



# Solar radiation prediction using multi-gene genetic programming approach

Hatice Citakoglu<sup>1</sup> · Bilal Babayigit<sup>2</sup> · Nese Acanal Haktanir<sup>1</sup>

Received: 23 December 2019 / Accepted: 12 August 2020 / Published online: 18 August 2020  
© Springer-Verlag GmbH Austria, part of Springer Nature 2020

## Abstract

Accurate estimation of solar radiation both spatially and temporally is important for engineering studies related to climate and energy. The multi-gene genetic programming (MGGP) is proposed as a new compact method for this purpose, which is verified to yield more accurate solar radiation estimations in Turkey. Meteorological data such as extraterrestrial solar radiation, sunshine duration, mean of monthly maximum sunny hours, long-term mean of monthly maximum air temperatures, long-term mean of monthly minimum air temperatures, monthly mean air temperature, and monthly mean moisture data are selected as the MGGP model inputs. In the prediction models, the meteorological data measured from 163 stations in seven climate areas of Turkey over the period 1975–2015 are used. The MGGP model results for solar radiation prediction are found to be more accurate than the values given by some conventional empirical equations such as Abdalla, Angstrom, and Hargreaves–Samani. The performance of MGGP is also assessed for Turkey by single-data and multi-data models. The multi-data models of MGGP and the calibrated empirical equations are found to be more successful than the single-data models for solar radiation prediction.

**Keywords** Solar radiation · Empirical equations · Multi-gene genetic programming · Turkey

## 1 Introduction

Solar radiation (SR) is the primary energy source of water cycle, and thus, it is one of the key meteorological data having direct effects on developmental stages of plant growth and evapotranspiration (ET). ET is the summation of evaporation from the soil surface and transpiration from the plant leaves, and it is a significant parameter in agricultural and environmental implementations. ET is calculated from various meteorological data including SR. Because of its being an endless form of natural energy, and along with technologic developments such as photovoltaic solar panels, in the last 20 years, SR has been being used as a source of clean and inexhaustible energy in many countries. And, apparently, solar energy will keep being used at an increasing rate. Hence, SR is another kind of renewable energy used in residential establishments

and industrial facilities (Kalogirou et al. 1999). Therefore, SR is a critical variable influencing the hydrological, environmental, agricultural, ecological, and the industrial activities.

Satellite-based instruments have directly measured SR since 1978 and demonstrate that on mean, about 1361 W/m<sup>2</sup> achieves the top of the Earth's atmosphere. Parts of the Earth's surface, air pollution and clouds in the atmosphere act as mirrors and reflect about 30% of this energy back into space. Higher SR values are recorded when the sun is more active. Changes in lighting follow a sunspot cycle of approximately 11 years, with SR values fluctuating on average around 0.1% in recent cycles (Stocker et al. 2013).

SR strongly affects the evaporation from terrestrial surfaces. Llasat and Snyder (1998) noted that minor changes in SR might have distinctive impacts on reference evapotranspiration (ET<sub>0</sub>). Bois et al. (2008) demonstrated that ET<sub>0</sub> computed for Southern France by FAO-56-PM method was highly sensitive to SR. Cobaner (2011) employed the method of adaptive neuro-fuzzy inference system (ANFIS) for ET<sub>0</sub> estimation and used SR and air temperature (Ta) as the input data. Citakoglu et al. (2014) obtained more accurate ET<sub>0</sub> estimations when only the SR was taken as the input data than the models taking wind speed (WS), moisture, and Ta as explanatory variables.

✉ Hatice Citakoglu  
hccitakoglu@erciyes.edu.tr

<sup>1</sup> Department of Civil Engineering, Erciyes University, Kayseri, Turkey

<sup>2</sup> Department of Computer Engineering, Erciyes University, Kayseri, Turkey

There are significant correlations between SR and some meteorological variables. For instance, Guan et al. (2007) demonstrated positive correlations of solar energy as related with humidity, atmospheric pressure, WS, and  $T_a$ . Ododo et al. (1995) noted that besides the other meteorological data, SR was also dependent on maximum temperature ( $T_{\max}$ ). Hargreaves and Samani (1982) presented a simple equation for SR estimation as related to  $T_a$  and position of the sun. ElNesr et al. (2015) used relevant data of 29 climate stations in Saudi Arabia and noticed a reverse relationship of SR with the altitude. Rahimikhoob (2010) worked on artificial neural network (ANN) models to calculate SR and used  $T_a$  as the input variable.

Various empirical models have been proposed using sun hours,  $T_{\max}-T_{\min}$ , and moisture-like meteorological data in order to calculate SR. Angstrom (1924) used sunny hours as inputs for model development and proposed two models in estimating SR. Tymvios et al. (2006) and Mubiru and Banda (2008) used different meteorological data as input variables in ANN models for estimation of SR. Rehman and Mohandes (2008) used different input combinations with  $T_a$ , moisture, and days of the year and developed three ANN models for Saudi Arabia. Authors demonstrated that models with  $T_a$  and moisture variables provide adequate performance for SR estimation. Alam et al. (2009) developed an ANN model for SR estimation in India. Dastorani et al. (2010) proposed empirical equations, multiple regression, and artificial intelligence (AI) models for SR estimation and noted that the outcomes of AI models were superior to the other models. Since  $ET_0$  varies both spatially and temporally, latitude ( $La$ ), longitude ( $Lo$ ), and altitude ( $Al$ ) were also incorporated into ANN models developed for SR estimation (Rehman and Mohandes 2008; Mohandes et al. 1998; Hontoria et al. 2005; Siqueira et al. 2010). Linares-Rodríguez et al. (2011) selected  $La$ ,  $Lo$ , the number of days of total cloud cover,  $T_a$ , total water vapor, and total ozone data as the input variables of an ANN model for SR estimation with the aid of an ANN model with plausible success. Fadare et al. (2010) designed multi-layer, feed forward back propagation ANN models with different architectures by using  $La$ ,  $Lo$ ,  $Al$ , month, mean sunshine duration ( $n$ ),  $T_{\text{mean}}$ , and moisture as the input data and showed that fuzzy genetic model yielded better outcomes than ANN and ANFIS models. Olatomiwa et al. (2015) used  $n$ ,  $T_{\max}$ , and  $T_{\min}$  as the input data for SR estimation and employed support vector machines (SVM), firefly algorithm (FFA), ANN, and genetic programming (GP) models and demonstrated that the SVM–FFA model was an efficient machine learning approach for SR estimation in Tabas province of Iran. Mohammadi et al. (2015a) used ANFIS model to estimate SR by using the year as the single input and showed that the ANFIS model was successful for SR estimation. Mohammadi et al. (2015b) employed  $n$ ,  $T_a$ , moisture,  $T_{\text{mean}}$ , and extraterrestrial SR as input data for daily and monthly SR estimation in

coastal zones of Iran and developed SVM with wavelet transform algorithm (WT), ANN, and GP models. Mashaly and Alazba (2016) modeled instant thermal efficiency of the sun with the aid of meteorological and operational data in ANN, multivariate regression (MVR), and stepwise regression (SWR) methods. Researches indicated that ANN model had better performance values than the MVR and SWR models. Mehdizadeh et al. (2016) used meteorological data and compared the performance of AI models such as gene expression programming (GEP), ANN, and ANFIS in SR estimation and demonstrated that SR and meteorological parameter-dependent scenarios of ANN and ANFIS models achieved better accuracy than the empirical equations. Nasruddin et al. (2017) measured global SRs in Jakarta province of Indonesia for two whole years and calculated monthly SR with the aid of four empirical models in literature and linear regression model, and they concluded that the empirical models could be preferred in monthly SR estimation in Indonesia. These researchers also indicated the Allen equation as the best model for monthly SR estimation in Indonesia. Meenal and Selvakumar (2018) employed SVM, ANN, and empirical SR models for various experimental input combinations to estimate SR of different provinces of India. By using the Waikato Environment for Knowledge Analysis (WEKA) software, Meenal and Selvakumar (2018) observed that month,  $La$ ,  $T_{\max}$ , and shiny hours were the most effective parameters and the moisture as the least effective parameter for SR estimation. And, these researchers also indicated that the SVM model had a better performance than both ANN and empirical models. Keshtegara et al. (2018) conducted a research for SR estimation with four different heuristic regression methods, which were Kriging, response surface method (RSM), multivariate adaptive regression (MARS), and M5 model tree (M5Tree), using meteorological data as explanatory variables. These researchers demonstrated that M5Tree model estimations yielded erroneous outcomes compared with both maximum errors and the other models for minimum agreement. They also noted that the Kriging models had better performance values than the MARS, RSM, and M5Tree models. Ozoegwu (2018) stated that because Third World countries have too few measured meteorological data, single-data models had greater use than the multi-data models, and therefore, he indicated that temperature necessarily has to be the only distinctive data to be used in models to be developed for SR estimation. Since the Hargreaves–Samani equation depends on  $T_{\max}$  and  $T_{\min}$  in SR estimation, Ozoegwu (2018) developed temperature-dependent models for Nigeria and indicated that single meteorological data model was statistically more reliable. Kaba et al. (2018) applied a new model obtained by a combination of deep learning (DL) and multiple ANN methods to 34 stations of Turkey. They used SR data of the years 2001–2007 for model train and test and compared the simulated and actual data and demonstrated that

