



Accurate total solar irradiance estimates under irradiance measurements scarcity scenarios

María Laura López · Luis E. Olcese ·
Gustavo G. Palancar · Beatriz M. Toselli

Received: 23 April 2019 / Accepted: 6 August 2019 / Published online: 15 August 2019
© Springer Nature Switzerland AG 2019

Abstract Accurate estimates of total global solar irradiance reaching the Earth's surface are relevant since routine measurements are not always available. This work aimed to determine which of the models used to estimate daily total global solar irradiance (TGSI) is the best model when irradiance measurements are scarce in a given site. A model based on an artificial neural

network (ANN) and empirical models based on temperature and sunshine measurements were analyzed and evaluated in Córdoba, Argentina. The performance of the models was benchmarked using different statistical estimators such as the mean bias error (MBE), the mean absolute bias error (MABE), the correlation coefficient (r), the Nash-Sutcliffe equation (NSE), and the statistics t test (t value). The results showed that when enough measurements were available, both the ANN and the empirical models accurately predicted TGSI (with MBE and MABE $\leq |0.11|$ and $\leq |1.98|$ kWh m⁻² day⁻¹, respectively; NSE ≥ 0.83 ; $r \geq 0.95$; and $|t$ values| $< t$ critical value). However, when few TGSI measurements were available (2, 3, 5, 7, or 10 days per month) only the ANN-based method was accurate ($|t$ value| $< t$ critical value), yielding precise results although only 2 measurements per month were available for 1 year. This model has an important advantage over the empirical models and is very relevant to Argentina due to the scarcity of TGSI measurements.

Highlights

- Performance of models to estimate total global solar irradiance (TGSI) is assessed.
- Frequency of measurements needed to calibrate models is analyzed.
- A minimum number of TGSI measurements per month are enough for ANN training.
- When TGSI measurements are scarce, empirical methods (EM) are not accurate.
- An ANN model is an overriding option with respect to the analyzed EM if TGSI measurements are scarce.

M. L. López (✉)
Facultad de Matemática, Astronomía, Física y Computación,
Universidad Nacional de Córdoba, Ciudad Universitaria,
5000 Córdoba, Argentina
e-mail: laulopez@famaf.unc.edu.ar

M. L. López
Instituto de Física Enrique Gaviola (IFEG)/CONICET, Ciudad
Universitaria, 5000 Córdoba, Argentina

L. E. Olcese · G. G. Palancar · B. M. Toselli
Facultad de Ciencias Químicas, Departamento de Físico Química,
Universidad Nacional de Córdoba, Ciudad Universitaria,
5000 Córdoba, Argentina

L. E. Olcese · G. G. Palancar · B. M. Toselli
Instituto de Investigaciones en Físicoquímica de Córdoba
(INFIQC)/CONICET/CLCM, Ciudad Universitaria,
5000 Córdoba, Argentina

Keywords Artificial neural network · Scarce measurements · Solar energy · Solar radiation estimation

Introduction

The amount of total global solar irradiance (TGSI) and its temporal distribution are important factors to determine water availability and to study the impact of changes in radiation levels on climate change, ecosystems, and economical activities. Also, these factors are

critical to optimize the design of solar energy systems (e.g., Wang et al. 2015; Widén et al. 2015). This latter use has become relevant to recent research studies since solar energy is one of the most important renewable energies.

The most accurate way to obtain solar radiation levels is by using ground-based measurements. However, due to the high cost of measuring equipment and its maintenance, a variety of methods to estimate solar radiation from different variables have been proposed and analyzed (e.g., Sonmete et al. 2011; Elani 2007; Barron et al. 2009). Since clouds are the most important meteorological factors that attenuate TGSI (e.g., Wacker et al. 2015), sunshine duration is the most direct variable to estimate TGSI. Ångström was the first to show a linear relation between the duration of bright sunshine and TGSI (Ångström 1924). Some years later this relation was modified by Prescott, who proposed the so-called Ångström-Prescott equation (Prescott 1940). Unlike Ångström equation, it introduces the use of the extraterrestrial irradiance on a horizontal surface to estimate TGSI. Different authors have introduced modifications to the original Ångström-Prescott model in order to apply it to different sites and climatic conditions. A review of the works using bright sunshine measurements for estimating TGSI can be found in Bakirci (2009a).

Although sunshine data are easier to obtain than TGSI measurements, they are not available at most of the meteorological stations. For this reason, models based only on geographical locations and easily available meteorological variables are attractive options. As in the case of sunshine hours, the difference between the daily maximum and minimum temperatures can be a good indicator of cloudiness, since cloud cover decreases the maximum air temperature (due to the reduction in transmissivity) and increases the minimum air temperature (due to cloud emissivity). Hargreaves and Samani (1982) were the first to introduce this variable to estimate TGSI through its relation with the extraterrestrial irradiance. Since then, many modifications to the Hargreaves and Samani model have been proposed. These modifications include the use of a different mathematical relation between the extraterrestrial irradiance and the maximum and minimum temperatures and the use of more meteorological variables such as pressure, relative humidity, and atmospheric precipitable water vapor. Besides the models based on sunshine measurements and meteorological data, other methods estimate

TGSI using data related to the presence of clouds such as cloud cover and cloud types. These models involve the routine observation of the sky conditions. A review of the empirical models proposed to estimate TGSI can be found in Besharat et al. (2013).

Soft computing techniques have been widely used in environmental studies since they allow both handling a large amount of data and making estimates, although the underlying physical processes involved in these estimations are not completely understood (He et al. 2014). For example, they have been used to predict the amount of precipitation (Choubin et al. 2018), to assess the risk of groundwater contamination (Sajedi-Hosseini et al. 2018) and to model meteorological drought (Rafiei-Sardooi et al. 2018). Artificial neural network (ANN) (Basheer and Hajmeer 2000) is one of the soft computing techniques commonly used to estimate global and spectral irradiance (Moreno-Sáez and Mora-López 2014; Khosravi et al. 2018; Voyant et al. 2017). One of the advantages of this tool compared with the empirical models is that the user does not need to define how each input parameter is related to the obtained value. However, once the ANN model has been trained with TGSI measurements and generalized for a particular site, its application to different sites is not straightforward. Although empirical models are simple to apply, they determine coefficients that are accurate only for the location where they were obtained. To retrieve and assess these coefficients, TGSI measurements are needed. Therefore, these measurements should be performed by covering the entire range of local meteorological conditions to be representative of the site of study.

In the literature, there are several studies focused on estimating solar radiation for a given site by using new empirical models. These models require TGSI measurements to calibrate the empirical coefficients and to validate the obtained estimations. Similarly, an ANN model also requires these measurements. For example, Quej et al. (2016) proposed an empirical model to estimate TGSI in six sites in Mexico by using, as average, TGSI measurements recorded over a 10-year period. In a later work, Quej et al. (2017) used the same set of measurements to develop an ANN model for TGSI estimations. De Souza et al. (2016) used measurements recorded over a 3-year period in a state of Brazil to calibrate a new model and measurements recorded over a 1-year period to validate it. Besharat et al. (2013) proposed a new model to estimate TGSI in a city in Iran by using TGSI measurements recorded from 1988 to

2008. In conclusion, all these works use a set of TGSIs measurements recorded on a daily basis for many years.

The choice of a particular model is not an easy task when this dataset has a limited number of data. Therefore, the present work aimed to determine which of the models can be used to estimate daily TGSIs when these measurements are scarce. Besides the ANN models, the empirical models based on temperature and sunshine measurements selected from the literature were analyzed and evaluated. A sensitivity analysis of the number of TGSIs measurements needed to obtain accurate estimations was performed for both the empirical and the ANN models. Statistical estimators were used to evaluate the accuracy of the model regarding the measurements performed in Córdoba, a city located in the center of Argentina.

