



OPEN ACCESS

EDITED BY

Georgios E. Arnaoutakis,
Hellenic Mediterranean University, Greece

REVIEWED BY

Nikolaos Savvakis,
Technical University of Crete, Greece
Ityona Amber,
Robert Gordon University, United Kingdom

*CORRESPONDENCE

Harry D. Kambezidis,
✉ harry@noa.gr,
✉ harry@univwa.gr

RECEIVED 14 November 2024

ACCEPTED 08 January 2025

PUBLISHED 29 January 2025

CITATION

Kambezidis HD (2025) An in-depth analysis of two common methodologies used to derive hourly solar radiation values from daily ones: pros and cons.

Front. Environ. Sci. 13:1528355.
doi: 10.3389/fenvs.2025.1528355

COPYRIGHT

© 2025 Kambezidis. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An in-depth analysis of two common methodologies used to derive hourly solar radiation values from daily ones: pros and cons

Harry D. Kambezidis^{1,2*}

¹Institute of Environmental Research and Sustainable Development, National Observatory of Athens, Lofos Nymphon, Athens, Greece, ²Laboratory of Soft Energies Applications and Environmental Protection, Department of Mechanical Engineering, University of West Attica, Aigaleo, Greece

Solar radiation comprises the primary renewable source of energy on Earth and has so been exploited in the last 20 years. Despite this, solar radiation measurements are scarce worldwide, thus giving space to modelling. Nevertheless, modelling solar radiation at an hourly level is nowadays required for a short-term output forecast from solar installations. The daily global solar radiation decomposition models are one category of solar models that convert daily solar radiation values to hourly ones. The Collares-Pereira and Rabl (CPR) and Collares-Repeira, Rabl and Gueymard (CPRG) models have shown to provide a better performance than others at individual sites without exhibiting any sign of universality on the other hand. The current study looks at this gap. In this regard, twelve sites are selected around the world. To estimate hourly values, the CPR and CPRG models are applied to daily solar radiation estimates for each site in particular years. Hourly data sets that are openly accessible provide the daily values. Additionally, daily and monthly values are derived from the estimated and observed hourly values. The hourly, daily, and monthly scales are used to compare the two models' performances. The CPR model outperforms the CPRG model across all sites and time scales. A universal coefficient of correction is used to further enhance the CPR performance, bringing the CPR-estimated solar radiation very close to the observed one.

KEYWORDS

solar radiation, decomposition models, hourly values, daily values, monthly values, universal applicability

1 Introduction

Solar energy is the primary source on Earth for controlling various fields such as the atmospheric environment (Giesen et al., 2008), terrestrial climate (Larsen et al., 2007), and terrestrial ecosystems (Bojinski et al., 2014). The abundance of solar radiation on Earth also makes it the most significant renewable energy source. As a result, it has been used in a variety of solar energy applications, primarily in the form of photovoltaic (PV) energy, such as (Kambezidis, 2021; Kambezidis, 2022). The substantial contribution that solar radiation makes to the Earth's thermal (energetic) balance is another crucial function. The hydrological cycles employed for irrigation, agriculture, and water-resource management are impacted by this equilibrium in terms of evapotranspiration (Boscaini

et al., 2020). For these reasons, the knowledge of solar radiation availability at a location is important. Furthermore, using solar energy requires this kind of understanding (Khatib et al., 2012). However, even now, there is a lack of data based on surface solar radiation measurement stations (Kambezidis, 2022). This disadvantage has long since made the creation of models of solar radiation necessary.

The models may be statistical procedures (e.g., heuristic, fuzzy logic), mathematical formulations (e.g., linear, polynomial), or artificial intelligence (e.g., artificial neural networks, generalised regression, feedback-back forward, cascade-forward back-propagation, neuro-fuzzy, and optimised ANNs) (Kambezidis, 2022; Khatib et al., 2012). The purpose of the models is to synthesise solar radiation time series in locations without solar data availability. Predicting the short-term (few hours ahead) solar radiation at a place is another goal, particularly for the AI-based models. This short-term knowledge is highly significant for the market of renewable energy (Nwokolo et al., 2022).

Hourly readings are crucial for solar modelling assessments or in-depth analyses of the solar availability at the site of interest, even though many meteorological stations worldwide now record horizontal solar radiation on a daily basis (Kambezidis, 2022). With the intention of providing hourly values of global horizontal solar radiation in conjunction with meteorological records or to fill in gaps, the so-called daily global solar radiation decomposition (DGSRD) models (Yao et al., 2015) were developed back in the 1950s, for example Whillier (1956). Depending on the input parameters that are employed, the DGSRD models are separated into two groups. Solar geometry (i.e., solar hour angle, solar declination, day length, and time, either local or solar) is used in the first category. Whillier's work is a typical example (Whillier, 1956). Liu and Jordan (1960), Collares-Pereira and Rabl (1979), Gueymard (1986), which is a modification of the CPR, although this latter researcher later developed an own model (Gueymard, 2000), Garg and Garg (1987), and Newell (1983) are other models of this type. The second group comprises models that utilise the Gaussian function, including representative models created by Jain (1984), Jain (1988), Baig et al. (1991), El shazly (1996).

Using data from specific sites, the accuracy of the aforementioned DGSRD models has been demonstrated. These sites include four in South Africa (Whillier, 1956), one in the United States (Liu and Jordan, 1960), five in the United States (Collares-Pereira and Rabl, 1979), one in Canada (Gueymard, 1986; Gueymard, 2000), four in India (Garg and Garg, 1987), one in the United States (Newell, 1983), one in Italy (Jain, 1984), one in Canada (Jain, 1988), one in Pakistan (Baig et al., 1991), and one in Egypt (El shazly, 1996). All of these studies have looked at data from one or more sites in a single country, with the exception of the work by Gueymard (2000), which collected data from 135 locations worldwide to develop and assess his own model. However, none of these studies have assessed the most popular DGSRD models of categories 1 and 2 using global data and drawn conclusions regarding their effectiveness. Therefore, bridging this gap is the goal of the current work. Only models that have demonstrated accuracy in the majority of research have been selected here for simplicity's sake; the CPR and CPRG models satisfy the requirements, as demonstrated by Yao et al. (2015). Thus, these two DGSRD models have been considered in the current work.

The article is divided into sections. Section 2 deploys the data and methodologies used. Section 3 provides the results of the study. Section 4 gives the conclusions of the work and discusses its main achievements, while it refers to future work in order to expand the efficiency of the models. A nomenclature is also deployed to provide a summary of the abbreviations and main symbols used in the study. Acknowledgements and references follow.