DL model yielded quite accurate outcomes for daily global SR estimation. They also indicated that DL model yielded more successful outcomes than various other methods in literature and also pointed out that DL model may constitute an alternative method and be used reliably in similar regions. Samadianfard et al. (2019) noted that SR was mostly predicted by artificial intelligence techniques or by empirical equations, and they applied SWR, ANFIS, GEP, and model trees (MT) approaches to detect the relations between SR and several meteorological variables in Tabriz region of Iran. As a result of their analyses, Samadianfard et al. (2019) observed that the relationship between the SR and  $n$  was pretty strong, and according to the performance evaluations, the SVR-6 model turned out to be better than the other models in predicting SR. Antonopoulos et al. (2019) used empirical equations, ANNs, and multi-linear regression methods (MLR) in order to estimate SR using daily meteorological measurements in Northern Greece. They have tried combinations of different input variables for ANNs and MLR models. They indicated that according to RMSE criteria, the results of the ANN models were in harmony with the results of the MLR models with the same input data. Fan et al. (2019a) used empirical equations, ANNs, ANFIS, SVM, MT, and machine learning models to estimate SR in different region of China. They demonstrated that the machine learning models performed better than the empirical models. They noted that ANFIS was highly amendable, while multivariate adaptive regression spline (MARS) and extreme gradient boosting (XGBoost) were also promising models to estimate daily SR. Fan et al. (2019b) used 72 existing empirical equations to evaluate for forecasting diffuse SR in different regions of China. They stated that sunshine duration-based equations and daily-based equations gave similar results. Chen et al. (2019) studied empirical equations available in literatures by using the commonly measured meteorological data and geographic factors. Chen et al. (2019) collected the entire 294 different empirical equations and divided the equations into 37 groups according to input data. They provided an overview of advanced empirical equations in the literature with their study, and they pointed out the more suitable equations in the three Gorges Reservoir area in China.

ANN has shown its success in many applications in various fields of civil engineering (e.g., Coulibaly et al. 2001; Coppola et al. 2003; Uncuoglu et al. 2008; Laman and Uncuoglu 2009; Bilgili and Sahin 2010; Altun and Dirikgil 2013; Cobaner et al. 2014; Citakoglu 2015; Bayram et al. 2016; Bayram and Al-Jibouri 2016). ANN is an effective algorithm for solving engineering problems and mainly consists of input layer, output layer, and one or more hidden layers. But how ANN defines the relationships between the input and the output is a black box, and there is no general framework for the selection of ANN structures and parameters. Consequently, it cannot easily obtain an explicit formulation of the relation between the input and output by using ANN.

GP is an approach of machine learning branch of artificial intelligence. The idea of GP was first proposed by Koza (1992) to handle challenging problems by making use of automatically evolving computer programs. Although GP has biological evolution inspiration like genetic algorithm, it has solutions represented by tree structures rather than binary strings. Also, GP does not require prior form of the existing relationships to achieve simplified prediction equations. Recent relevant literature shows that GP models successfully simulate the behavior of various branches of civil engineering problems (e.g, Aytek and Kisi 2008; Azamathulla et al. 2008; Kashid and Maity 2012; Gandomi et al. 2016; Kurugodu et al. 2018; Mehr et al. 2018). Recently, due to Searson et al. (2010), multi-gene genetic programming (MGGP) has emerged as a promising variant of traditional GP. MGGP can deal with non-linearity and maps the relationships among all involved factors by combining the model structure selection capability of the traditional GP and parameter estimation power of classical regression. Because MGGP uses small trees to compose a generalized model, it provides simpler models than that of traditional GP. Because of its success, there are several applications of MGGP in civil engineering problems (Gandomi and Alavi 2012a, 2012b; Bayazidi et al. 2014; Kumar et al. 2014; Cobaner et al. 2016a, b; Hadi and Tombul 2018; Pedrino et al. 2019).

In the present study, extraterrestrial radiation ( $R_a$ ), sunshine durations ( $n$ ), mean of monthly maximum sunny hours ( $N$ ), long-term mean of monthly maximum air temperatures ( $T_{\max}$ ), long-term mean of monthly minimum temperatures ( $T_{\min}$ ), monthly mean air temperature ( $T_{\text{mean}}$ ), and monthly mean moisture (relative humidity) ( $\text{RH}_{\text{mean}}$ ) data over the period from 1975 through 2015 are used as the input data to estimate SR values of Turkey. The main objective of this study is to obtain a prediction model by MGGP approach using SR values of Turkey. To achieve this objective, the following aspects are considered: (i) to investigate the accuracy of three different original empirical equations in SR estimation, (ii) to modify the coefficients of these three empirical equations for Turkey, (iii) to obtain more practical equations with the aid of MGGP method, (iv) to obtain new equations appropriate for Turkey with the aid of data used in empirical equations and compare these new equations with those calibrated three empirical equations, (v) to obtain a succinct equation suitable for Turkey needing only easy-to-acquire data like  $T_{\max}$  and  $T_{\min}$ , and (vi) to compare single-data models with multi-data models. The ultimate goal of the study summarized in this paper is to obtain an efficient and compact model for SR.

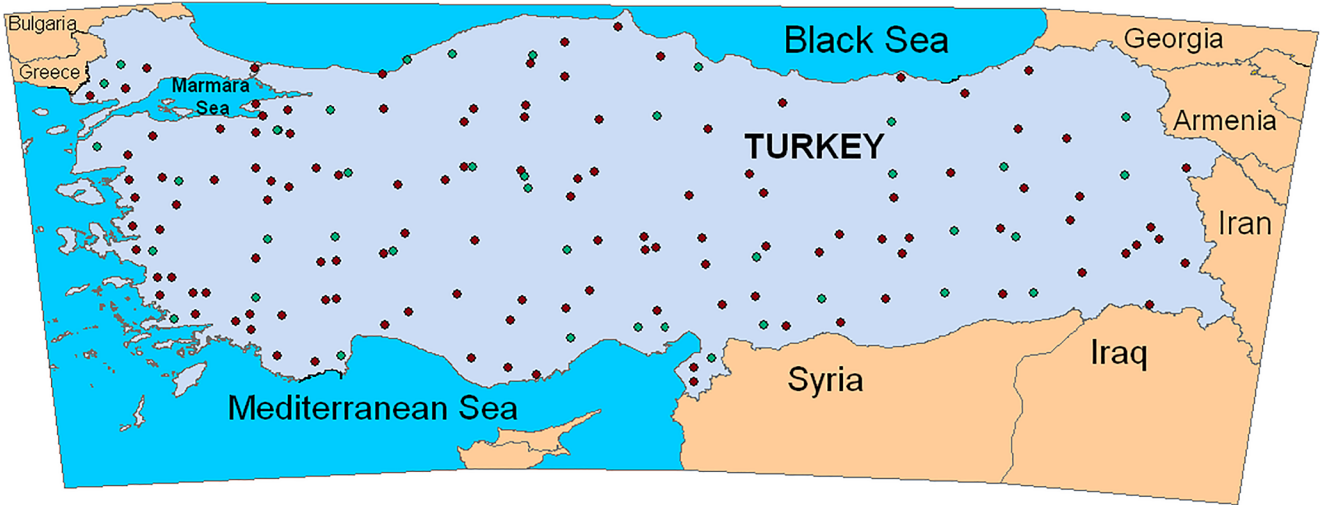


Fig. 1 Locations of the meteorological stations used in train are colored in red rounds and used in test are colored in green rounds

## 2 Materials and methods

### 2.1 Material

Long-term monthly means of meteorological data used in this study were measured in 163 stations operated by the General Directorate of State Meteorological Services (MGM). Measurements cover the period between the years 1975 and 2015. Turkey is located between 26'–45'' E longitudes and 36'–42'' N latitudes. The stations used cover all of seven climate zones of Turkey. The altitudes of the stations range between 3 and 2400 m. Location of the metrological stations in Turkey is illustrated in Fig. 1. As can be seen in Fig. 1, the stations used in the train and test are distributed uniformly over Turkey and are colored in red and green, respectively.