Materials and methods

Data

Site of study

This work was performed in Córdoba City (31.4° S; 64.18° W; 470 m.a.s.l.), which has an estimated population of 1.3 million inhabitants (2010 National Demographic Census). It is a Mediterranean city in Argentina located in the center of the country in a semi-arid region. The climate is sub-humid and the prevailing wind direction is from the NE. The annual average precipitation is about 700 mm. However, the area is affected by severe and persistent dry periods that occur cyclically in wintertime. Snowfalls are infrequent. Thus, precipitations consist mostly of rain, which is concentrated mainly in summer. The study was carried out between 2000 and 2013. So, a wide range of meteorological conditions were covered during that period.

Meteorological data

Minimum, maximum, and mean ambient temperatures, relative humidity, wind speed, atmospheric pressure, and sunshine hours were provided by a National Weather Service station located near the irradiance measurement site. Sunshine hour data, also provided by the National Weather Service, were measured using a Campbell-Stokes heliograph. All of these meteorological data were obtained on a daily timescale for all the

days between 2000 and 2013. For these days, solar radiation measurements are available (see the “[Solar radiation measurements](#)” section).

Solar radiation measurements

Total global solar irradiance (TGSIs) surface measurements (300–3000 nm) were obtained with a pyranometer YES (Yankee Environmental System, Inc.) model TSP-700 operated by our research team. The pyranometer was placed in a wide-open area on the University Campus in Córdoba. Broadband observations were recorded as half a minute average values throughout the day. These measurements were integrated throughout the day following the trapezoidal rule method in order to obtain the daily TGSIs values. It should be mentioned that the TGSIs station that provided the values is not an automatic station. So, the pyranometer was mounted every day at the mentioned site early in the morning and removed at night. The measurements and data from the heliograph were manually checked for errors and inconsistencies and quality control was performed to eliminate inaccurate measurements. Finally, a total of 696 days were used. TGSIs measurements recorded during these days represent a scenario with high data availability if a model is to be calibrated and the accuracy of its TGSIs estimates is to be evaluated. Therefore, to simulate scenarios with limited data availability, only a small number of these TGSIs measurements were selected to calibrate the models. This procedure will be carefully detailed in the “Measured vs. estimated irradiance: large number of available TGSIs measurements” and “Measured vs. estimated irradiance: scarce TGSIs measurements” sections.

Figure 1 shows the location of Córdoba City in the center of Argentina, as well as a map of the city showing the location of the meteorological station and the irradiance measurement site. As can be seen, both sites are approximately 2.5 km away and are located at the same elevation.

Empirical models

The most commonly used empirical models based on sunshine duration and meteorological data were selected. Models with fixed coefficients were discarded. Following this criteria, fifteen models were selected to obtain estimated TGSIs values. All of the selected models use extraterrestrial irradiance (H_0), which was

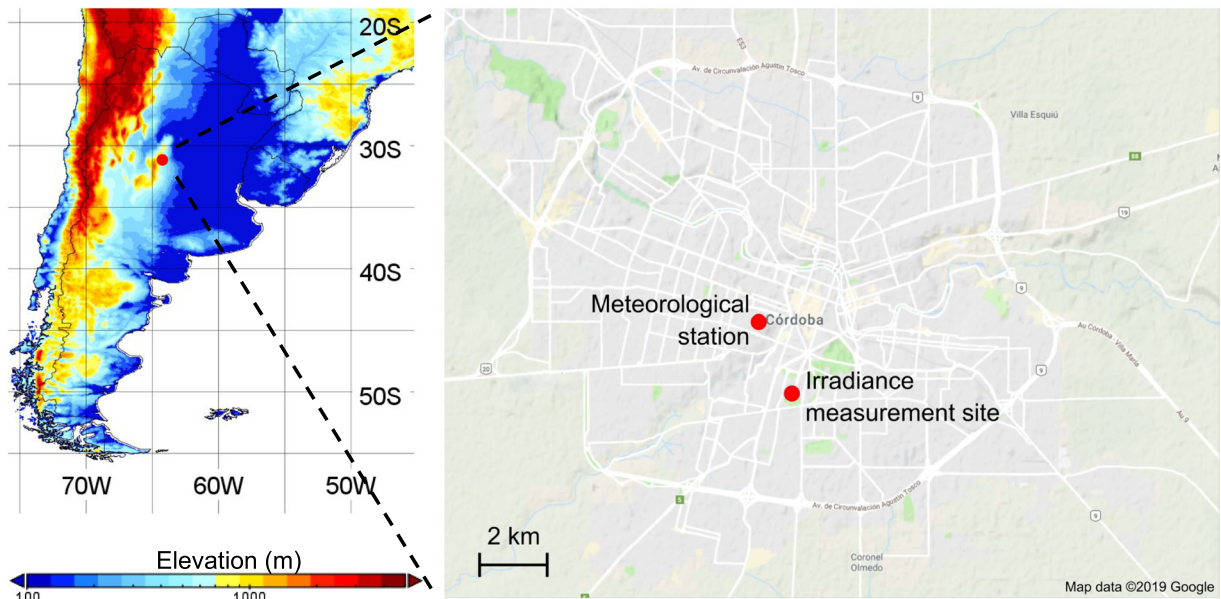


Fig. 1 Location of Córdoba City on a shaded relief map. On the right, the location of the meteorological station and the irradiance measurement site in the city

calculated from the following widely used equation (e.g., Duffie and Beckman 1991):

$$H_o = \frac{24I_{SC}}{\pi} \left[1 + 0.033 \cos\left(\frac{360n_o}{365}\right) \right] \left[\cos\varphi \cos\delta \sin w_s + \left(\frac{2\pi w_s}{360}\right) \sin\varphi \sin\delta \right] \tag{1}$$

where I_{SC} is the solar constant (1367 W m^{-2}); n_o is the Julian day; φ is the latitude of the site (in degrees); δ is the solar declination angle (in degrees); and w_s is the sunset hour angle (in degrees). The two latter variables were calculated as follow:

$$\delta = 23.45 \sin[360(n_o + 284)/365] \tag{2}$$

$$w_s = \cos^{-1}(-\tan\varphi \tan\delta) \tag{3}$$

Figure 2 shows the annual evolution of the measured TGSI and the extraterrestrial irradiance. The effect of clouds on TGSI measurements was evident on days with lower than expected TGSI values.

The models used in this work are listed in Table 1. In this table, H_e is the estimated total global surface solar irradiance; n is the day length; N is the maximum possible sunshine duration; a , b , c , and d are the empirical coefficients; T_{max} and T_{min} are the maximum and minimum daily temperature, respectively ($^{\circ}\text{C}$); Kr is an empirical coefficient; P is the mean atmospheric

pressure of the site (hPa); i and $(i + 1)$ represent the day of interest and the next day, respectively; T_{avg} is the average temperature (in $^{\circ}\text{C}$); $f(T_{avg})$ and $f(T_{min})$ are both functions of T_{avg} and T_{min} , respectively. For all the models used in this work, n was obtained from the heliograph, while N was calculated as:

$$N = 2w_s/15 \tag{4}$$

Artificial neural network

ANNs have been designed to mimic the human brain and are structurally formed by processing units called neurons. In a typical ANN, each input variable is jointly connected to neurons placed in one or several intermediate layers, which are then connected to the outputs. The weight of each individual connection is a parameter adjusted during the ANN training. The training process is iterative. In each loop, the error between the known and the calculated output is evaluated, and consequently, the connection weights are modified until the network can correctly predict the already known outputs.

