intelligence approaches. *Applied Artificial Intelligence*, 28, 793–813.
- Shaddel, M., Javan, D. S., & Baghernia, P. (2016). Estimation of hourly global solar irradiation on tilted absorbers from horizontal one using artificial neural network for case study of Mashhad. *Renewable and Sustainable Energy Reviews*, 53, 59–67.
- Sharifi, S. S., Rezaverdinejad, V., & Nourani, V. (2016). Estimation of daily global solar radiation using wavelet regression, ANN, GEP and empirical models: A comparative study of selected temperature based approaches. *Journal of Atmospheric and Solar–Terrestrial Physics*, 149, 131–145.
- Singh, G., Sachdeva, S. N., & Pal, M. (2016). M5 model tree based predictive modeling of road accidents on non-urban sections of highways in India. *Accident Analysis and Prevention*, 96, 108–117.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14, 199–222.
- Solomatine, D., & Dulal, K. N. (2003). Model trees as an alternative to neural networks in rainfall-runoff modeling. *Hydrological Sciences Journal*, 48, 399–411.
- Sun, H., Zhao, N., Zeng, X., & Yan, D. (2015). Study of solar radiation prediction and modeling of relationship between solar radiation and meteorological variables. *Energy Conversion and Management*, 105, 880–890.
- Swartman, R. K., & Ogunlade, O. (1967). Solar radiation estimates from common parameters. *Solar Energy*, 11, 170–172.
- Taormina, R., Chau, K. W., & Sivakumar, B. (2015). Neural network river forecasting through baseflow separation and binary-coded swarm optimization. *Journal of Hydrology*, 529(3), 1788–1797.
- Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(7), 7183–7192.
- Tymvios, F. S., Jacovides, C. P., Michaelides, S. C., & Scouteli, C. (2005). Comparative study of Angströms and artificial neural networks' methodologies in estimating global solar radiation. *Solar Energy*, 78, 752–762.
- URL1. Retrieved from <https://www.google.com/maps/@34.3808769,50.7895208,2502254m/data=!3m1!1e3?hl=en>
- Vakili, M., Sabbagh-Yazdi, S. R., Khosrojerdi, S., & Kalhor, K. (2017). Evaluating the effect of particulate matter pollution on estimation of daily global solar radiation using artificial neural network modeling based on meteorological data. *Journal of Cleaner Production*, 141, 1275–1285.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer-Verlag, ISBN 0-387-98780-0.
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79–82.
- Willmott, C. J., Robeson, S. M., & Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, 32(13), 2088–2094.
- Wu, C. L., & Chau, K. W. (2011). Rainfall-runoff modeling using artificial neural network coupled with singular spectrum analysis. *Journal of Hydrology*, 399(3-4), 394–409.
- Wu, W., & Liu, H. B. (2012). Assessment of monthly solar radiation estimates using support vector machines and air temperatures. *International Journal of Climatology*, 32, 274–285.
- Zhang, D., & Tsai, J. J. P. (2007). *Advances in machine learning applications in software engineering*. Hershey, PA: IGI Publishing.