Fifty-six percent of Turkey has an altitude above 1000 m. The lands with a slope greater than 15% constitute 62% of country's total area. Twenty-five percent of the country is overlain by the first-, second-, and third-quality class soils, and 90% of these lands are used for agriculture. Of a total of 80 million hectares of land area of Turkey, 26.3 million hectares are used for cultivation activities. Turkey has quite large agricultural fields in different climate conditions. Therefore, different plant species are grown in terrestrial and coastal sections of the country. Fruits, vegetables, and citrus species both in open fields and in greenhouses are dominant in the Mediterranean region; olive, tobacco, and citrus are common in the Aegean region; sunflower is dominant in the Marmara region; cereals and sugar beet are the common plants in the Central Anatolia region; and finally, cotton and cereals are the major crops in the Southeastern Anatolia region.

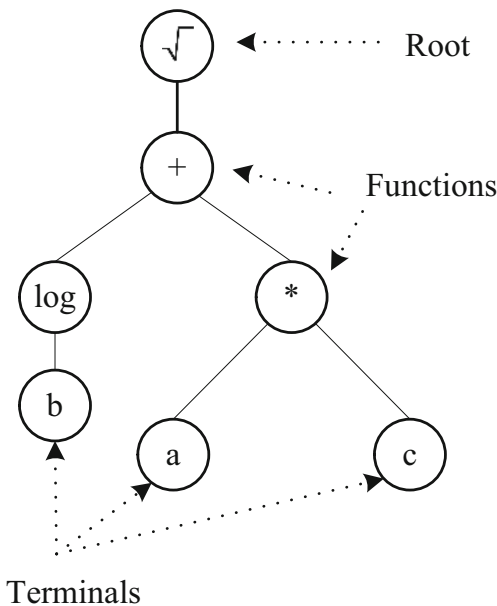


Fig. 2 A tree structure of a traditional GP model

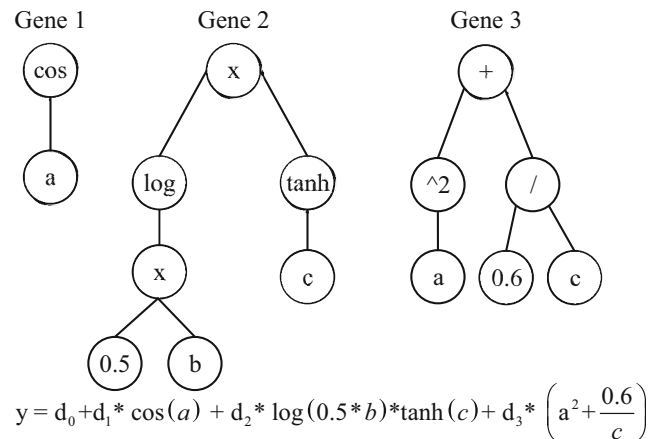


Fig. 3 An MGGP model tree structure



**Table 1** Correlation coefficients indicating the relationships between each input data and SR for different climate zones of Turkey

Regions	$N$ (h)	$n$ (h)	$T_{\max}$ (°C)	$T_{\min}$ (°C)	$T_{\text{mean}}$ (°C)	$RH_{\text{mean}}$ (%)	$R_a$ (MJ/m <sup>2</sup> )
Mediterranean region	0.962	0.910	0.775	0.625	0.742	-0.551	0.960
Blacksea region	0.944	0.916	0.760	0.567	0.743	-0.379	0.945
Ege region	0.966	0.928	0.846	0.732	0.830	-0.782	0.967
Marmara region	0.958	0.935	0.838	0.755	0.843	-0.691	0.960
Southeastern Anatolia region	0.986	0.954	0.894	0.867	0.893	-0.860	0.985
Eastern Anatolia region	0.948	0.826	0.762	0.760	0.776	-0.792	0.948
Central Anatolia region	0.935	0.916	0.861	0.796	0.859	-0.861	0.940
Entire Turkey	0.935	0.899	0.766	0.633	0.755	-0.697	0.943

## 2.2 Methods

### 2.2.1 Multi-gene genetic programming (MGGP)

Genetic programming (GP) is a symbolic regression technique developed to derive predictive models for challenging computational problems (Koza 1992). Different from other forms of regression such as linear regression, in GP, the form of the model is not defined a priori and does not require any assumptions to develop models. GP automatically evolves computer programs for solving a specific task through an evolutionary process. First, GP starts with a randomly generated individual program. Then, genetic operators like crossover and mutation are applied to select individual programs based on their fitness values to produce new individuals for the next generations. The programs are evolved and improved until better fitness values are obtained. Traditional GP model has a single-tree structure and is constructed by a set of functional or terminal sets of elements. A functional set can be arithmetic operators (+, \*, /, or -), mathematical functions (sin, cos, tanh, log, etc.), Boolean operators (AND, OR, NOT, etc.), or even logical expressions (IF or THEN). The terminal set can include variables (a, b, c, etc.) or constant values. Random selections are used for functions and terminals to provide and express tree structures. A tree structure of a traditional GP model ( $\sqrt{\log(b) + a*c}$  symbolic expression) is depicted in Fig. 2.

Multi-gene genetic programming (MGGP) is an automatic programming technique, and like GP, MGGP uses terminals and functions (Searson et al. 2010). MGGP can design effective prediction models because of its outstanding adaptability and versatility. Because it integrates the ability of traditional GP for selecting the model structure by the least squares technique, this peculiarity provides MGGP to generate models by using low-depth GP trees and combining low-order non-linear transformations of the input-output variables. Traditional GP uses a single-tree structure, but MGGP models are derived from several genes (each gene is a tree structure). Each gene is weighted ( $d_1, d_2, d_3, \dots$ ), and then, a random bias term ( $d_0$ ) is added to predict an output ( $y$ ). The schematic representation of an MGGP model tree structure is shown in Fig. 3.

The maximum gene number and tree depth are two essential control parameters of the MGGP model. These two parameters have significant effect on the complexity of the evolved models. Detailed explanations of MGGP parameters and operators are given in publications by Searson et al. (2010) and Gandomi and Alavi (2012a).