For the present work, and given that the ANNs are trained in a semi-automatic way, all the possible combinations among the following meteorological factors were evaluated: T_{max} , T_{min} , $T_{min}(i + 1)$, $T_{max}(i + 1)$, P_{max} , P_{min} , HR_{max} , HR_{min} , P_{month} , and T_{month} . By analogy

Fig. 2 Annual variation of daily TGSI and daily extraterrestrial irradiance

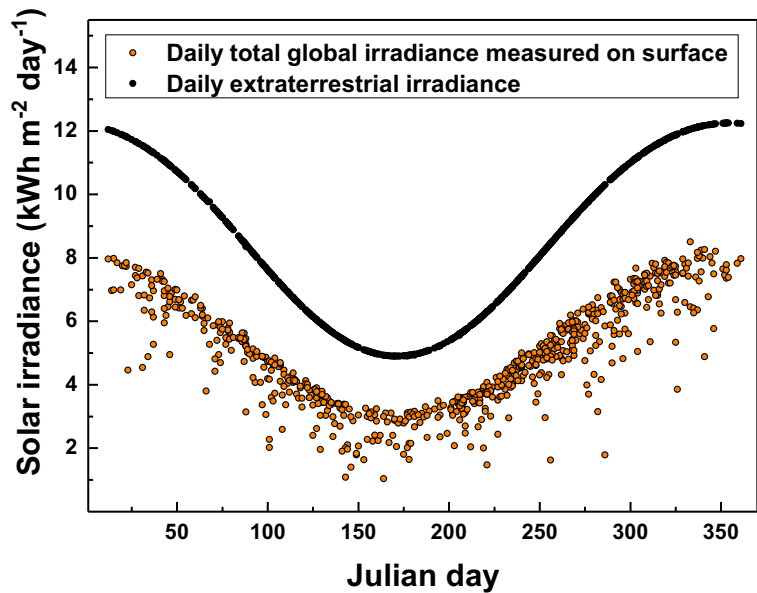


Table 1 Equations of the models used to estimate TGSI. H_e is the estimated total global surface solar irradiance; n is the day length; N is the maximum possible sunshine duration; $a, b, c, d,$ and Kr are the empirical coefficients; T_{max} and T_{min} are the maximum and

minimum daily temperature, respectively ($^{\circ}C$); P is the mean atmospheric pressure of the site (hPa); i and $(i + 1)$ are the day of interest and the next day, respectively; T_{avg} is the average temperature ($^{\circ}C$)

Model	Equation
Ångström-Prescott (Ångström 1924; Prescott 1940)	$H_e = H_o[a + b(n/N)]$
Akinoglu and Ecevit (Akinoglu and Ecevit 1990)	$H_e = H_o[a + b(n/N) + c(n/N)^2]$
Samuel (Samuel 1991)	$H_e = H_o[a + b(n/N) + c(n/N)^2 + d(n/N)^3]$
Newland (Newland 1988)	$H_e = H_o[a + b(n/N) + c \log(n/N)]$
Ampratwum and Dorvlo (Ampratwum and Dorvlo 1999)	$H_e = H_o[a + b \log(n/N)]$
Bakirci linear-exponential (Bakirci 2009b)	$H_e = H_o[a + b(n/N) + c \exp(n/N)]$
Bakirci exponent (Bakirci 2009b)	$H_e = H_o[a(n/N)^b]$
Almorox et al. exponential (Almorox et al. 2005)	$H_e = H_o[a + b \exp(n/N)]$
Hargreaves and Samani (Hargreaves and Samani 1982)	$H_e = H_o [(Kr\sqrt{T_{max}-T_{min}})]$
Allen (Allen 1997)	$H_e = H_o [(a\sqrt{(P/1013)}\sqrt{T_{max}-T_{min}})]$
Samani (Samani 2000)	$H_e = H_o[(a(T_{max} - T_{min})^2 + b(T_{max} - T_{min}) + c)(T_{max} - T_{min})]$
Bristow and Campbell (Bristow and Campbell 1984)	$H_e = H_o a [1 - \exp(-b\Delta\Delta^c)]$ $\Delta T = T_{max(i)} - 0.5(T_{min(i)} - T_{min(i+1)})$
Donatelli and Campbell (Donatelli and Campbell 1998)	$H_e = H_o a [1 - \exp(-b(f(T_{avg})\Delta\Delta^2 f(T_{min})))]$ $T_{avg} = \frac{T_{max} - T_{min}}{2}$ $f(T_{avg}) = 0.017 \exp[\exp(-0.053T_{avg})]$ $f(T_{min}) = \exp(T_{min}/c)$
Chen et al. (Chen et al. 2004)	$H_e = H_o [a(\sqrt{T_{max}-T_{min}}) + b]$
Almorox et al. meteorological (Almorox et al. 2013)	$H_e = aT_{max} + bT_{min} + cH_o + d$

with the nomenclature previously used, $T_{\max(i+1)}$ symbolizes the maximum temperature of the next day; P_{\max} and P_{\min} represent, respectively, the maximum and minimum atmospheric pressure of the i day; and HR_{\max} and HR_{\min} are, respectively, the maximum and minimum relative humidity of the i day. Finally, P_{month} and T_{month} are the monthly mean atmospheric pressure and temperature, respectively. Besides H_o , a certain number, from one to all, of these meteorological parameters were used as input parameters to estimate TGSI, resulting in 1024 possible combinations. Each of these combinations was separately trained with a feed-forward network using the Levenberg-Marquardt back-propagation training algorithm (Marquardt 1963; Hagan and Menhaj 1994) in order to find which of the combinations yielded the best results. In this network, the first layer has a connection from the network input, then a single hidden layer has a connection from the previous layer and the final layer produces the network's output. Also, for each combination of variables, several ANNs were trained using a different number of neurons in the hidden layer in order to find the best network architecture for each data set. The number of neurons of the intermediate layer was set to 3, 5, or 10, while the transfer function (i.e., how the input and output of the neurons are correlated) was set as tan-sigmoid or linear, resulting in 6144 different network configurations.

The results presented here were obtained by using the best network selected among all the possible variable combinations and architectures. This network was chosen in terms of the average root mean square error between the irradiance values obtained using the ANN model and the measured values considering all the subsets described in the "Results and discussion" section. All the ANN calculations were done using the Matlab® Neural Network Toolbox (version 2011Rb). Among all the 6144 configurations employed (different inputs and different numbers of neurons in the hidden layer), the selected ANN configuration used the following input parameters: T_{\min} , P_{\min} , HR_{\min} , $T_{\min(i+1)}$, P_{month} , and T_{month} , 10 neurons in the hidden layer and tan-sigmoid activation function. This configuration, hereinafter referred to just as ANN configuration, was the only one used in this work.

To better understand the results obtained, it is worth mentioning some aspects of the working process to estimate TGSI. The common approach to estimate TGSI using empirical models is to split the measurements into two sub-datasets. The first one is

used to determine the empirical coefficients of the models, while the second one is used to evaluate the accuracy of the models by comparing the estimated with measured TGSI. Both groups are frequently referred to as calibration and validation group, respectively (Besharat et al. 2013). However, the use of an ANN also requires a validation stage, but as a part of the training process. Hereinafter, the term "evaluation" will be used to refer to the process that analyzes the performance of the models by considering the group of data not included in the determination of the coefficients of the models or in the ANN training and by comparing the estimates with the measurements.

Comparison techniques

To compare the estimated with measured TGSI, the recommendations by Yorukoglu and Celik (2006) were followed. To evaluate the accuracy and performance of the derived correlations (in predicting TGSI), the statistical estimators shown in Table 2 were used. In this table, $H_{i,e}$ and $H_{i,m}$ are, respectively, the measured and estimated TGSI for the i day; n_d is the number of data; $H_{a,m}$ and $H_{a,e}$ are the average of the measured and estimated TGSI, respectively, and (n_d-1) are the degrees of freedom.