2 Materials and methods

2.1 Data collection

To implement the goals of the study, the same 12 sites around the world with those selected by Kambezidis et al. (2024) were used. The selection criteria were: different environmental characteristics (i.e., different climatological conditions), different terrain features (i.e., different topographical characteristics), and distribution across the continents (provided that available measurements exist). This way, all the climatological, terrain and distribution patterns are reflected in the solar radiation data. Table 1 shows the selected sites in alphabetical order together with their geographical coordinates, environmental description, climatological classification, and period of usable data, while Figure 1 provides a world map indicating the location of the 12 sites.

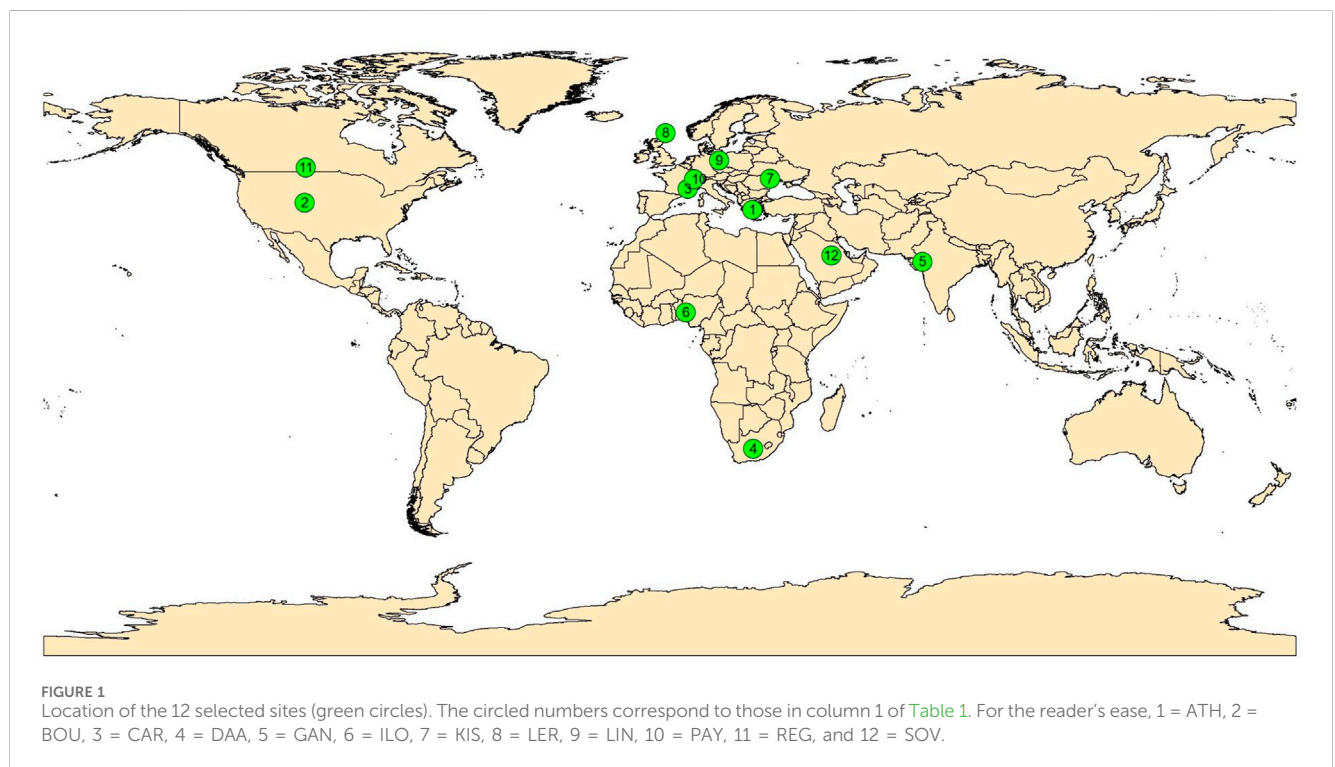
For the purpose of the study, the global horizontal irradiance, H_g (in Wm^{-2}), included in the data sets of Kambezidis et al. (2024) for the 12 sites, was also used. The selection of the sites and their solar radiation data were based on the BSRN (Baseline Surface Radiation Network), except for Athens (ATH); in this case, data from the Actinometric Station of the National Observatory of Athens (ASNOA) not belonging to BSRN was used. The abbreviations of the sites (except for Athens) are those provided by the BSRN typology. A description of the BSRN operation can be found in Driemel et al. (2018). All data was downloaded from the BSRN network on permission, except for the Athens data, which is part of the solar radiation components measured at ASNOA, Greece, in continuous operation since 1952.

As shown in Table 1, single years were selected for analysis at each site instead of a period of years. As mentioned in Kambezidis et al. (2024) "...That was performed on the rationale that no climatological analysis was intended to be conducted within the scope of this study ... For this reason, the year for each site was selected in the period 1999 to 2020 from the BSRN list of stations, with an additional restriction that the individual years cover the mentioned period as broadly as possible. This way, any weather peculiarities occurring over an extended area within a specific year would be avoided."

All data values are hourly averages. The BSRN network provides its data in UTC (universal time coordinated), while the data from ASNOA is in LST (local standard time). Therefore, a transformation of all UTC data into LST ones for the 12 sites, except ATH, took place in the data elaboration phase. In this stage, the corresponding solar hour angle, ω (positive degrees after solar noon, $-180^\circ \leq \omega \leq 180^\circ$), solar altitude, γ (positive degrees above the local horizon, $-90^\circ \leq \gamma \leq 90^\circ$), solar azimuth, ψ (positive degrees after solar noon, $-180^\circ \leq \psi \leq 180^\circ$), and solar declination, δ (in degrees, $-23.5^\circ \leq \delta \leq 23.5^\circ$) values over the selected year for

TABLE 1 The 12 selected sites of the present study in alphabetical order. The symbols ϕ , λ , and z denote the geographical latitude (positive degrees north of the Equator), the geographical longitude (positive degrees east of the Greenwich meridian), and the altitude of the site (amsl = above mean sea level), respectively. In column 6, I denotes rural, and II denotes urban environment. N = north, S = south, E = east, W = west. All geographical coordinates have been rounded to the second decimal digit.