### 2.2.2 Empirical equations

There are several empirical equations in literature to define quantitative relations between SR and meteorological data. FAO recommends Angstrom and Hargreaves–Samani

**Table 2** MAE, MARE, MSE, RMSE, and  $R^2$  statistics of original equations for train and test stages

Equations	MAE (MJ/m <sup>2</sup> )	MARE (%)	MSE (MJ <sup>2</sup> /m <sup>4</sup> )	RMSE (MJ/m <sup>2</sup> )	$R^2$	
Training data	Abdalla equation	41.97	311.03	1930.9	43.94	0.910
	Angstrom equation	1.60	11.32	5.02	2.24	0.917
	Hargreaves–Samani equation	17.26	132.55	326.48	18.07	0.850
Test data	Abdalla equation	41.98	317.42	1931.30	43.95	0.906
	Angstrom equation	1.66	12.02	4.99	2.23	0.927
	Hargreaves–Samani equation	17.34	137.28	330.97	18.19	0.809

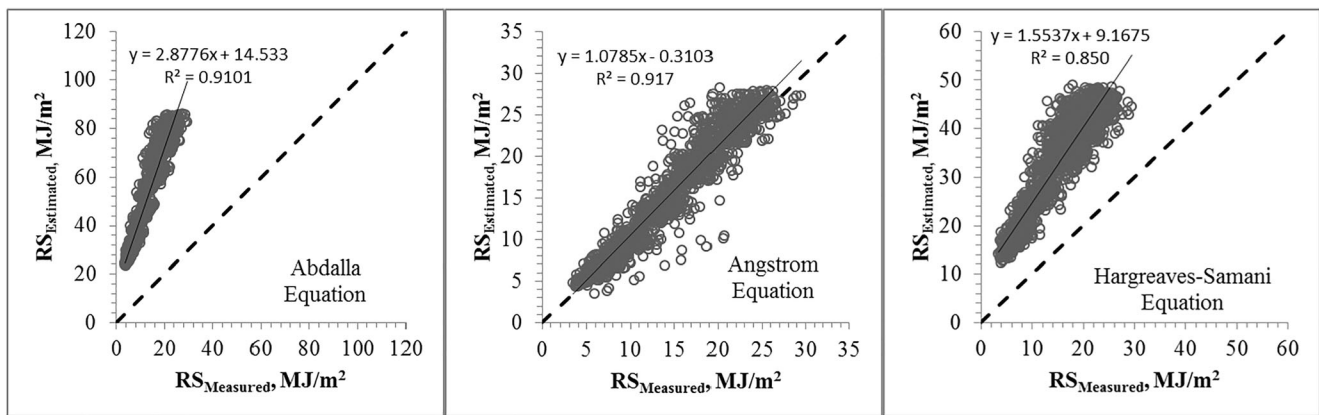


Fig. 4 SR values calculated with the original equations and measured by MGM for train data

equations for SR estimation. First, we have used the Angstrom equation for estimation of SR. This equation related SR with extraterrestrial radiation reaching to earth and sunshine durations ( $n$ ) as expressed below:

$$R_s = \left( a + b \cdot \frac{n}{N} \right) \cdot R_a \quad (1)$$

where  $R_s$  is solar radiation ( $\text{MJ}/\text{m}^2$ ),  $R_a$  is extraterrestrial radiation ( $\text{MJ}/\text{m}^2$ ),  $n$ ; sunshine durations,  $N$  is maximum monthly mean sunshine durations, and  $a$  and  $b$  are coefficients for which the values recommended in [11] are  $a = 0.25$ ,  $b = 0.50$ .

Hargreaves and Samani (1982) developed an equation for estimation of SR as related to  $T_{\max}$ ,  $T_{\min}$ , and  $R_a$ . Since it used data which is easiest to gauge ( $T_a$ ) as explanatory variables, this equation has become the most practical one, and it is rewritten below:

$$R_s = a \cdot R_a \cdot \sqrt{T_{\max} - T_{\min}} \quad (2)$$

where  $T_{\max}$  is maximum long-term monthly mean temperature ( $^{\circ}\text{C}$ ),  $T_{\min}$  is minimum long-term monthly mean temperature ( $^{\circ}\text{C}$ ), and  $a$  is a coefficient, for which the numerical values of 0.19 for coastal zones and 0.17 for terrestrial zones were recommended by Hargreaves and Samani (1982).

In recent years, several researchers developed different empirical equations for SR estimation relating it to various meteorological data. For instance, Abdalla (1994) recommended a new equation for SR estimation from sunshine durations ( $n$ ), maximum monthly mean sunshine durations ( $N$ ), mean temperature, and relative humidity, which is:

$$R_s = \left( a + b \cdot \frac{n}{N} + c \cdot T_{\text{mean}} + d \cdot \text{RH}_{\text{mean}} \right) \cdot R_a \quad (3)$$

where  $T_{\text{mean}}$  is monthly mean temperature;  $\text{RH}_{\text{mean}}$  is monthly mean moisture; and  $a$ ,  $b$ ,  $c$ ,  $d$  are coefficients, for which the numerical values  $a = 1.943$ ,  $b = 0.577$ ,  $c = -0.01483$ , and  $d = -0.12129$  are recommended by Abdalla (1994).

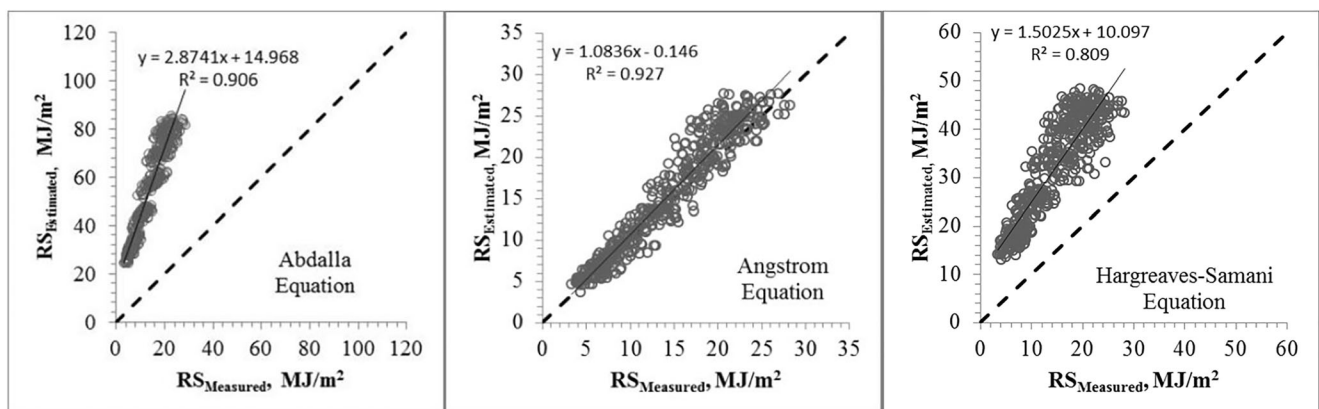


Fig. 5 SR values calculated with the original equations and measured by MGM for test data

**Table 3** The coefficients of the original and the calibrated equations

	Coefficients of original equation				Coefficients of calibrated equation			
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
Abdalla equation	1.943	0.577	-0.01483	-0.12129	0.54	0.29	-0.0026	-0.0024
Angstrom equation	0.25	0.50	—	—	0.34	0.30	—	—
Hargreaves–Samani equation	0.17–0.19	—	—	—	0.089	—	—	—

### 2.2.3 Calibration with Microsoft Excel Solver

We have used the Excel Solver to find the optimal value of the equation in the target cell. The Excel Solver interoperates with a cell group directly or indirectly related to the equation in the target cell. Constraints can be set for the values to be used in the model by the Solver, and such constraints can also be applied to the other cells influencing the equation of the target cell [Excel–help].

The enlisted steps below were followed to find out numerical relations between SR and the other meteorological data with the aid of Excel Solver.

1. Open the Solver parameters dialog box from Solver command in analysis group of data tab.
2. In the set target cell box, enter data points for objective cell or enter into Excel worksheet. The target cell should contain an equation. In this study, as the target cell, the cell containing the evaluation parameters of mean absolute error (MAE), root mean square error (RMSE), and mean absolute relative error (MARE) are selected.
3. From the target cell menu, the “minimum” option is selected for each one of the MAE, RMSE, and MARE parameters, because of the obvious reason that the closer the parameter to zero the better the tested model.
4. From the changing variables menu, a name is entered for each decision variable cell range. Variable cells should be directly or indirectly related to the target cell. In this study, the cells containing the values of *a*, *b*, *c*, and *d* coefficients

will be selected as variable cells. Initial values of *a*, *b*, *c*, and *d* coefficients are given as (1,1,1,1).

5. From the Solver parameters dialog box, proper one of three algorithms or solving method is selected. The non-linear generalized reduced gradient (GRG) scheme is used for smooth non-linear problems, while the simple LP method is used for linear problems, and the expansion method is used for non-smooth problems. The GRG method is used in this study.
6. Following the click on the solve button, the equation is solved by the Solver and those numerical values of *a*, *b*, *c*, and *d* coefficients with the least error values are obtained.