The statistics t test was performed to determine the existence of significant differences between the estimated TGSI following each of the models and the TGSI measurements. The t value obtained from the expression in Table 2 (e.g., Muzathik et al. 2011) was compared with a critical t value (t_{crit}) obtained from standard statistical tables for a given level of significance (α) and (n_d-1) degrees of freedom (e.g., Walpole et al. 2012). In this work, the t table was used for a two-sided test using a value of 0.05 for α . Thus, the null hypothesis is accepted if:

$$-t_{\text{crit}_{\frac{\alpha}{2}}(n_d-1)} < t < t_{\text{crit}_{\frac{\alpha}{2}}(n_d-1)} \quad (5)$$

The results showed that there were not significant differences between the TGSI values obtained using the model and the measured TGSI.

Table 2 Equations of the statistical estimators used to compare the estimated with measured TGSI. $H_{i,e}$ and $H_{i,m}$ are the measured and estimated TGSI for the i day, respectively; n_d is the number of data; $H_{a,m}$ and $H_{a,e}$ are the average of the measured and estimated TGSI, respectively; (n_d-1) are the degrees of freedom

Statistical estimator	Equation
Mean bias error (MBE)	$MBE = \frac{1}{n_d} \sum_{i=1}^{n_d} (H_{i,e} - H_{i,m})$
Mean percentage error (MPE)	$MPE = \frac{100}{n_d} \sum_{i=1}^{n_d} \left(\frac{H_{i,m} - H_{i,e}}{H_{i,m}} \right)$
Mean absolute bias error (MABE)	$MABE = \frac{1}{n_d} \sum_{i=1}^{n_d} (H_{i,m} - H_{i,e})$
Mean absolute percentage error (MAPE)	$MAPE = \frac{100}{n_d} \sum_{i=1}^{n_d} \left(\frac{ H_{i,m} - H_{i,e} }{H_{i,m}} \right)$
Root mean square error (RMSE)	$RMSE = \left[\frac{1}{n_d} \sum_{i=1}^{n_d} (H_{i,m} - H_{i,e})^2 \right]^{\frac{1}{2}}$
Correlation coefficient (r)	$r = \left[\frac{SS_{em}}{\sqrt{S_e} \sqrt{S_m}} \right]^{\frac{1}{2}}, \text{ where:}$ $SS_{em} = \sum_{i=1}^{n_d} (H_{i,m} - H_{a,m})(H_{i,e} - H_{a,e})$ $S_e = \sum_{i=1}^{n_d} (H_{a,e} - H_{i,e})^2; S_m = \sum_{i=1}^{n_d} (H_{a,m} - H_{i,m})^2$ $H_{a,m} = \frac{1}{n_d} \sum_{i=1}^{n_d} H_{i,m}; H_{a,e} = \frac{1}{n_d} \sum_{i=1}^{n_d} H_{i,e}$
Nash-Sutcliffe equation (NSE)	$NSE = 1 - \frac{\sum_{i=1}^{n_d} (H_{i,m} - H_{i,e})^2}{\sum_{i=1}^{n_d} (H_{i,m} - H_{a,m})^2}$
Statistics t test	$t = \left[\frac{(n_d-1)(MBE)^2}{(RMSE)^2 - (MBE)^2} \right]^{\frac{1}{2}}$

Results and discussion

The “Measured vs. estimated irradiance: large number of available TGSI measurements” section discusses the comparison between the measured and estimated TGSI when 85% of the TGSI measurements were used to calibrate the empirical models and train ANN. The results obtained showed that the models were adequate to estimate TGSI when a large number of TGSI measurements were available, which is not the most common scenario.

The accuracy of the models was analyzed using the statistical estimators described in the “Comparison

techniques” section. The empirical coefficients obtained for each model for Córdoba City were analyzed and compared with those previously reported for other sites.

The “Measured vs. estimated irradiance: scarce TGSI measurements” section describes a sensibility analysis performed to determine the number of daily TGSI measurements needed to calibrate the empirical models and ANNs. Conversely to the “Measured vs. estimated irradiance: large number of available TGSI measurements” section, the results described in this section are useful in scenarios where TGSI measurements are scarce.

Measured vs. estimated irradiance: large number of available TGSI measurements

To calibrate the empirical models, train ANN, and then evaluate all of them, the measured TGSI was divided into two groups. For the calibration and training stage, the set of daily TGSI measurements was randomly selected. A number of measurements corresponding to 85% of the total data were included in this calibration/training group. The remaining 15% was formed by the evaluation group.

Table 3 shows the empirical coefficients obtained for each model in the calibration stage.

Figure 3 shows the relation between the estimated and measured TGSI for all the analyzed methods applied to the evaluation data set. Red lines indicate the 1:1 relation.

In order to select the methods in good agreement with the measurements, several statistical estimators were analyzed. Table 4 shows the statistical parameters obtained for each method. For a higher modeling accuracy, MBE, MPE, MABE, MAPE, and RMSE values should be close to zero, while r and NSE coefficients should be close to one.

The values of these parameters show the agreement between the estimated and measured TGSI (Table 4). These results can also be qualitatively inferred from a visual inspection of Fig. 3. As can be seen, for all the models, except for the Bristow and Campbell model, the calculated t values were less than t_{crit} (1.983 at $\alpha = 5\%$). Thus, excluding the Bristow and Campbell model and considering that some models were more precise than others, it can be concluded that there were no significant

differences between the TGSI values obtained using all the evaluated and measured models for the site of study.

The empirical coefficients previously shown in Table 3 were compared with those obtained for different sites and consequently for different meteorological conditions. The obtained coefficients were in the range of values previously reported (e.g., Muzathik et al. 2011, reported $a = 0.2207$ and $b = 0.5249$ for Ångström-Prescott model in Malaysia, Almorox et al. 2013, reported $a = 0.7685$, $b = -0.0714$, and $c = 1.0919$ for Bristow and Campbell model in Argentina and Besharat et al. 2013, reported $Kr = 0.1746$ in Iran).

From all these models, the Almorox et al. meteorological model was chosen because it was developed and evaluated in Cañada de Luque City, which is approximately 100 km from Córdoba City. Because of its similar topography and proximity to Córdoba City, the empirical coefficients for the Almorox et al. (2013) meteorological model were expected to be approximately the same for both cities. Considering this hypothesis, the empirical coefficients obtained by Almorox et al. (2013) for Cañada de Luque were used to estimate the TGSI values for Córdoba City. Table 5 shows the empirical coefficients reported for Cañada de Luque (Almorox et al. 2013) and the t and p values calculated from the comparison between the estimated and measured TGSI for Córdoba City.