Number#, site, country (abbreviation)	ϕ (deg)	λ (deg)	z (m amsl)	Terrain features (topography)	Terrain type	Köppen-Geiger climate classification	Period
#1, Athens, Greece (ATH)	37.97 N	23.72 E	107	Shrubs, trees (hilly)	II	Csa (temperate, dry summer, hot summer)	2000
#2, Boulder, United States (BOU)	40.05 N	105.01 W	1,577	Grass (flat)	I	BSk (arid, steppe, cold)	1998
#3, Carpentras, France (CAR)	44.08 N	5.06 E	100	Cultivated land (hilly)	I	Csa (temperate, dry summer, hot summer)	2018
#4, De Aar, S. Africa (DAA)	30.67 S	23.99 E	1,287	Sand (flat)	I	BSk (arid, steppe, cold)	2017
#5, Gandhinagar, India (GAN)	23.11 N	72.63 E	65	Shrubs (flat)	II	BSh (arid, steppe, hot)	2020
#6, Ilorin, Nigeria (ILO)	8.53 N	4.57 E	350	Shrubs (flat)	I	Aw (tropical, savannah)	2003
#7, Kishinev, Moldova (KIS)	47.00 N	28.82 E	205	Grass (flat)	II	Dfb (cold, no dry season, warm summer)	2020
#8, Lerwick, UK (LER)	60.14 N	1.18 W	80	Grass (hilly)	I	Cfc (temperate, no dry season, cold summer)	2003
#9, Lindenberg, Germany (LIN)	52.21 N	14.12 E	125	Cultivated land (hilly)	I	Dfb (cold, no dry season, warm summer)	2018
#10, Payerne, Switzerland (PAY)	46.82 N	6.94 E	491	Cultivated land (hilly)	I	Dfb (cold, no dry season, warm summer)	2013
#11, Regina, Canada (REG)	50.21 N	104.71 W	578	Cultivated land (flat)	I	Dfb (cold, no dry season, warm summer)	2003
#12, Solar Village, S. Arabia (SOV)	24.91 N	46.41 E	650	Desert, sand (flat)	I	BWh (arid, desert, hot)	2002



each station were calculated by using the XRONOS.bas code (Kambezidis and Papanikolaou, 1990; Kambezidis and Tsangrassoulis, 1993), (*xronos* means *time* in Greek with *x* being

spelled as *ch*). XRONOS.bas is an improvement to the original SUNAE code (Walraven, 1978). A discontinuity in the estimation of ψ at the sunrise and sunset moments was recently discovered and

TABLE 2 Statistical estimators per site for the evaluation of the CPR-estimated hourly $H_{g,h}$ values with the observed ones by using linear regression analysis. All regression analyses were derived at the 99.9% CI ($\alpha = 0.001$). The overbars and $\sigma()$ denote the annual average and standard deviation, respectively, while the subscripts obs and est refer to the corresponding observed and estimated values. All numbers have been rounded to the second decimal digit except for p, which is integer.

Site	RMSE (Wm ⁻²)	MBE (Wm ⁻²)	MAE (Wm ⁻²)	R ²	d	α	p	Sig	$\overline{H_{g,h,obs}}$ (Wm ⁻²)	$\overline{H_{g,h,est}}$ (Wm ⁻²)	$\sigma(H_{g,h,obs})$ (Wm ⁻²)	$\sigma(H_{g,h,est})$ (Wm ⁻²)
ATH	67.69	-0.92	43.65	0.94	0.99	0.001	0	Yes	379.50	378.58	284.20	269.81
BOU	137.70	3.18	110.99	0.76	0.97	0.001	0	Yes	335.32	338.50	283.26	251.94
CAR	98.81	3.72	78.10	0.87	0.98	0.001	0	Yes	310.25	313.97	278.52	263.67
DAA	97.94	1.18	74.91	0.91	0.99	0.001	0	Yes	452.59	453.76	334.97	326.94
GAN	91.99	0.31	74.88	0.89	0.99	0.001	0	Yes	360.15	360.46	271.35	250.14
ILO	132.86	0.97	108.38	0.80	0.98	0.001	0	Yes	382.59	383.56	295.68	256.64
KIS	33.91	1.12	12.26	0.89	0.99	0.001	0	Yes	120.03	121.15	100.22	97.13
LER	68.14	2.71	43.27	0.87	0.98	0.001	0	Yes	163.99	166.70	184.98	165.22
LIN	64.47	2.47	41.14	0.93	0.99	0.001	0	Yes	254.45	256.92	243.70	234.09
PAY	87.35	3.94	66.66	0.89	0.99	0.001	0	Yes	285.76	289.70	269.19	253.45
REG	111.18	4.59	90.63	0.81	0.99	0.001	0	Yes	300.49	305.08	254.70	240.95
SOV	56.98	-0.56	38.37	0.97	1.00	0.001	0	Yes	488.61	488.04	346.01	335.95

solved by providing an update version of XRONOS [XRONOS.ma in the Matlab environment (Kambezidis et al., 2022)]. For the purpose of the present work, the latest version of XRONOS was used to calculate the γ , ψ , and δ values at the 12 sites halfway between two consecutive hours (i.e., at n:30' between hours n and n+1). Then, these values were assigned to the H_g value at the n+1 h. After completing these steps, a quality-control procedure took place over all the 12 data sets: values of $H_{g,h} \leq 0$ Wm⁻² were rejected; also no calculation was made for $\gamma < 5^\circ$. The remaining hourly values on each day were summed up to give the daily ones in Whm⁻².

2.2 Methodology

As mentioned in the Introduction, two specific DGSRD models are considered in the present analysis, i.e., the CPR and CPRG ones. Their mathematical formulations are the following.

2.2.1 The CPR model

$$\frac{H_{g,h}}{H_{g,d}} = r_{CPR} \tag{1}$$

$$r_{CPR} = \frac{\pi}{24} \cdot (a + b \cdot \cos \omega) \cdot \frac{\cos \omega - \cos \omega_{ss}}{\sin \omega_{ss} - \left(\frac{\pi - \omega_{ss}}{180}\right) \cdot \cos \omega_{ss}} \tag{2}$$

$$a = 0.4090 + 0.5016 \cdot \sin(\omega_{ss} - 60^\circ) \tag{3}$$

$$b = 0.6609 - 0.4767 \cdot \sin(\omega_{ss} - 60^\circ) \tag{4}$$

$$\omega_{ss} = \arccos(-\tan \varphi \cdot \tan \delta) \tag{5}$$

where $H_{g,h}$ and $H_{g,d}$ are the hourly and daily solar radiation values in Wm⁻² and Whm⁻², respectively; r_{CPR} is the ratio of hourly to daily global horizontal solar radiation (in h⁻¹); ω_{ss} is the sunset solar hour angle (in degrees); φ is the geographical latitude of the site (positive

degrees in the northern hemisphere); a, b are the constants of Equation 2.