There are several equations available in literature for SR estimation. In this study, Angstrom, Abdalla, and Hargreaves–Samani equations are considered incorporated with the mentioned meteorological data in Turkey. The explanatory meteorological variables for these equations include maximum, minimum, and mean temperatures; extraterrestrial radiation; mean moisture; *n*; and monthly maximum sunshine durations. The data measured in 163 meteorological stations are divided into two portions, 75% of which (randomly selected 1467 elements) are used to develop the models and the remaining 25% (randomly selected 489 data) are used to test the power of the models. The stations used for model train are indicated in red circles, and the stations used for model tests are indicated in green circles

**Table 4** MAE, MARE, MSE, RMSE, and  $R^2$  statistics of calibrated equations for train and test stages

	Equations	MAE	MARE	MSE	RMSE	$R^2$
		(MJ/m <sup>2</sup> )	(%)	(MJ <sup>2</sup> /m <sup>4</sup> )	(MJ/m <sup>2</sup> )	
Training data	Abdalla equation	1.161	9.047	2.428	1.558	0.938
	Angstrom equation	1.277	9.892	2.903	1.704	0.925
	Hargreaves–Samani equation	2.063	17.155	6.638	2.577	0.850
Test data	Abdalla equation	1.293	10.088	2.876	1.696	0.927
	Angstrom equation	1.252	10.111	2.736	1.654	0.930
	Hargreaves–Samani equation	2.323	20.141	8.097	2.845	0.809

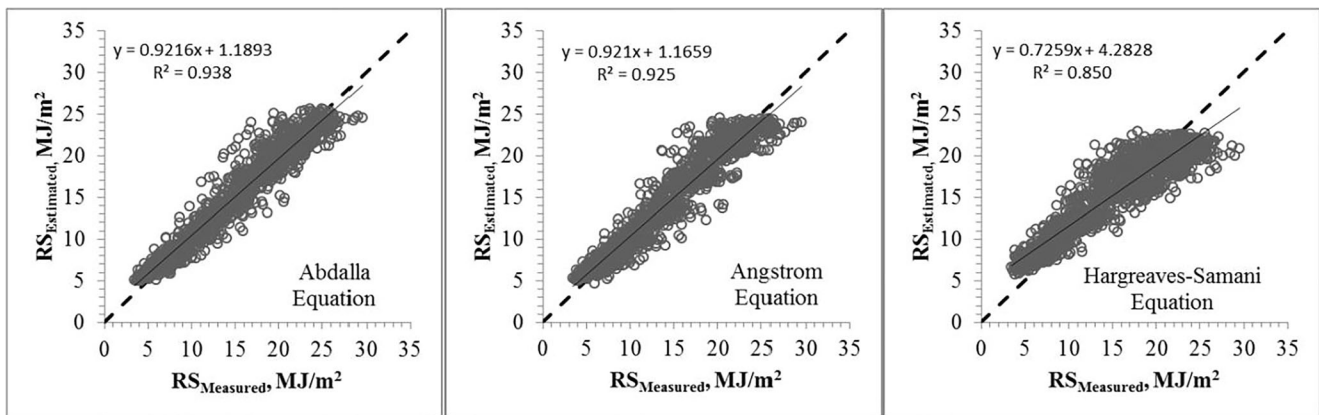


Fig. 6 SR values calculated with the calibrated equations and measured by MGM for train data

(Fig. 1). In the first stage, train and test data sets are used to calculate SR values for Turkey with the aid of original equations. In the second stage, *a*, *b*, *c*, and *d* coefficients of original equations were computed for 1975–2015 period, and actual SR values by using the long-term monthly mean values of the meteorological data were recorded at 163 stations with the aid of Solver menu of Microsoft Excel. In the third stage, the meteorological data of Angstrom, Abdalla, and Hargreaves–Samani equations are used to develop new equations with the aid of the MGGP method.

2.2.4 Performance assessment criteria

In this study, extraterrestrial radiation ( $R_a$ ), sunshine durations ( $n$ ), mean of monthly maximum sunny hours ( $N$ ), long-term mean of monthly maximum air temperatures ( $T_{max}$ ), long-term mean of monthly minimum temperatures ( $T_{min}$ ), monthly mean air temperature ( $T_{mean}$ ), and monthly mean moisture ( $RH_{mean}$ ) data measured at 163 stations of Turkey are used as input for (1) multi-gene genetic programming (MGGP), (2) calibrated three

empirical equations, and (3) original empirical equations to estimate SR. Performance assessment criteria include mean absolute error (MAE), mean absolute relative error (MARE), mean square error (MSE), and root mean square error (RMSE). In this study, MAE, MARE, MSE, and RMSE are defined as follows:

$$MAE = \frac{1}{N_s} |RS_e - RS_m| \tag{4}$$

$$MARE = 100 \frac{1}{N_s} \left| \frac{RS_e - RS_m}{RS_m} \right| \tag{5}$$

$$MSE = \frac{1}{N_s} (RS_e - RS_m)^2 \tag{6}$$

$$RMSE = \sqrt{\frac{1}{N_s} (RS_e - RS_m)^2} \tag{7}$$

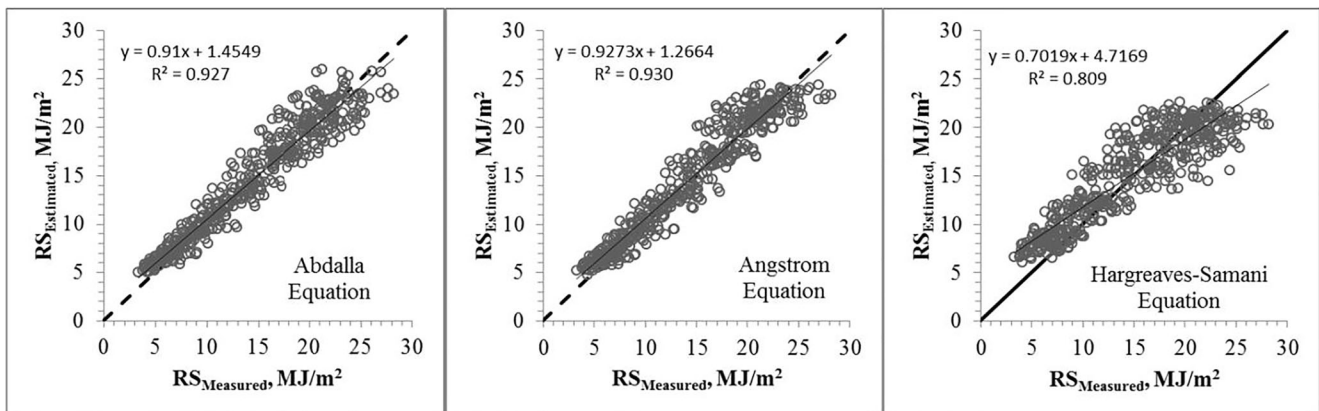


Fig. 7 SR values calculated with the calibrated equations and measured by MGM for test data



**Table 5** The parameter settings of the MGGP model

Parameter	Values
Function node set	+, -, *, sin, exp., tan
Population size	700
Maximum number of generations	100
Probability of reproduction	0.05
Probability of mutation	0.1
Probability of crossover	0.85
Elitism	0.02
Maximum gene depth	4–5
Maximum number of genes	3–4
Tournament size	4
Tree building method	Ramped half and half
Method of selection for tournament	Lexicographic

### 3 Results and discussions

Turkey lies in a climatologically heterogeneous area of about 780 thousand km<sup>2</sup> in the northern hemisphere. Accordingly, first, simple regressions between SR against each one of these input variables are done in order to have an initial assessment of dependence of SR individually on these explanatory variables. The correlation coefficients of the relationships between SR and each input data for different climate areas of Turkey are given in Table 1. The correlation coefficients for the minimum temperature and moisture of the Black Sea region are lower than those of the other regions. However, in general, the correlation coefficients for all of seven regions of Turkey are not that much different from each other, and they are also close to those for Turkey handled as one region. So, instead of trying separate models for each region individually, a generalized model is attempted for Turkey as a whole.