In conclusion, despite the closeness of Cañada de Luque to Córdoba city, the statistical comparison resulted in a t value = 4.501. This value was higher than the critical t value, indicating the high spatial variability of the empirical coefficients. Therefore, the coefficients obtained for a

Table 3 Adjusted coefficients for all the empirical models

Model	Kr	a	b	c	d
Ångström-Prescott		0.290	0.388		
Akinoglu and Ecevit		0.185	0.770	- 0.303	
Samuel		0.210	0.553	0.143	- 0.263
Newland		0.383	0.286	0.047	
Ampratwum and Dorvlo		0.637	0.144		
Bakirci lineal-exponential		0.661	0.404		
Bakirci exponent		0.535	1.044	- 0.343	
Almorox et al. exponential		0.163	0.194		
Hargreaves and Samani	0.158				
Allen		0.158			
Samani		- 0.00007	- 0.00147	0.198	
Bristow and Campbell		0.618	0.016	2.039	
Donatelli and Campbell		0.618	0.465	7.955	
Chen et al.		0.067	0.344		
Almorox meteorological		0.244	- 0.304	0.698	- 6.155

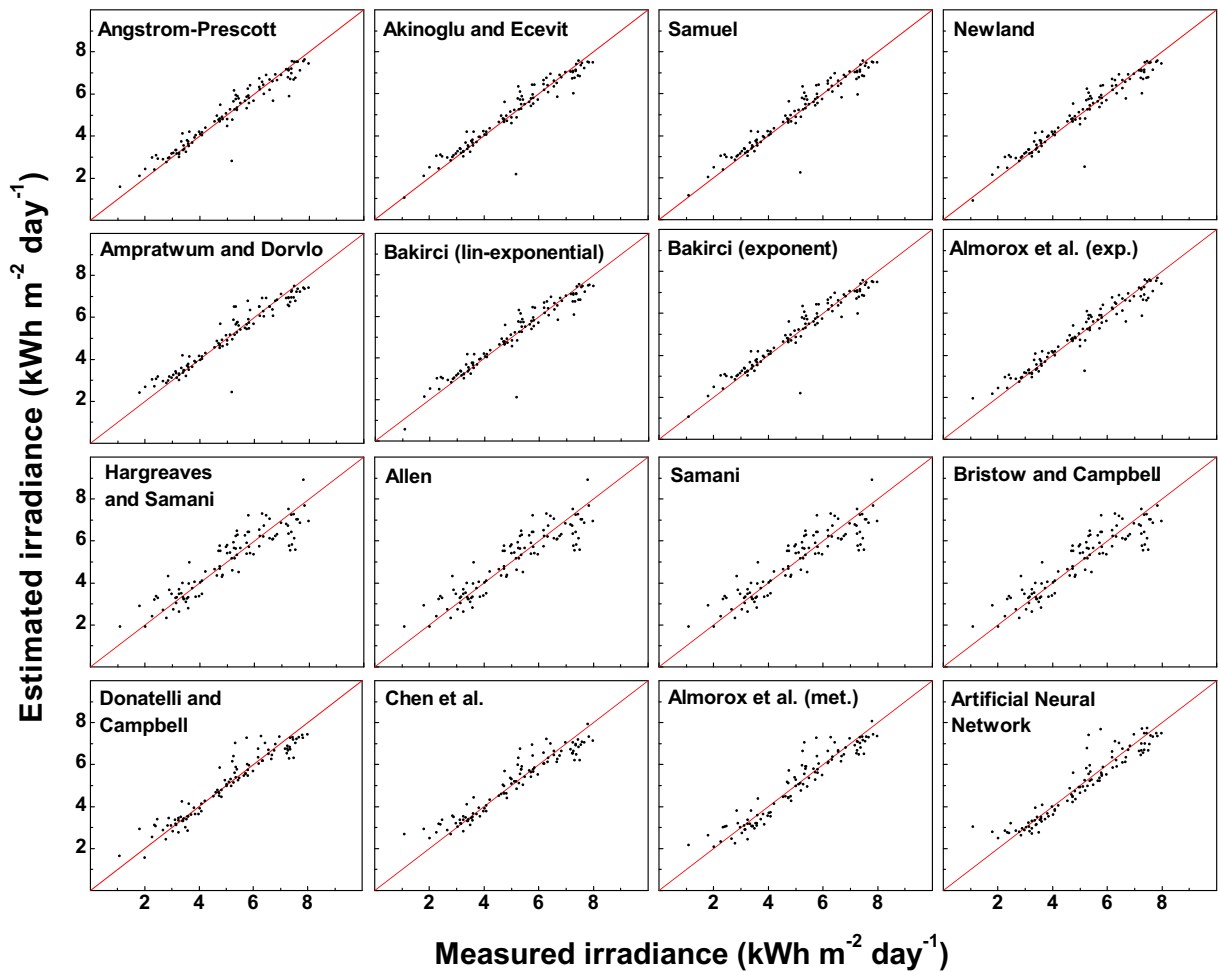


Fig. 3 Relation between the measured and estimated TGSi for the evaluation data set

given site should not be blindly applied to a different site, no matter how close the site is. This spatial dependence on the empirical coefficients has been previously reported in the literature and led to studies of spatial sensitivity of the models. For example, Liu et al. (2014) focused on the spatial modeling of the parameters of the Ångström-Prescott model in order to extend its application to different sites in China. Urraca et al. (2017) also analyzed the spatial performance of the TGSi estimates of a model in Spain.

Measured vs. estimated irradiance: scarce TGSi measurements

Given that empirical coefficients should be obtained from local TGSi measurements, an analysis was performed to determine the minimum number of TGSi measurements needed to calibrate the empirical models or to train the ANN models. This analysis was carried out by following the same

procedure for the two stages of calibration and training (for the empirical and ANN models, respectively) described in the “Measured vs. estimated irradiance: large number of available TGSi measurements” section and by subsequent evaluation. However, different subsets with different numbers of TGSi measurements were tested. The subsets of TGSi measurements were randomly selected. However, in order to consider all the seasons and meteorological conditions, all months had the same number of days. Subsets containing 2, 3, 5, 7, and 10 days per month were formed. That is to say, the numbers of days used were 24, 36, 60, 84, and 120. These numbers were randomly selected and divided into 85% for calibration and 15% for validation processes. In this analysis, the days in a given month came from different years (these days were also randomly selected). For these different groups, the calibration/validation processes were as in the “Measured vs. estimated irradiance: large number of available TGSi measurements” section (the

Table 4 Statistical estimators for each of the models (* = [kWh m⁻² d⁻¹])

Model	MBE*	MPE	MABE*	MAPE	RMSE*	r	NSE	<i>t</i> value	<i>p</i> value
Ångström-Prescott	-0.04	-1.89	1.05	6.58	1.57	0.983	0.933	0.251	0.802
Akinoglu and Ecevit	-0.07	-1.77	1.06	6.40	1.69	0.980	0.922	0.411	0.682
Samuel	-0.09	-1.94	1.06	6.38	1.68	0.981	0.924	0.517	0.606
Newland	-0.03	-1.44	1.05	6.43	1.61	0.982	0.930	0.204	0.839
Ampratwum and Dorvlo	-0.01	-0.89	1.18	8.13	1.77	0.978	0.915	0.083	0.934
Bakirci linear-exponential	-0.08	-1.83	1.06	6.35	1.69	0.980	0.923	0.452	0.652
Bakirci exponent	-0.03	-1.23	1.08	6.88	1.71	0.980	0.921	0.186	0.853
Almorox et al. exponential	-0.02	-2.00	1.06	6.89	1.52	0.984	0.937	0.113	0.910
Hargreaves and Samani	-0.03	-2.68	1.96	12.15	2.46	0.956	0.834	0.122	0.903
Allen	-0.03	-2.74	1.96	12.18	2.47	0.955	0.833	0.142	0.887
Samani	0.11	-2.66	1.53	10.27	2.10	0.971	0.879	1.430	0.878
Bristow and Campbell	-0.03	-2.77	1.98	12.30	2.39	0.961	0.843	2.023	0.021
Donatelli and Campbell	-0.04	-1.92	1.30	8.45	1.77	0.978	0.915	0.236	0.814
Chen et al.	-0.09	-3.68	1.41	9.76	1.87	0.977	0.905	0.484	0.630
Almorox et al. meteorological	0.08	-0.92	1.44	9.68	1.84	0.977	0.908	0.417	0.678
ANN	-0.02	-2.90	1.58	11.22	1.26	0.963	0.859	0.184	0.854

weights of the ANN model when 2, 3, 5, 7, and 10 days per month were used for the calibration/validation processes are shown in Appendix). These values were used to estimate TGSI and to compare them with the TGSI measurements included in the evaluation data set (all the available measurements). All the statistical estimators were calculated. However, only the analysis of the *t* values is shown. Considering the amount of data in the evaluation group, t_{crit} was 1.963 ($\alpha = 5\%$).