2.2.2 The CPRG model

$$\frac{H_{g,h}}{H_{g,d}} = r_{CPRG} \tag{6}$$

$$r_{CPRG} = \frac{r_0}{f_c} \cdot (a + b \cdot \cos \omega) \tag{7}$$

$$a = 0.4090 + 0.5016 \cdot \sin(\omega_{sr} - 60^\circ) \tag{8}$$

$$b = 0.6609 - 0.4767 \cdot \sin(\omega_{sr} - 60^\circ) \tag{9}$$

$$r_0 = \frac{\pi}{24} \cdot \frac{\cos \omega - \cos \omega_{sr}}{\sin \omega_{sr} - \left(\frac{\pi - \omega_{sr}}{180}\right) \cdot \cos \omega_{sr}} \tag{10}$$

$$f_c = a + 0.5 \cdot b \cdot \frac{\frac{\pi - \omega_{sr}}{180} - \sin \omega_{sr} \cdot \cos \omega_{sr}}{\sin \omega_{sr} - \frac{\pi - \omega_{sr}}{180} \cdot \cos \omega_{sr}} \tag{11}$$

$$\omega_{sr} = \arccos(\tan \varphi \cdot \tan \delta) \tag{12}$$

where r_{CPRG} is the ratio of hourly to daily global horizontal solar radiation (in h⁻¹); ω_{sr} is the sunrise solar hour angle (in degrees); a, b are the constants of Equation 7.

2.2.3 Data processing

For each of the 12 sites, the hourly ratios r_{CPR} and r_{CPRG} were calculated from the expressions (Equations 2, 7) for the CPR and CPRG models, respectively. Then, the criteria mentioned in Section 2.1 were applied to the 12 data sets, which left 4527, 4757, 4858, 4719, 4547, 4540, 4831, 4964, 4852, 4910, 4724, and 4771 h free of errors on annual basis for the ATH, BOU, CAR, DAA, GAN, ILO, KIS, LER, LIN, PAY, REG, and SOV sites, respectively. After cleaning the data sets, a summation of the $H_{g,h}$ hourly values for each day in the year corresponding to each site took place to derive the daily $H_{g,d}$ values. The daily $H_{g,d}$ values were afterwards

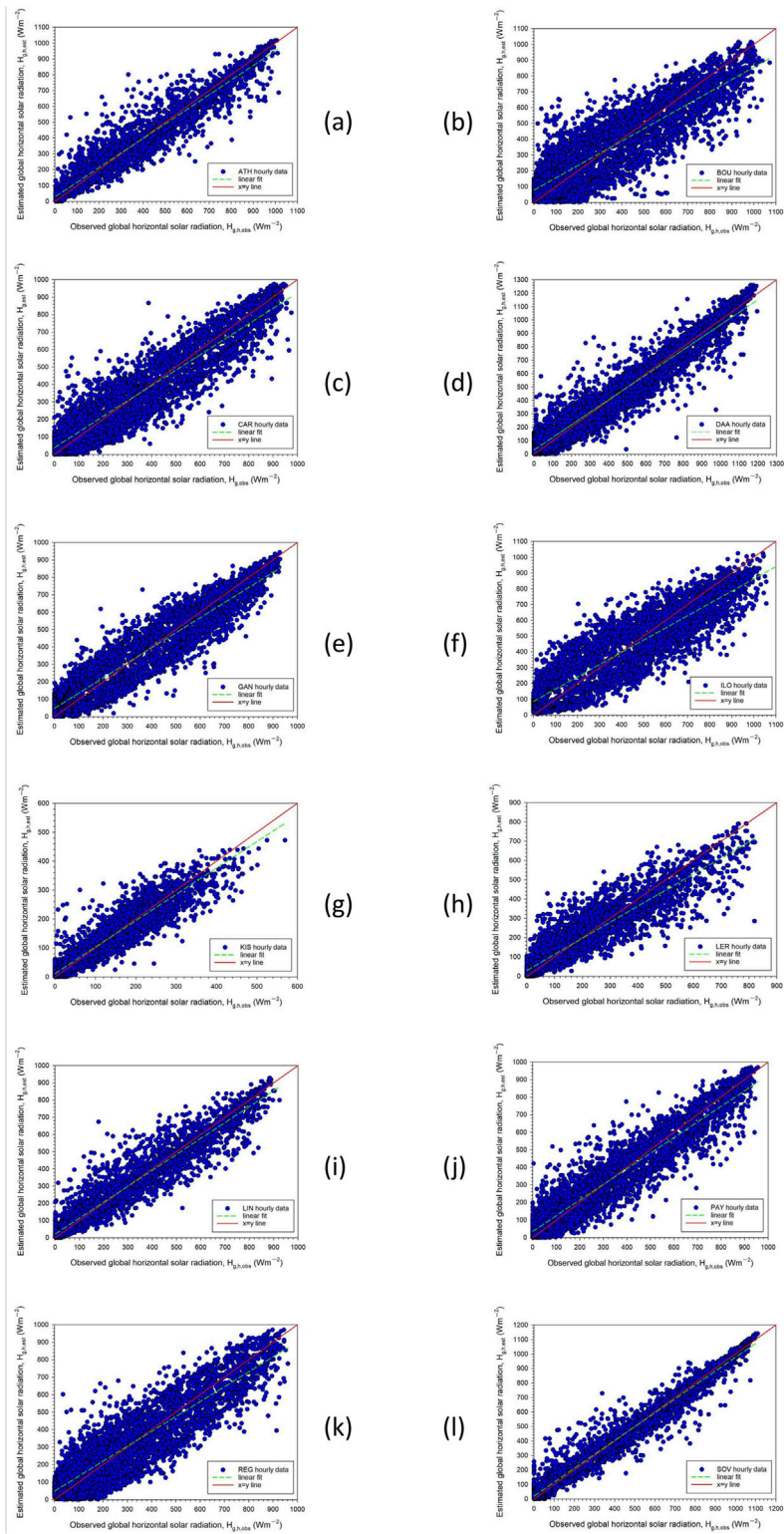


FIGURE 2 Scatter plots of the hourly mean $H_{g,h,est}$ estimated values from the CPR model versus the $H_{g,h,obs}$ observed ones for the 12 sites (panels (A-L), respectively) during the year for each of them. The solid red lines indicate the 1:1 lines and the green dotted ones the linear fits to the $H_{g,h,est} - H_{g,h,obs}$ data pairs. All the linear regressions are significant at the 99.9% CI.

TABLE 3 Regression equations for the observed versus the estimated solar radiation hourly data at all the 12 sites at the 99.9% CI by using the CPR-model expressions (1)–(5) and the CPRG-model ones (6)–(12); the corresponding R² values have been taken from Tables 2, 4, respectively. All numbers have been rounded to the second decimal digit.