Before developing new equations by MGGP and numerical optimization methods, comparison criteria (error values) are determined for the original equations considered in this study (Table 2). Comparison criteria of the original equations are given in Table 2. As can be seen in Table 2, the comparison criteria of the train and test data are similar. While original Abdalla and Hargreaves–Samani equations have quite large error values, original Angstrom equation yields reasonable ones. The comparison of the measured and estimated SR values from train and test data is presented in Figs. 4 and 5. As can be seen in Figs. 4 and 5, Angstrom equation yields better estimations for both test data sets. The values computed by Abdalla and Hargreaves–Samani equations are greater than the measured values. As seen in Figs. 4 and 5, Abdalla and Hargreaves–Samani equations have high

determination coefficients. But, it can be observed that no points estimated Abdalla and Hargreaves–Samani equations drop on the  $y = x$  (45°) line. It is clear from Figs. 4 and 5 that the values computed by Abdalla and Hargreaves–Samani equations are not in compliance with the measured values, while the values yielded by Angstrom equation are more consistent.

The magnitudes of  $a$ ,  $b$ ,  $c$ , and  $d$  coefficients of Angstrom, Abdalla, and Hargreaves–Samani equations are calibrated for entire Turkey with the aid of Excel Solver. In this stage, three different equations are used, and hence, three different calibrations are performed. Coefficients of the original and the calibrated equations are given in Table 3. According to Table 3, the new coefficients are different from the original ones. As mentioned before, the coefficients of the original Hargreaves–Samani equation are different for coastal and inland areas. However, such a difference is not obtained in this study. The comparison criteria for three different calibration equations are given in Table 4. According to Table 4, the error values of the calibrated equations are appreciably lower than those of the original equations. There is a distinct decrease in error values of the calibrated Abdalla and Hargreaves–Samani equations. As can be seen from Figs. 6 and 7, the decrease in error values is reflected also on scatter diagrams of the calibrated equations. Yet, deviations from  $y = x$  (45°) line by the calibrated Hargreaves–Samani equation are still more than those of the other two equations in scatter plots of the train and test stages (Figs. 6 and 7).

MGGP method is an alternative method employed in this study. SR values are also estimated by the MGGP method. The MGGP parameters listed in Table 5 are used to generate the prediction models for SR.

The selection of parameter values of the MGGP approach is very important because it influences the generalization capability of the models to be formed. Therefore, the parameters given in Table 5 are determined through the preliminary studies. Especially, the population size and the number of generations are selected based on the complexity of the problem. The larger the values of these two parameters, the longer the study will take. The maximum number of genes ( $G_{max}$ ) and the maximum gen depth ( $D_{max}$ ) influence the size of the search universe and the solutions to be discovered in this search universe. More successful outcomes are achieved by increasing these parameters, but the complexity of the problem and the study duration increases in that case. Different  $G_{max}$  and  $D_{max}$  values are tried to make the model produce more

**Table 6** MAE, MARE, MSE, RMSE, and  $R^2$  statistics of equations obtained by the MGGP method for train and test stages

Equations		MAE ( M J / m <sup>2</sup> )	MARE (%)	MSE ( M J <sup>2</sup> / m <sup>4</sup> )	RMSE ( M J / m <sup>2</sup> )	$R^2$
Training data	Abdalla Equation	1.038	7.489	2.068	1.438	0.947
	Angstrom Equation	1.184	8.835	2.557	1.599	0.934
	Hargreaves–Samani Equation	1.315	9.741	3.138	1.771	0.919
Test data	Abdalla Equation	1.149	8.319	2.410	1.552	0.938
	Angstrom Equation	1.201	9.324	2.585	1.608	0.934
	Hargreaves–Samani Equation	1.408	10.751	3.387	1.840	0.913

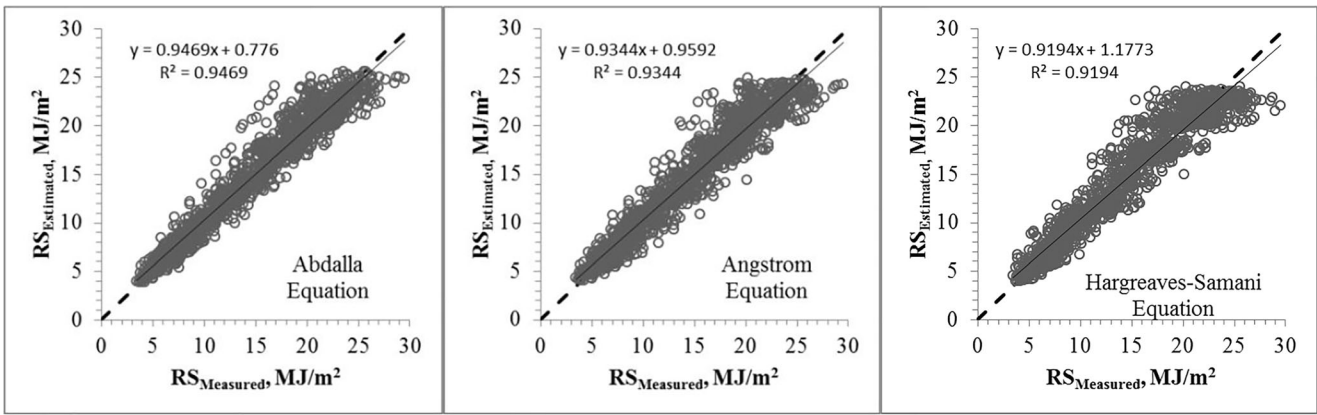


Fig. 8 SR values calculated with MGGP and measured by MGM for train data

accurate outcomes, and the best models are obtained by  $D_{max}$  of 4–5 and  $G_{max}$  of 3–4. MSE values are used for goodness-of-fit, and the model with the least MSE is identified as the best model. The maximum number of generations is used as the termination criterion. MAE, MARE, MSE, RMSE, and determination coefficient  $R^2$  criteria are used to assess the prediction models.

Equations generated by the MGGP method by using the data of Abdalla equation ( $n, N, T_{mean}, RH_{mean}, R_a$ ), of

$$R_S = 0.8348 \times R_a - 0.4174 \times T_{mean} - 0.8348 \times N + 0.8348 \times n + 0.01629 \times n \times (RH_{mean} + 3 \times n) - 0.003428 \times RH_{mean} \times (Ra - T_{mean} + 5 \times n + 4.65) - 0.01629 \times N \times (2 \times n + \sin(N)) + 0.001393 \times [T_{mean} - n + RH_{mean}] \times (T_{mean} - n + \sin(n)) \times (N - n + \sin(n - N) + \sin(n)) + 5.099 \quad (8)$$

The equation generated with the data of Angstrom equation is:

$$R_S = 0.8738 \times R_a - 0.4369 \times n - 0.4369 \times N - 0.1311 \times N^2 - 0.4369 \times \sin(N + 0.9479) - 0.1311 \times (N - 3.58) \times (N - n) - (12.27 \times (N - R_a + n \times R_a)) / \exp(n + 3.735) + 3.618 \quad (9)$$

Angstrom equation ( $n, N, R_a$ ), and of Hargreaves–Samani equation ( $T_{max}, T_{min}, R_a$ ) are given below:

The equation generated with the data of Abdalla equation is:

The equation generated with the data of Hargreaves–Samani equation is:

$$R_S = 0.263 \times R_a - 0.5171 \times T_{max} + 0.007334 \times \sin(0.6695 \times T_{max}) + 0.8905 \times \sin(0.6695 \times R_a) - 1.881 \times \sin(R_a + \sin(R_a)) - 0.07122 - 0.0113 \times (0.2697 \times R_a - \sin(R_a)) \times (T_{min} - 2 \times R_a + 10.69) + 0.007334 \times (T_{max} + 9.993) \times (T_{max} + 10.61) + 5.074$$