Figure 4 shows the *t* values obtained for each model when different numbers of days per month were used in the subset. Red horizontal lines indicate the $\pm t_{crit}$ values.

As can be seen, all the empirical models presented some *t* values higher than t_{crit} . However, all the *t* values for the ANN models were lower than t_{crit} . This behavior shows that 2, 3, 5, 7, and 10 days per month are enough to estimate TGSI.

In the case of the empirical models, the Samuel, Allen, Samani, Chen et al. and Almorox et al. meteorological models were adequate to estimate daily TGSI when almost 600 measurements of TGSI were available for calibration

(the “Measured vs. estimated irradiance: large number of available TGSI measurements” section). However, when the calibration was carried out using a subset of a limited number of TGSI measurements per month, they presented *t* values higher than t_{crit} (see Fig. 4). This fact shows the strong relationship between the number of calibration days and the accuracy of the models, since more days were needed to obtain the convergence of the empirical coefficients.

The increase in the number of days used for calibration and training was expected to lead to *t* values closer to zero. This behavior was effectively verified for the ANN models, but not for any of the empirical models. This shows the improvement in the performance of the ANN models as the amount of data used for training was increased (Table 4 shows that the *t* value was 0.184 for about 600 days in the training group).

The innovative character of this research does not lie in the fact that the performance of the ANN model is better than that of the empirical models. In fact, this better performance has been widely demonstrated in previous works

Table 5 Empirical coefficients of Almorox et al. meteorological model reported for Cañada de Luque (Almorox et al. 2013) and Córdoba (this work) and the *t* and *p* values for Córdoba City

Site of study	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>t</i> value	<i>p</i> value
Cañada de Luque (Almorox et al. 2013)	0.62	-0.66	0.62	-11.07	4.501	< 0.0001
Córdoba (this work)	0.244	-0.304	0.698	-6.155	0.417	0.678

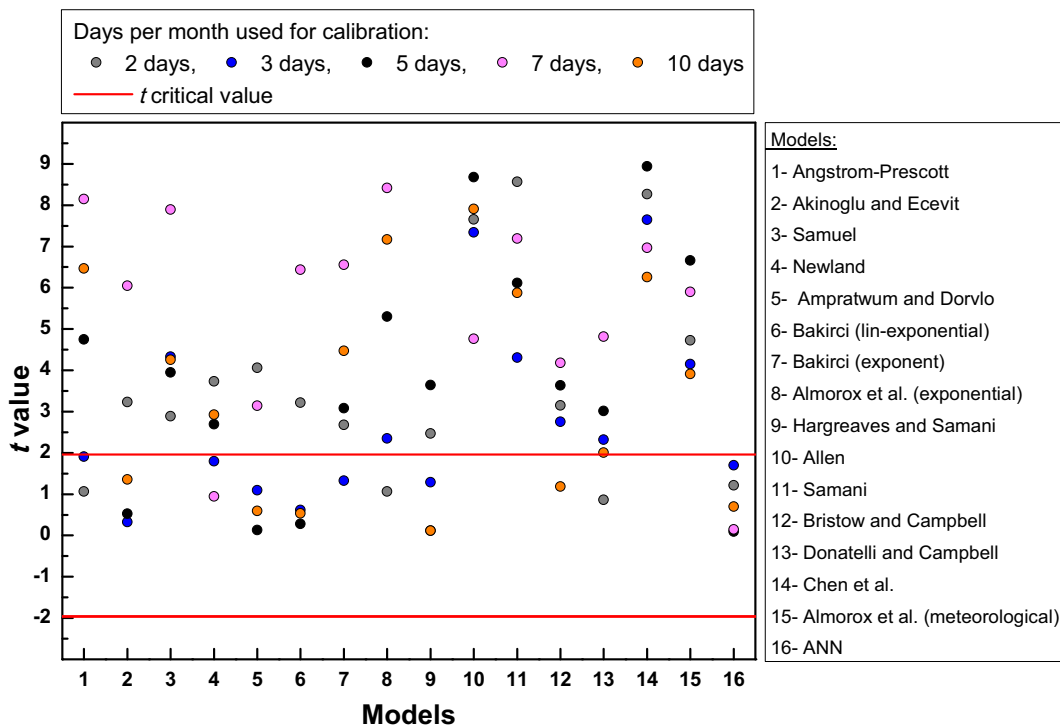


Fig. 4 t values calculated by using a different number of daily TGSI measurements per month to calibrate the empirical models and to train the ANN models

(e.g., Sharifi et al. 2016; Citakoglu 2015; Antonopoulos et al. 2019). The innovative nature of this study lies in the fact that the daily TGSI measurements performed for only 10 days per month (or even less) for 1 year were enough to train the ANN models and to obtain estimates statistically equal to the measurements for the entire period. This is an important result for sites like Argentina where these measurements are scarce. However, the main disadvantage of this method is that its application is limited to sites where TGSI measurements are recorded.

Lastly, the results obtained in this work can also be used to complement studies analyzing the spatial sensitivity of the models. For example, Urraca et al. (2017) evaluated different models in relation to the spatial performance at nearly forty ground stations in central Spain from 2001 to 2013. They focused on studying the spatial variability of irradiance measurements and on obtaining irradiance estimates by means of interpolation techniques and found that the accuracy of interpolation at each location depends on the distance of the closest pyranometer. If the determination of the distance between the pyranometers needed to cover a certain area is complemented by the determination of the minimum number of TGSI measurements needed to obtain accurate estimates, it would constitute an important

advance in optimizing the resources and, at the same time, in generating more accurate solar radiation maps.

Summary and concluding remarks

This work analyzed the performance of the empirical and ANN models in order to estimate TGSI in Córdoba City between the years 2000 and 2013. From the statistical analysis, it was found that all the analyzed empirical and ANN models can be used to estimate TGSI in Córdoba City with a good accuracy provided that a large number of measurements are available to calibrate these models. However, sites without TGSI estimates do not always meet this requirement. In addition, although the calibration/training and evaluation stages for a given site are performed using the appropriate number of measurements, the results cannot be applied to other nearby sites. This expected result was statistically shown when the empirical coefficients obtained in the present work for the Almorox et al. meteorological model were compared to those previously reported in the literature for a site located almost 100 km away.

If the spatial closeness between two sites does not guarantee accurate TGSI estimates by using the same empirical coefficients, then TGSI measurements are required to calibrate the models. This fact led us to investigate the minimum number of daily TGSI measurements required to calibrate/train each model. The number of the daily TGSI measurements per month used to calibrate the empirical models and to train the ANN configurations was varied. From the analysis of the results, it was concluded that the ANN model is the only model that with 10 measurements available per month for 1 year (or even less) can be used to obtain reliable estimates of daily TGSI. This represents an important advantage over the empirical models because they can be easily applied. However, the main weakness of this method is that its application is still limited to sites where TGSI measurements are recorded.

Acknowledgments We thank Secretaría de Ciencia y Tecnología de la Universidad Nacional de Córdoba (UNC), Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), and Agencia Nacional de Promoción Científica (FONCYT) for their support.

Appendix

The tables presented in this section show the weights found for the ANN model when 2, 3, 5, 7, and 10 days per month were used for the calibration/training processes (Tables 6, 7, 8, 9, and 10, respectively).