Site	Model	Regression equations	R ²
ATH	CPR	$H_{g,h,est} = 0.92 \cdot H_{g,h,obs} + 28.57$	0.94
	CPRG	$H_{g,h,est} = 0.14 \times 10^{-2} \cdot H_{g,h,obs}^2 - 0.21 \cdot H_{g,h,obs} + 163.68$	0.87
BOU	CPR	$H_{g,h,est} = 0.78 \cdot H_{g,h,obs} + 77.82$	0.76
	CPRG	$H_{g,h,est} = 0.10 \times 10^{-2} \cdot H_{g,h,obs}^2 + 159.03$	0.71
CAR	CPR	$H_{g,h,est} = 0.89 \cdot H_{g,h,obs} + 39.32$	0.87
	CPRG	$H_{g,h,est} = 0.14 \times 10^{-2} \cdot H_{g,h,obs}^2 + 61.40$	0.81
DAA	CPR	$H_{g,h,est} = 0.93 \cdot H_{g,h,obs} + 31.23$	0.91
	CPRG	$H_{g,h,est} = 1.16 \cdot H_{g,h,obs} + 38.16$	0.91
GAN	CPR	$H_{g,h,est} = 0.87 \cdot H_{g,h,obs} + 48.05$	0.89
	CPRG	$H_{g,h,est} = 0.05 \times 10^{-2} \cdot H_{g,h,obs}^2 + 0.49 \cdot H_{g,h,obs} + 92.34$	0.78
ILO	CPR	$H_{g,h,est} = 0.78 \cdot H_{g,h,obs} + 86.78$	0.80
	CPRG	$H_{g,h,est} = 0.83 \cdot H_{g,h,obs} + 82.17$	0.79
KIS	CPR	$H_{g,h,est} = 0.91 \cdot H_{g,h,obs} + 11.63$	0.89
	CPRG	$H_{g,h,est} = 0.17 \times 10^{-2} \cdot H_{g,h,obs}^2 + 0.43 \cdot H_{g,h,obs} + 44.89$	0.64
LER	CPR	$H_{g,h,est} = 0.83 \cdot H_{g,h,obs} + 30.09$	0.87
	CPRG	$H_{g,h,est} = 0.20 \times 10^{-2} \cdot H_{g,h,obs}^2 + 259.37$	0.14
LIN	CPR	$H_{g,h,est} = 0.93 \cdot H_{g,h,obs} + 21.19$	0.93
	CPRG	$H_{g,h,est} = 0.16 \times 10^{-2} \cdot H_{g,h,obs}^2 + 132.16$	0.70
PAY	CPR	$H_{g,h,est} = 0.89 \cdot H_{g,h,obs} + 35.18$	0.89
	CPRG	$H_{g,h,est} = 0.16 \times 10^{-2} \cdot H_{g,h,obs}^2 - 0.19 \cdot H_{g,h,obs} + 149.37$	0.77
REG	CPR	$H_{g,h,est} = 0.85 \cdot H_{g,h,obs} + 48.52$	0.81
	CPRG	$H_{g,h,est} = 0.16 \times 10^{-2} \cdot H_{g,h,obs}^2 - 0.23 \cdot H_{g,h,obs} + 179.87$	0.66
SOV	CPR	$H_{g,h,est} = 0.96 \cdot H_{g,h,obs} + 20.07$	0.97
	CPRG	$H_{g,h,est} = 0.09 \times 10^{-2} \cdot H_{g,h,obs}^2 + 0.10 \cdot H_{g,h,obs} + 131.69$	0.94

multiplied by the hourly r_{CPR} or r_{CPRG} values of the same day to produce the estimated $H_{g,h}$ values for the CPR and CPRG models, respectively. Annual and monthly values were also calculated for the observed and modelled H_g values at all the 12 sites. The same averaging process was applied to the r_{CPR} and r_{CPRG} hourly values; these monthly and annual r values in association with the corresponding H_g values were used to see if a better performance of the CPR and CPRG models could be achieved in terms of the statistical metrics discussed in the following Section. All solar radiation values refer to all-sky conditions.

2.2.4 Model performance

To evaluate the model-estimated values (i.e., the performance of the CPR and CPRG models) the following statistical metrics were considered: the root mean-square error (RMSE), the mean-bias error (MBE), the mean-absolute error (MAE), the coefficient of determination (R^2), and the index of agreement (d). Further, a linear (or non-linear in some cases) regression analysis took place between the observed and estimated hourly, daily and monthly H_g values at the 99.9% confidence interval (CI). The formulations of the metrics are given below.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (E_i - O_i)^2}{N}} \tag{13}$$

where E_i and O_i are the i th estimated and observed value, respectively; N is the total number of data points as deployed in Section 2.2.3 for the 12 sites. The RMSE is a measure of the average difference between the predicted values from a model and the actual ones and provides information on the short-term performance of the model. Mathematically, it is the standard deviation of the residuals, which represent the distance between the regression line and the data points. The RMSE, therefore, quantifies how dispersed these residuals are, revealing how tightly the observed data is around the predicted values. The closer the data points to the regression line, the lower the RMSE value and the more errorless the model. This way, a model with less error produces more precise estimations.

$$MBE = \frac{1}{N} \sum_{i=1}^N (E_i - O_i) \tag{14}$$

The MBE serves as a metric to identify the average bias in the estimations of a model as it provides information on the long-term performance of the model. A negative MBE value gives the average amount of underestimation in the calculated value. So, one drawback of this statistical metric is that overestimation of an individual observation may cancel underestimation in a separate one. Nevertheless, the lower the MBE, the higher the model's accuracy.

$$MAE = \frac{1}{N} \sum_{i=1}^N |E_i - O_i| \tag{15}$$

The MAE is a measure of the errors between the estimations from a model and the actual values. It is used to measure how close the estimated values are to the observed ones. It is a simple but powerful metric used to evaluate the accuracy of regression models. It measures the average absolute difference between the predicted and actual values. The lower the MAE, the better the performance of the model.

$$R^2 = 1 - \frac{\sum_{i=1}^N (E_i - O_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \tag{16}$$

The R^2 shows the proportion of the variation in the estimated variable from the observed one. It is a statistical measure that indicates how well a model predicts an outcome. It varies between 0 and 1. A higher R^2 value indicates a better fit of the model to the data, meaning that the model can explain a larger proportion of the variation in the observed variable. The overbar over O indicates the average of the observed variable.

$$d = 1 - \frac{\sum_{i=1}^N (E_i - O_i)^2}{\sum_{i=1}^N (|E_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{17}$$

The d is a statistical measure that assesses the level of agreement between the observed and modelled values. It ranges between 0 and 1, where 1 indicates perfect agreement and 0 indicates no agreement at all. It is often used in fields like hydrology, meteorology, and environmental science to evaluate the accuracy of models and

TABLE 4 Statistical estimators per site for the evaluation of the CPRG-estimated hourly $H_{g,h}$ values with the observed ones by using non-linear regression analysis. All regression analyses were derived at the 99.9% ($\alpha = 0.001$) CI. The overbars and $\sigma()$ denote the annual average and standard deviation, respectively, while the subscripts obs and est refer to the corresponding observed and estimated values. All numbers have been rounded to the second decimal digit except for p, which is integer.