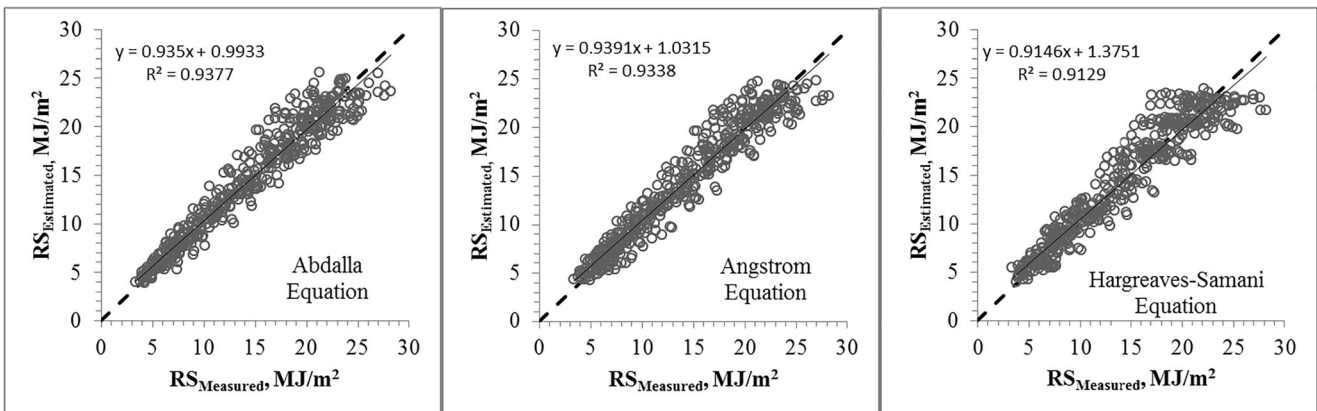


Fig. 9 SR values calculated with MGGP and measured by MGM for test data

Comparisons of MAE, MARE, MSE, RMSE, and  $R^2$  values of the computed results of Eqs. 8, 9, and 10 against the measured values at train and test stages are provided in Table 6. According to Table 6, with the use of MGGP method, MARE values are decreased by 7–10% for train and 8–11% for test, MSE values decreased by 2–3.3% for train and test, MARE and RMSE values decreased by 1–1.8% for train and 1.5–1.8% for test. It is observed in scatter plots in Figs. 8 and 9 that SR values obtained by the MGGP method are much scattered around the trend line than original equations.

There are three different equations considered in this study: Abdalla, Angstrom, and Hargreaves–Samani. In the last stage of the study, first, comparisons are made for each one of these three equations separately. Then, all equations are compared. While SR estimation by the original Hargreaves–Samani equation turns out to be the least successful one, numerical optimization and MGGP methods yield much closer estimations to the actual values. Considering the RMSE values in Tables 2, 4, and 6, it is observed that the MGGP method has much lower RMSE values than the original Hargreaves–Samani equation and the calibrated Hargreaves–Samani equation. Therefore, the MGGP method is found to be much more successful than the other methods. Although the original *Angstrom* equation has lower error values than the other two original equations, it is not sufficiently successful in the estimation of SR values obtained with the aid of numerical optimization and the MGGP methods. It is observed in Tables 2, 4, and 6 that the RMSE values of three different methods are close to each other. The original Abdalla equation is identified as the weakest method in SR estimation. The MARE value of the original Abdalla equation for test data is identified as 317.42 (Table 2). However, the MARE values after numerical optimization and the MGGP methods for test data drop down to 10.09 and 8.32 (Tables 4 and 6), respectively. When the numerical optimization and the MGGP methods are compared, it is observed that the MGGP method has lower error values.

Comparing the magnitudes of the parameters used for goodness criteria for all of the methods and equations in Tables 2, 4, and 6, it is obvious that the MGGP method has the least error values and the best method for SR estimation. The equations obtained with MGGP approach are much practical and easy than the original and calibrated empirical equations.

## 4 Conclusions

In this paper, solar radiation values are estimated for Turkey using extraterrestrial radiation, maximum temperature ( $T_{\max}$ ), minimum temperature, mean temperature sunshine durations, maximum monthly mean sunshine durations, and mean moisture data as independent variables. Original Abdalla, Angstrom, and Hargreaves–Samani equations, calibrated Abdalla, Angstrom, and Hargreaves–Samani equations and new equations improved by MGGP method are used for solar radiation prediction. These

three different equations and three different methods are compared based on the commonly used performance criteria of MAE, MARE, MSE, RMSE, and  $R^2$  statistics. The following conclusions are drawn from the present study:

- When the original Abdalla, Angstrom, and Hargreaves–Samani equations are compared, the Angstrom equation yields closer estimations to the SR values measured by the General Directorate of Meteorology of Turkey (MGM) than the other two equations.
- When the estimation performance of the original, calibrated, and MGGP-derived equations are compared at train and test stages, it is observed that the SR values estimated by the MGGP method are closer to the SR values measured by MGM.
- The MGGP-derived equation with the use of data of Angstrom equation is not found to be sufficiently successful in SR estimation as compared with the other two methods.
- Of the equations developed by the MGGP method by using the data of the original equations, the new equation obtained through the use of the data of Abdalla equation yields better outcomes in terms of the error criteria than the new equations obtained through the use of the data of the other two methods.
- Although multi-data models are not preferred for the third world countries, in MGGP model, multi-data models yield better outcomes than the single-data models.

**Acknowledgments** The authors wish to thank the Turkish State Meteorological Service for providing the long-term monthly mean of meteorological data.

## References

- Abdalla YAG (1994) New correlation of global solar radiation with meteorological parameters for Bahrain. *Int J Solar Energy* 16:111–120
- Alam S, Kaushik S, Garg N (2009) Assessment of diffuse solar energy under general sky condition using artificial neural network. *Appl Energy* 86:554–564
- Altun F, Dirikgil T (2013) The prediction of prismatic beam behaviours with polypropylene fiber addition under high temperature effect through ANN, ANFIS and fuzzy genetic models. *Compos Part B Eng* 52:362–371
- Angstrom A (1924) Solar and terrestrial radiation. *QJR Meteorol Soc* 50: 121–126
- Antonopoulos VZ, Papamichail DM, Aschonitis VG, Antonopoulos AV (2019) Solar radiation estimation methods using ANN and empirical models. *Comput Electron Agric* 160:160–167
- Aytek A, Kisi O (2008) A genetic programming approach to suspended sediment modelling. *J Hydrol* 351(3–4):288–298
- Azamathulla H, AbGhani A, Zakaria NA, Lai SH, Chang CK, Leow CS, Abuhasan A (2008) Genetic programming to predict ski-jump bucket spill-way scour. *J Hydrodyn* 20(4):477–484
- Bayazidi, A.M., Wang, G.-G., Bolandi, H., Alavi, A.H., Gandomi, A.H., 2014. Multigene genetic programming for estimation of elastic