Table 6 Weights of the ANN model when 2 days per month were used for the calibration/validation processes

Neuron	Hidden layer									Output layer	
	Weights									Bias	Weights
1	-0.483	-0.591	-0.622	-0.423	0.275	-0.842	-0.830	-0.102	1.581	-0.374	0.275
2	1.282	0.841	-0.048	-0.055	0.482	0.133	0.697	-0.808	-1.421	0.338	
3	-0.460	0.659	-0.087	0.101	0.921	-1.271	0.996	-0.068	1.220	0.627	
4	0.137	1.068	-0.178	1.067	0.233	0.543	-0.884	-0.086	-0.659	0.139	
5	-0.721	0.637	-0.845	-0.876	0.803	0.398	0.319	-0.440	0.107	-0.358	
6	0.928	1.029	0.397	0.309	0.294	-0.052	0.562	-0.138	-0.103	0.250	
7	-0.115	-0.834	1.192	-0.102	-0.070	-0.579	-0.228	0.341	-0.624	-0.511	
8	0.842	-0.224	-0.225	0.044	0.792	-0.846	-0.058	-1.104	0.859	-0.169	
9	-0.896	0.452	-0.504	-1.110	0.124	0.275	-0.963	0.291	-1.197	0.861	
10	-0.804	-0.219	0.585	0.719	-0.884	-0.788	-0.432	0.448	-1.974	-0.511	

Table 7 Weights of the ANN model when 3 days per month were used for the calibration/validation processes

Neuron	Hidden layer									Output layer	
	Weights									Bias	Weights
1	0.670	-0.118	0.044	0.454	-0.642	1.231	-1.012	0.262	-2.542	0.974	0.364
2	0.401	-0.524	-0.402	-0.283	-1.069	0.413	0.520	1.254	-1.705	-0.553	
3	0.305	-1.016	0.765	-0.132	-0.553	0.704	0.717	-0.278	-1.209	0.443	
4	-0.980	-0.327	-0.267	0.308	-1.377	0.200	-0.579	-0.266	-0.099	-0.469	
5	-1.139	-0.078	-0.205	1.888	-0.553	0.246	0.515	0.436	0.430	0.225	
6	-1.111	0.077	0.146	-0.097	-0.455	-0.001	0.134	-1.289	-1.165	-0.236	
7	0.806	0.728	0.197	0.261	-0.230	-0.343	-0.005	-1.242	0.864	0.516	
8	-0.157	0.844	0.119	-0.447	1.128	0.784	1.265	-0.474	-1.536	-0.849	
9	0.088	0.105	0.461	-0.758	0.604	-0.616	-1.123	0.922	1.197	0.217	
10	-0.165	1.334	-0.247	0.178	-1.266	-0.226	0.564	0.334	-2.193	0.973	

Table 8 Weights of the ANN model when 5 days per month were used for the calibration/validation processes

Neuron	Hidden layer									Output layer	
	Weights									Bias	Weights
1	1.588	1.623	- 1.928	- 0.261	2.596	1.495	2.363	1.449	- 3.358	0.991	0.499
2	1.930	- 1.319	1.234	0.773	- 0.371	0.901	- 0.521	- 2.380	- 0.535	0.149	
3	- 0.452	0.838	1.021	0.157	2.151	- 1.552	- 1.115	- 1.257	- 0.620	- 0.717	
4	2.997	1.517	0.698	- 0.498	0.050	2.429	0.094	- 0.495	- 1.732	- 0.266	
5	- 0.740	- 0.910	- 0.439	1.326	1.432	- 2.416	0.599	- 1.010	- 0.134	0.446	
6	- 1.468	- 0.719	- 1.020	0.503	- 0.796	1.665	- 1.420	- 0.532	- 1.510	- 0.328	
7	- 2.061	- 2.252	- 0.193	1.152	- 3.178	0.711	0.952	- 0.703	- 1.020	- 0.674	
8	- 1.457	- 0.933	2.890	- 1.382	2.124	0.970	2.243	1.142	- 1.224	0.118	
9	- 0.732	- 0.761	0.969	- 0.003	2.728	- 1.587	1.499	- 0.046	0.785	- 0.523	
10	- 0.514	- 0.062	- 0.896	- 0.552	- 0.164	- 0.169	- 0.395	0.536	- 2.857	0.113	

Table 9 Weights of the ANN model when 7 days per month were used for the calibration/validation processes

Neuron	Hidden layer									Output layer	
	Weights									Bias	Weights
1	- 0.824	- 0.032	- 2.556	2.765	0.402	- 0.458	- 1.734	0.518	- 2.925	0.341	- 0.254
2	0.321	1.897	- 2.115	0.570	0.248	- 0.894	1.467	0.022	0.970	0.216	
3	- 1.114	- 0.324	- 0.335	3.055	- 0.386	0.141	0.513	0.187	0.532	0.025	
4	- 0.768	- 0.141	- 2.481	- 3.842	- 2.073	- 1.583	2.207	- 1.034	- 0.102	0.263	
5	0.390	- 1.146	- 1.435	- 0.351	- 1.125	0.133	- 0.398	0.830	- 0.671	- 0.218	
6	0.412	- 1.155	- 0.462	- 0.663	- 0.644	- 0.127	0.954	- 0.155	- 0.161	- 0.807	
7	- 0.451	- 1.574	- 0.150	0.365	2.796	- 0.563	- 0.964	- 0.629	0.676	- 0.257	
8	0.165	- 0.696	0.870	3.610	3.421	- 0.309	1.433	0.979	- 1.893	- 0.178	
9	0.166	1.391	- 2.741	- 0.683	0.193	- 0.204	0.480	0.229	0.838	- 0.118	
10	- 1.158	0.884	1.471	0.836	0.280	1.170	- 0.801	- 1.806	- 1.163	- 0.479	

Table 10 Weights of the ANN model when 10 days per month were used for the calibration/validation processes

Neuron	Hidden layer									Output layer	
	Weights									Bias	Weights
1	1.639	- 0.010	- 0.192	- 0.023	0.6450	- 0.201	- 1.190	- 1.399	- 1.858	0.634	- 0.648
2	- 0.061	- 0.415	- 0.166	0.726	- 1.725	- 0.526	3.106	- 1.849	0.263	- 0.084	
3	0.288	- 1.356	0.117	0.474	0.722	0.463	- 0.165	- 1.130	- 0.311	- 0.042	
4	- 0.429	- 0.153	- 0.100	- 0.568	0.290	- 0.152	- 0.197	- 0.036	0.361	- 0.888	
5	0.109	- 0.519	1.975	- 0.814	1.867	0.196	0.163	- 0.558	- 0.513	- 0.094	
6	- 0.046	- 0.646	1.160	1.111	0.165	- 0.205	- 0.570	- 0.862	1.391	0.112	
7	0.024	- 1.608	1.063	0.580	- 0.497	- 0.834	2.112	1.915	1.219	0.361	
8	0.003	- 0.392	1.173	1.111	- 0.296	0.010	- 0.131	- 0.685	- 0.152	- 0.605	
9	0.342	- 0.667	- 1.712	- 0.336	- 1.650	- 0.571	0.028	1.038	- 1.651	- 0.444	
10	0.197	- 0.363	- 0.694	1.320	- 0.740	0.347	0.544	2.011	- 2.235	- 0.682	