Site	RMSE (Wm^{-2})	MBE (Wm^{-2})	MAE (Wm^{-2})	R^2	d	α	p	Sig	$\overline{H_{g,h,obs}}$ (Wm^{-2})	$\overline{H_{g,h,est}}$ (Wm^{-2})	$\sigma(H_{g,h,obs})$ (Wm^{-2})	$\sigma(H_{g,h,est})$ (Wm^{-2})
ATH	161.95	24.13	119.49	0.87	0.97	0.001	0	Yes	379.50	403.71	284.18	334.77
BOU	185.43	28.29	142.67	0.71	0.96	0.001	0	Yes	342.14	372.89	286.72	307.87
CAR	209.94	-4.86	161.55	0.81	0.95	0.001	0	Yes	314.76	311.25	279.93	362.99
DAA	169.63	110.03	139.51	0.91	0.98	0.001	0	Yes	450.68	550.99	335.54	397.71
GAN	137.50	6.67	102.61	0.78	0.98	0.001	0	Yes	360.36	364.15	271.53	264.05
ILO	135.75	3.83	109.66	0.79	0.98	0.001	0	Yes	382.59	386.42	295.68	260.22
KIS	80.54	19.56	57.22	0.64	0.93	0.001	0	Yes	120.03	139.59	100.23	125.76
LER	478.20	169.01	218.90	0.14	0.53	0.001	0	Yes	163.99	333.00	184.98	473.98
LIN	214.30	66.13	142.71	0.70	0.91	0.001	0	Yes	254.45	320.58	243.75	334.36
PAY	228.40	59.08	162.49	0.66	0.94	0.001	0	Yes	300.49	359.57	254.70	339.62
REG	111.18	4.59	90.63	0.81	0.92	0.001	0	Yes	281.22	285.95	97.00	95.95
SOV	122.11	15.46	100.53	0.94	0.99	0.001	0	Yes	488.61	504.06	346.01	363.68

predictions. Therefore, the closer the d to 1, the higher the model's accuracy.

Further to the above statistical indicators, the percent error (PE) and the standard error (SE) were also adopted for the error analysis discussed in Section 3.4. These statistics are defined as follows.

$$PE (\%) = \frac{\bar{E} - \bar{O}}{\bar{O}} \times 100 \quad (18)$$

The PE is the relative size of the difference between an estimated value, and the true (observed) value. It compares the difference in values to the expected actual value and tells how far off the estimated value is. The overbar over E indicates the average of the estimated variable. PE values up to $\pm 5\%$ are acceptable and the modelled-estimated values are then considered satisfactory; values of $PE \geq |10\%|$ indicate failure of the estimated values (Christie et al., 2005).

$$SE = \frac{\sigma}{\sqrt{N}} \quad (19)$$

where σ is the standard deviation of the population (data points of the variable). The SE refers to the standard deviation of the distribution of the sample average taken from a population. The smaller the SE, the more representative the sample is of the overall population. And the more data points involved in the calculations of the mean, the smaller the SE tends to be. In cases where the SE is large, the data may have some notable irregularities. If the mean of the estimated values of the variable x is \bar{x} , the physical meaning of SE is, if N estimations of the variable were repeated, there would be a 68% probability that the mean value in the new data set will lie within $\bar{x} \pm SE$.

In the propagation of errors theory (Taylor, 1982), if two time series of data, A and B, are independent of each other with errors ΔA and ΔB , respectively, the error of a third time series C of the same variable will have error $\Delta C = \sqrt{(\Delta A^2 + \Delta B^2)}$. In the present study,

the data sets A, B, and C refer to the $H_{g,h,est}$ values for the same site and model.

3 Results

3.1 Evaluation of the CPR and CPRG models at the hourly scale

To perform this evaluation, the observed and estimated hourly $H_{g,h}$ values were subjected to the selected statistical metrics of Section 2.2.4. Tables 2, 4 give the outcomes of the statistical indicators for all the sites and for the CPR and CPRG models, respectively. In both cases the CI has been set to 99.9% ($\alpha = 0.001$), while the null hypothesis is that the estimated hourly $H_{g,h}$ values are not related significantly with the observed ones. However, the p-values are seen to be close to zero (and definitely less than α) at all the sites and both types of the DGSRD models; this rejects the null hypothesis and concludes that all the regression models are significant at the set CI. Despite this general remark, several specific issues can attract attention.

The CPR models (Table 2) at 8 (namely BOU, CAR, GAN, ILO, KIS, LER, PAY, and REG) out of the 12 sites provide $R^2 < 0.90$; the lowest accuracy is shown at the BOU site ($R^2 = 0.76$). Nevertheless, this is not the only criterion to decide about the CPR's performance. These 8 models (especially BOU, ILO, and REG) show higher values of RMSE and MAE than the remaining 4 sites (i.e., ATH, DAA, LIN, SOV), while the MBE values of the 8 sites are not that greater than those for the remaining 4 sites. The higher RMSE values of the 8 sites conclude that there is a great dispersion of the residuals (estimated hourly values – observed hourly values) and, therefore, the observed data is not tightly deployed around the predicted values. Along the same line, the great MAE values of the same sites imply that the estimated values are not so close to the observed ones. On