- modulus of concrete. *Mathematical problems in engineering*. Article ID 474289
- Bayram S, Al-Jibouri S (2016) Efficacy of estimation methods in forecasting building projects' costs. *J Constr Eng Manag* 142(11):050160121–050160129
- Bayram S, Ocal M, Laptali OE, Atis C (2016) Comparison of multi layer perceptron (MLP) and radial basis function (RBF) for construction cost estimation: the case of Turkey. *J Civil Eng Manag* 22:480–490
- Bilgili M, Sahin B (2010) Comparative analysis of regression and artificial neural network models for wind speed prediction. *Meteorog Atmos Phys* 109(1–2):61–72
- Bois B, Pieri P, Van Leeuwen C, Wald L, Huard F, Gaudillere JP, Saur E (2008) Using remotely sensed solar radiation data for reference evapotranspiration estimation at a daily time step. *Agric For Meteorol* 148:619–630
- Chen J, He L, Yang H, Ma M, Chen Q, Wu SJ, Xiao Z (2019) Empirical models for estimating monthly global solar radiation: a most comprehensive review and comparative case study in China. *Renew. Sust Energy Rev* 108:91–111
- Citakoglu H (2015) Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation. *Comput Electron Agric* 118:28–37
- Citakoglu H, Cobaner M, Haktanir T, Kisi O (2014) Estimation of monthly mean reference evapotranspiration in Turkey. *Water Resour Manag* 28(1):99–113
- Cobaner M (2011) Evapotranspiration estimation by two different neuro-fuzzy inference systems. *J Hydrol* 398(3–4):292–302
- Cobaner M, Citakoglu H, Kisi O, Haktanir T (2014) Estimation of mean monthly air temperatures in Turkey. *Comput Electron Agric* 109:71–79
- Cobaner M, Babayigit B, Dogan A (2016a) Estimation of groundwater levels with surface observations via genetic programming. *J Am Water Works Assoc* 108:E335–E338
- Cobaner M, Babayigit E, Babayigit B (2016b) Estimation of groundwater level with genetic programming using meteorological data. *Nigde Univ J Eng Sci* 5(2):177–187
- Coppola E Jr, Poulton M, Charles E, Dustman J, Szidarovsky F (2003) Application of artificial neural networks to complex groundwater management problems. *Nat Resour Res* 12(4):303–320
- Coulibaly P, Anctil F, Aravena R, Bobee B (2001) Artificial neural network modeling of water table depth fluctuations. *Water Resour Res* 37(4):885
- Dastorani MT, Moghadamnia A, Piri J, Rico Ramirez M (2010) Application of ANN and ANFIS models for reconstructing missing flow data. *Environ Monit Assess* 166(1–4):421–434
- ElNesr MN, Alazba AA, Amina MT (2015) Estimation of shortwave solar radiations in the Arabian Peninsula: a new approach. *Desalin Water Treat* 57(1):37–50
- Fadare DA, Irimese I, Oni AO, Falana A (2010) Modeling of solar energy potential in Africa using an artificial neural network. *Am J Sci Ind Res* 1(2):144–157
- Fan J, Wu L, Zhang F, Cai H, Zeng W, Wang X, Zou H (2019a) Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: a review and case study in China. *Renew Sust Energy Rev* 100:186–212
- Fan J, Wua L, Zhanga F, Caia H, Mab X, Baic H (2019b) Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China. *Renew. Sust. Energy. Rev.* 105:168–186
- Gandomi AH, Alavi AH (2012a) A new multi-gene genetic programming approach to nonlinear system modeling, part i: materials and structural engineering problems. *Neural Comput & Applic* 21(1):171
- Gandomi AH, Alavi AH (2012b) A New multi-gene genetic programming approach to nonlinear system modeling, part ii: geotechnical and earthquake engineering problems. *Neural Comput & Applic* 21(1):189
- Gandomi AH, Sajedi S, Kiani B, Huang Q (2016) Genetic programming for experimental big data mining: a case study on concrete creep formulation. *Autom Constr* 70:89–97
- Guan L, Yang J, Bell JM (2007) Cross-correlation between weather variables in Australia. *Build Environ* 42(3):1054–1070
- Hadi SJ, Tombul M (2018) Monthly streamflow forecasting using continuous wavelet and multi-gene genetic programming combination. *J Hydrol* 561:674–687
- Hargreaves, G.H., Samani, Z.A., 1982. Estimating potential evapotranspiration. *J. Irrig. Drain. Eng.* 108(3), 225–230
- Hontoria L, Aguilera J, Zufiria P (2005) An application of the multilayer perceptron: solar radiation maps in Spain. *Sol Energy* 79(5):523–530
- Kaba K, Sarıgül M, Avcı M, Kandirmaz HM (2018) Estimation of daily global solar radiation using deep learning model. *Energy*. 162(2018):126–135
- Kalogirou, S.A., Neocleous, C., Paschiardis, S., Schizas, C., 1999. Wind speed prediction using artificial neural networks. *European Symposium on Intelligent Techniques ESIT '99*, June 3–4, Crete (Greece)
- Kashid SS, Maity R (2012) Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using genetic programming. *J. Hydrol.* 454–455:26–41
- Keshtegara B, Mert C, Kisi O (2018) Comparison of four heuristic regression techniques in solar radiation modeling: Kriging method vs RSM, MARS and M5 model tree. *Renew. Sust Energy Rev* 81(2018):330–341
- Koza JR (1992) *Genetic programming: on the programming of computers by means of natural selection*. MIT Press, Cambridge
- Kumar B, Jha A, Deshpande V, Sreenivasulu G (2014) Regression model for sediment transport problems using multi-gene symbolic genetic programming. *Comput Electron Agric* 103:82–90
- Kurugodu HV, Bordoloi B, Hongc Y, Garg A, Garg A, Sreedeeep S, Gandomi AH (2018) Genetic programming for soil-fiber composite assessment. *Adv Eng Softw* 122:50–61
- Laman M, Uncuoglu E (2009) Prediction of the moment capacity of pier foundations in clay using neural networks. *Kuwait J Sci Eng* 36(1B):33–52
- Linares-Rodríguez A, Ruiz-Arias JA, Pozo-Vázquez D, Tovar-Pescador J (2011) Generation of synthetic daily global solar radiation data based on ERA-interim reanalysis and artificial neural networks. *Energy*. 36:5356–5365
- Llasat MC, Snyder RL (1998) Data error effects on net radiation and evapotranspiration estimation. *Agric For Meteorol* 91(3/4):209–221
- Mashaly, A.F., and Alazba A.A., 2016. Comparison of ANN, MVR, and SWR models for computing thermal efficiency of a solar still. *International J. Green Energy*, 13(10), 1016–1025
- Meenal R, Selvakumar AI (2018) Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. *Renew. Energy* 121:324–343
- Mehdizadeh S, Behmanesh J, Khalili K (2016) Comparison of artificial intelligence methods and empirical equations to estimate daily solar radiation. *J Atmos Solar Terr Phys* 146:215–227
- Mehr AD, Nourani V, Kahya E, Hrnjica B, Sattar AMA, Yaseeng ZM (2018) Genetic programming in water resources engineering: a state-of-the-art review. *J Hydrol* 566:643–667
- Mohammadi K, Shamshirband S, Tong CW, Alam KA, Petkovic D (2015a) Potential of adaptive neuro-fuzzy system for prediction of daily global solar radiation by day of the year. *Energy Convers Manag* 93:406–413
- Mohammadi K, Shamshirband S, Tong CW, Arif M, Petkovic D, Sudheer C (2015b) A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation. *Energy Convers Manag* 92:162–171



- Mohandes M, Rehman S, Halawani TO (1998) Estimation of global solar radiation using artificial neural networks. *Renew Energy* 14(1–4): 179–184
- Mubiru J, Banda EJKB (2008) Estimation of monthly average daily global solar irradiation using artificial neural networks. *Sol Energy* 82(2):181–187
- Nasruddin, Budiayanto MA, Nawara R (2017) Comparative study of the monthly global solar radiation estimation data in Jakarta. 2nd international Tropical Renew. Energy Conference. *Earth Environ Sci* 105:1–6
- Ododo JC, Sulaiman AT, Aidan J, Yuguda MM, Ogbu FA (1995) The importance of maximum air temperature in the parameterisation of solar radiation in Nigeria. *Renew Energy* 6(7):751–763
- Olatomiwa L, Mekhilef S, Shamshirband S, Mohammadi K, Petkovic D, Sudheer C (2015) A support vector machine–firefly algorithm-based model for global solar radiation prediction. *Sol Energy* 115: 632–644
- Ozoegwu CG (2018) New temperature-based models for reliable prediction of monthly mean daily global solar radiation. *J Renew Sustain Energy* 10(2):1–15
- Pedriño EC, Yamada T, Lunardi TR, de M Vieira JC (2019) Islanding detection of distributed generation by using multi-gene genetic programming based classifier. *Appl Soft Comput J* 74:206–215. <https://doi.org/10.1016/j.asoc.2018.10.016>
- Rahimikhoob A (2010) Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. *Renew Energy* 35(9):2131–2135
- Rehman S, Mohandes M (2008) Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy* 36(2):571–576
- Samadianfard S, Majnooni-Heris A, NomanQasem S, Kisi O, Shamshirband S, Chau K (2019) Daily global solar radiation modeling using data-driven techniques and empirical equations in a semi-arid climate. *Eng App Comput Fluid Mech* 13(1): 142–157
- Searson, D.P.; Leahy, D.E.; & Willis, M.J., 2010. GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. *Proceedings of the International Multi Conference of Engineers and Computer Sci.* 1:17. IMECS, Hong Kong
- Siqueira AN, Tiba C, Fraidenraich N (2010) Generation of daily solar irradiation by means of artificial neural networks. *Renew Energy* 35(11):2406–2414
- Tymvios FS, Jacovides CP, Michaelides SC, Scouteli C (2006) Comparative study of Angstrom’s and artificial neural networks’ methodologies in estimating global solar radiation. *Sol Energy* 78: 752–762
- Uncuoglu E, Laman M, Saglamer A, Kara HB (2008) Prediction of lateral effective stresses in sand using artificial neural network. *Soils Found* 48(2):141–153

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.