References

- Akinoglu, B. G., & Ecevit, A. (1990). Construction of a quadratic model using modified Ångström coefficients to estimate global solar radiation. *Solar Energy*, 45, 85–92.
- Allen, R. G. (1997). Self-calibrating method for estimating solar radiation from air temperature. *Journal of Hydrologic Engineering*, 2, 56–67.
- Almorox, J., Benito, M., & Hontoria, C. (2005). Estimation of monthly Ångström-Preccott equation coefficients from measured daily data in Toledo, Spain. *Renewable Energy*, 30, 931–936.
- Almorox, J., Bocco, M., & Willington, E. (2013). Estimation of daily global solar radiation from measured temperatures at Cañada de Luque, Córdoba, Argentina. *Renewable Energy*, 60, 382–387.
- Ampratwum, D. B., & Dorvlo, A. S. S. (1999). Estimation of solar radiation from the number of sunshine hours. *Applied Energy*, 63, 161–167.
- Ångström, A. (1924). Solar and terrestrial radiation. *Quarterly Journal of the Royal Meteorological Society*, 50, 121–125.
- Antonopoulos, V. Z., Papamichail, D. M., Aschonitis, V. G., & Antonopoulos, A. V. (2019). Solar radiation estimation methods using ANN and empirical models Author links open overlay panel. *Computers and Electronics in Agriculture*, 160, 160–167.
- Bakirci, K. (2009a). Models of solar radiation with hours of bright sunshine: a review. *Renewable and Sustainable Energy Reviews*, 13, 2580–2588.
- Bakirci, K. (2009b). Correlations for estimation of daily global solar radiation with hours of bright sunshine in Turkey. *Energy*, 34, 485–501.
- Barron, M. G., Vivian, D. N., Yee, S. H., & Santavy, D. L. (2009). Methods to estimate solar radiation dosimetry in coral reefs using remote sensed, modeled, and in situ data. *Environmental Monitoring and Assessment*, 151, 445–455.
- Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43, 3–31.
- Besharat, F., Dehghan, A. A., & Faghih, A. R. (2013). Empirical models for estimating global solar radiation: a review and case study. *Renewable and Sustainable Energy Reviews*, 21, 798–821.
- 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.
- Chen, R., Ersi, K., Yang, J., Lu, S., & Zhao, W. (2004). Validation of five global radiation models with measured daily data in China. *Energy Conversion and Management*, 45, 1759–1769.
- Choubin, B., Zehtabian, G., Azareh, A., Rafiei-Sardooi, E., Sajedi-Hosseini, F., & Kişi, Ö. (2018). Precipitation forecasting using classification and regression trees (CART) model: a comparative study of different approaches. *Environmental Earth Sciences*, 77(8), 314.
- Citakoglu, H. (2015). Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation. *Computers and Electronics in Agriculture*, 118, 28–37.
- De Souza, J. L., Bastos Lyra, G., Dos Santos, C. M., Araujo Ferreira Junior, R., Tiba, C., Bastos Lyra, G., & Maringolo Lemes, M. A. (2016). Empirical models of daily and monthly global solar irradiation using sunshine duration for Alagoas State, Northeastern Brazil. *Sustainable Energy Technologies and Assessments*, 14, 35–45.
- Donatelli, M., & Campbell, G. S. (1998). A simple model to estimate global solar radiation. *Proceedings of the fifth European society of agronomy congress, Nitra, Slovak Republic* (pp. 133–134).
- Duffie, J. A., & Beckman, W. A. (1991). *Solar engineering of thermal processes*. Hoboken: Wiley.
- Elani, U. A. (2007). Distribution of ultraviolet solar radiation at Riyadh Region, Saudi Arabia. *Environmental Monitoring and Assessment*, 124, 235–241.
- Hagan, M. T., & Menhaj, M. B. (1994). Training feed forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6), 989–993.
- Hargreaves, G. H., & Samani, Z. A. (1982). Estimating potential evapotranspiration. *Journal of Irrigation and Drainage Engineering*, 108, 223–230.
- He, Z., Wen, X., Liu, H., & Du, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology*, 509, 379–386.
- Khosravi, A., Koury, R. N. N., Machado, L., & Pabon, J. J. G. (2018). Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms. *Journal of Cleaner Production*, 176, 63–75.
- Liu, X., Li, Y., Zhong, X., Zhao, C., Jensen, J. R., & Zhao, Y. (2014). Towards increasing availability of the Ångström-Preccott radiation parameters across China: spatial trend and modeling. *Energy Conversion and Management*, 87, 975–989.
- Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2), 431–441.
- Moreno-Sáez, R., & Mora-López, L. (2014). Modelling the distribution of solar spectral irradiance using data mining techniques. *Environmental Modelling & Software*, 53, 163–172.
- Muzathik, A. M., Ibrahim, M. Z., Samo, K. B., & Wan Nik, W. B. (2011). Estimation of global solar irradiation on horizontal and inclined surfaces based on the horizontal measurements. *Energy*, 36, 812–818.
- Newland, F. J. (1988). A study of solar radiation models for the coastal region of South China. *Solar Energy*, 31, 227–235.
- Preccott, J. A. (1940). Evaporation from a water surface in relation to solar radiation. *Transactions of the Royal Society of South Australia*, 64, 114–118.
- Quej, V. H., Almorox, J., Ibrakhimov, M., & Saito, L. (2016). Empirical models for estimating daily global solar radiation in Yucatán Peninsula, Mexico. *Energy Conversion and Management*, 110, 448–456.
- Quej, V. H., Almorox, J., Arnaldo, J. A., & Saito, L. (2017). ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. *Journal of Atmospheric and Solar - Terrestrial Physics*, 155, 62–70.
- Rafiei-Sardooi, E., Mohseni-Saravi, M., Barkhori, S., Azareh, A., Choubin, B., & Jafari-Shalamzar, M. (2018). Drought modeling: a comparative study between time series and neuro-

- fuzzy approaches. *Arabian Journal of Geosciences*, 11(17), 487.
- Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., & Pradhan, B. (2018). A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Science of the Total Environment*, 644, 954–962.
- Samani, Z. (2000). Estimating solar radiation and evapotranspiration using minimum climatological data. *Journal of Irrigation and Drainage Engineering*, 126, 265–267.
- Samuel, T. D. M. A. (1991). Estimation of global radiation for Sri Lanka. *Solar Energy*, 47, 333–337.
- 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.
- Sonmete, M. H., Ertekin, C., Menges, H. O., Haciseferoğullari, H., & Evrendilek, F. (2011). Assessing monthly average solar radiation models: a comparative case study in Turkey. *Environmental Monitoring and Assessment*, 175, 251–277.
- Urraca, R., Martinez-de-Pison, E., Sanz-Garcia, A., Antonanzas, J., & Antonanzas-Torres, F. (2017). Estimation methods for global solar radiation: case study evaluation of five different approaches in central Spain. *Renewable and Sustainable Energy Reviews*, 77, 1098–1113.
- Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: a review. *Renewable Energy*, 105, 569–582.
- Wacker, S., Gröbner, J., Zysset, C., Diener, L., Tzoumanikis, P., Kazantzidis, A., Vuilleumier, L., Stöckli, R., Nyeki, S., & Kämpfer, N. (2015). Cloud observations in Switzerland using hemispherical sky cameras. *Journal of Geophysical Research – Atmospheres*, 120, 695–707.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probability & statistics for engineers & scientists*. Boston: Prentice Hall.
- Wang, J., Wang, E., Yin, H., Feng, L., & Zhao, Y. (2015). Differences between observed and calculated solar radiations and their impact on simulated crop yields. *Field Crops Research*, 176, 1–10.
- Widén, J., Carpman, N., Castellucci, V., Lingfors, D., Olauson, J., Remouit, F., Bergkvist, M., Grabbe, M., & Waters, R. (2015). Variability assessment and forecasting of renewables: a review for solar, wind, wave and tidal resources. *Renewable and Sustainable Energy Reviews*, 44, 356–375.
- Yorukoglu, M., & Celik, A. L. (2006). A critical review on the estimation of daily global solar radiation from sunshine duration. *Energy Conversion and Management*, 47, 2441–2450.

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