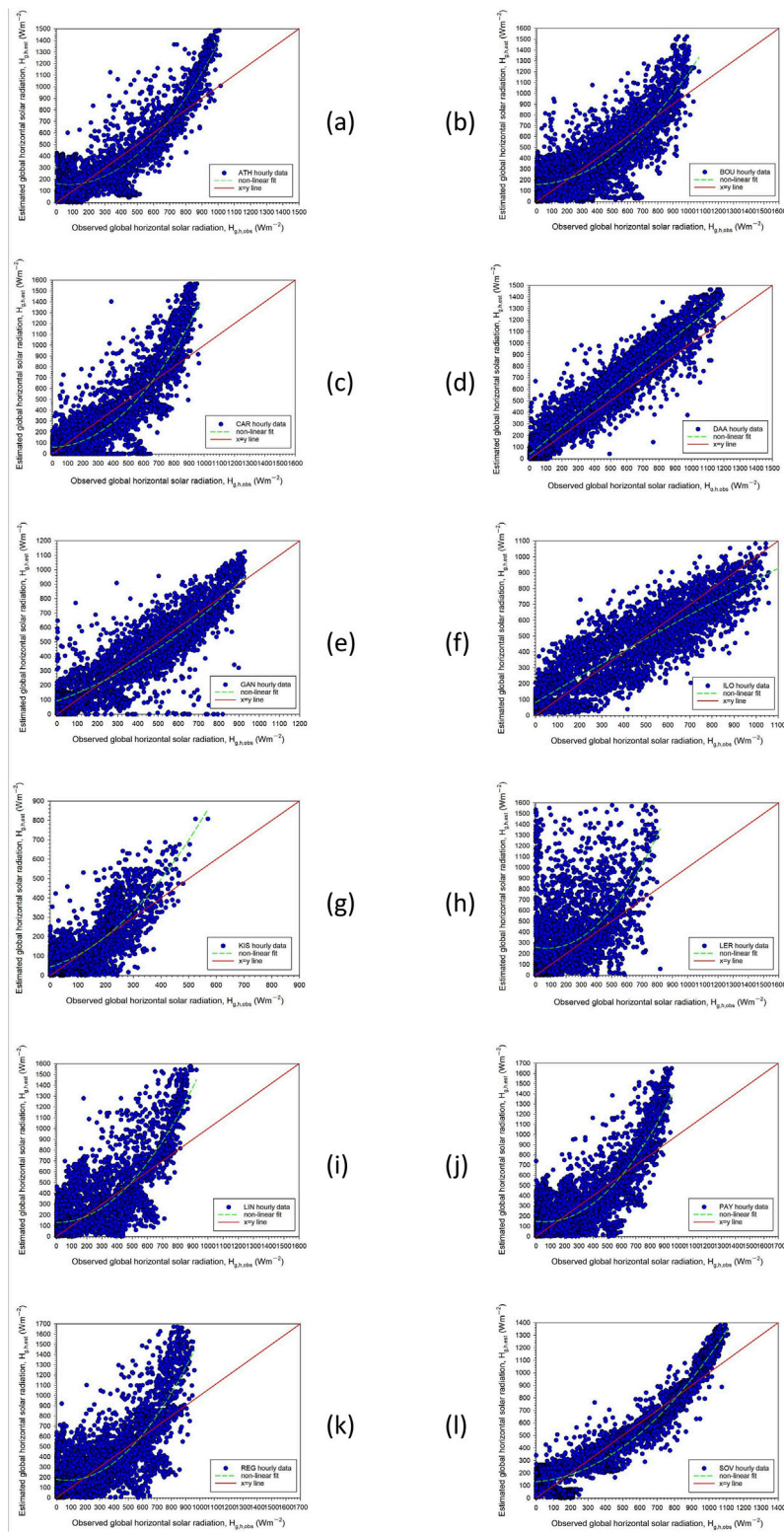


FIGURE 3 Scatter plots of the hourly mean $H_{g,h,est}$ estimated values from the CPRG model versus the $H_{g,h,obs}$ observed ones for the 12 sites (panels **(A-L)**, respectively) during the year for each of them. The solid red lines indicate the 1:1 lines and the green dotted ones the linear fits to the $H_{g,h,est} - H_{g,h,obs}$ data pairs. All the non-linear regressions are significant at the 99.9% CI.

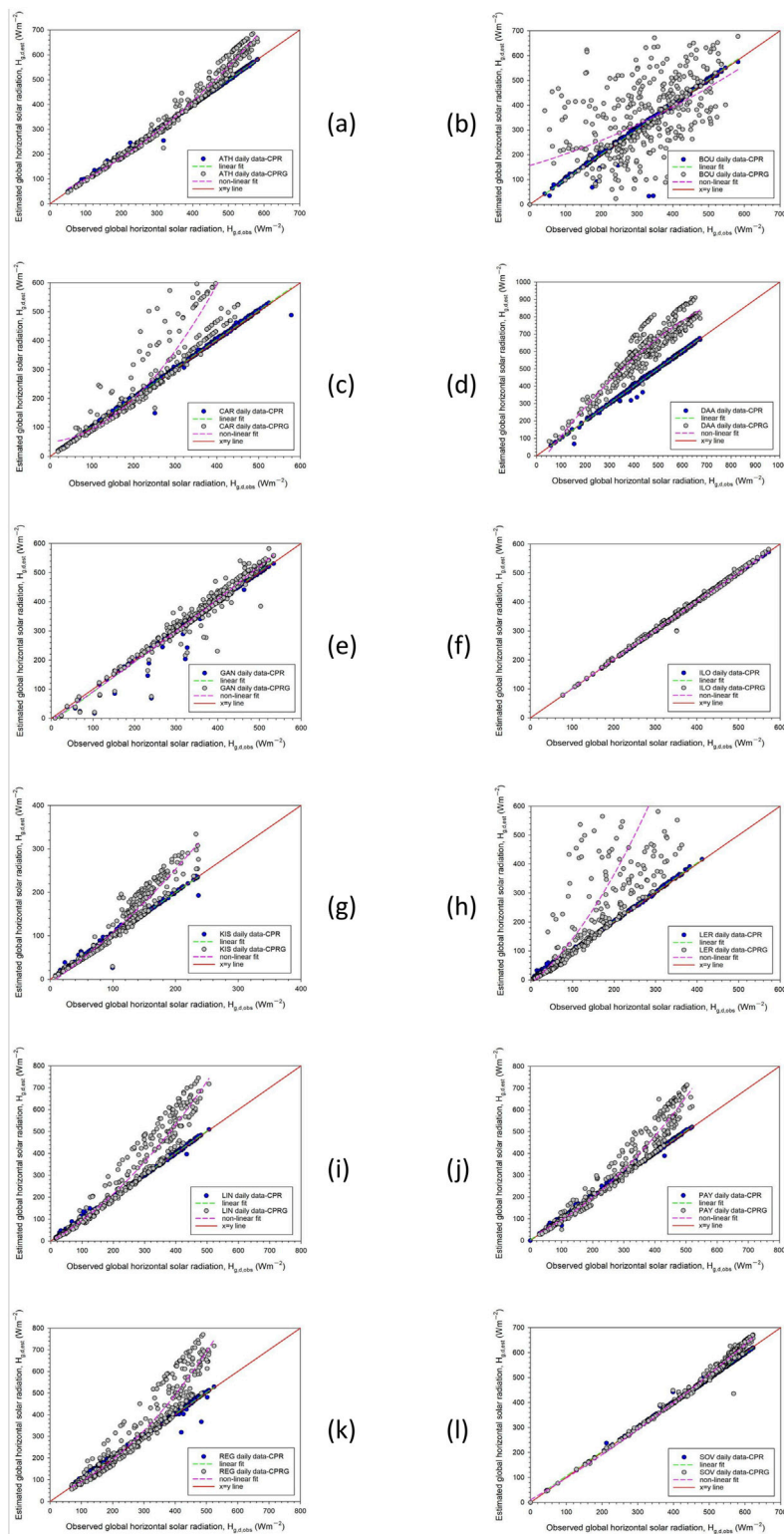


FIGURE 4
 Scatter plots of the daily mean $H_{g,d,est}$ estimated values from the CPR and CPRG models versus the $H_{g,d,obs}$ observed ones for the 12 sites (panels (A–L), respectively) during the year for each of them. The solid red lines indicate the 1:1 lines, the green dotted ones the linear fits to the $H_{g,d,est} - H_{g,d,obs}$ data pairs for the CPR and the pink dotted lines for the CPRG model. The R^2 values are as follows. CPR model: ATH (1.00), BOU (0.95), CAR (1.00), DAA (1.00), GAN (0.98), ILO (1.00), KIS (0.99), LER (1.00), LIN (1.00), PAY (1.00), REG (0.99), and SOV (0.99). CPRG model: ATH (0.98), BOU (0.29), CAR (0.88), DAA (0.89), GAN (0.92), ILO (1.00), KIS (0.93), LER (0.59), LIN (0.94), PAY (0.96), REG (0.92), and SOV (0.99). All the linear and non-linear regressions are significant at the 99.9% CI.

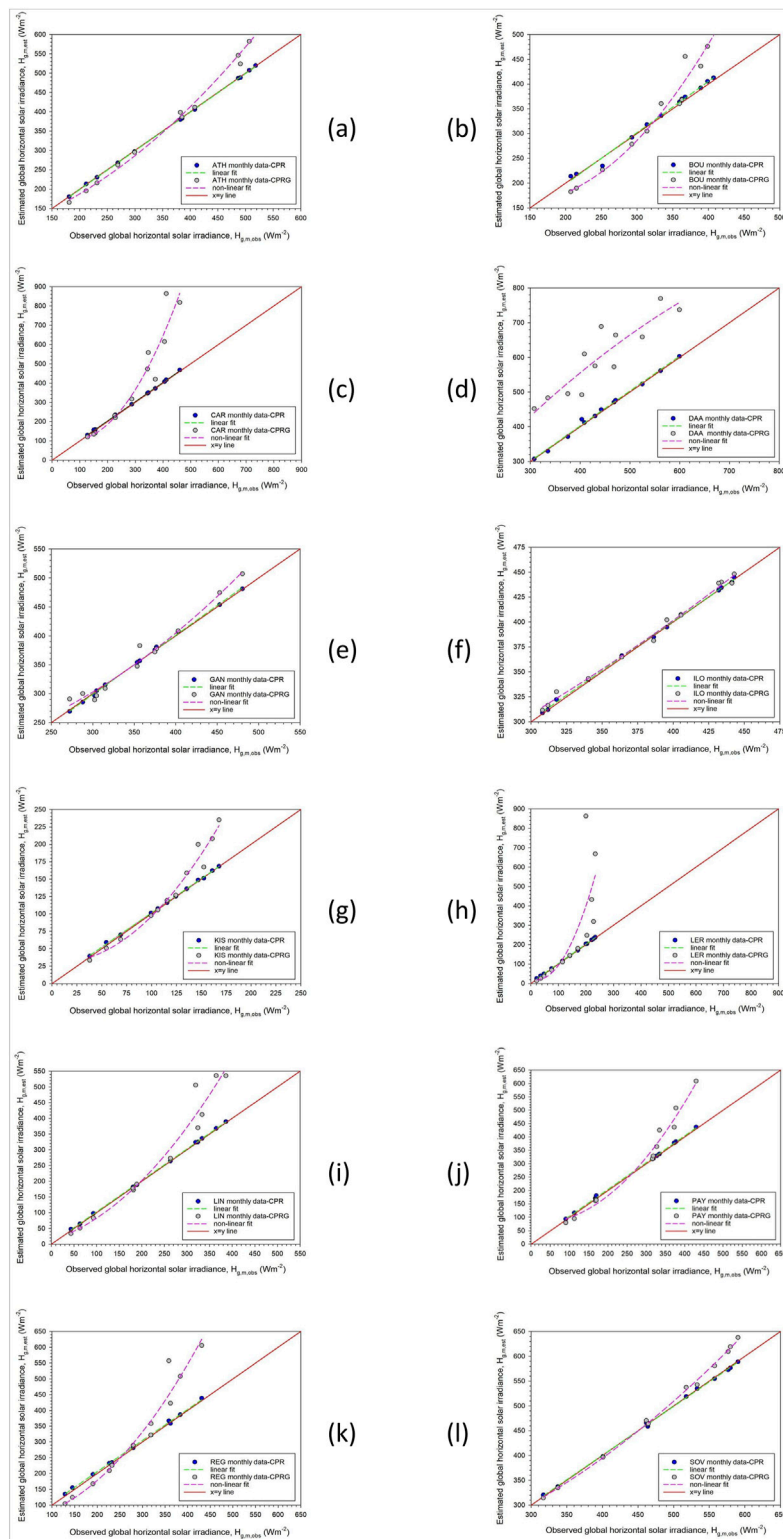


FIGURE 5

Scatter plots of the monthly mean $H_{g,m,est}$ estimated values from the CPR and CPRG models versus the $H_{g,m,obs}$ observed ones for the 12 sites (panels (A-L), respectively) during the year for each of them. The solid red lines indicate the 1:1 lines, the green dotted ones the linear fits to the $H_{g,m,est} - H_{g,m,obs}$ data pairs for the CPR and the pink dotted lines for the CPRG model. The R^2 values are as follows. CPR model: ATH (1.00), BOU (0.99), CAR (1.00), DAA (1.00), GAN (1.00), ILO (1.00), KIS (1.00), LER (1.00), LIN (1.00), PAY (1.00), REG (1.00), and SOV (1.00). CPRG model: ATH (0.99), BOU (0.96), CAR (0.92), DAA (0.81), GAN (0.98), ILO (0.99), KIS (0.97), LER (0.61), LIN (0.96), PAY (0.98), REG (0.95), and SOV (1.00). All the linear and non-linear regressions are significant at the 99.9% CI.

TABLE 5 Statistical indices of percent error (PE) and standard error (SE) for the error analysis of both DGSRD models at the 12 sites. The SE values were combined with the solar radiation estimated averages at each site for both models; the averages of $H_{g,h,est}$ have been taken from Tables 2, 4 for the CPR and CPRG models, respectively.

Site	PE (%)		$\overline{H_{g,h,est}} \pm SE (Wm^{-2})$	
	CRP model	CPRG model	CRP model	CPRG model
ATH	-0.24	+6.36	378.58 ± 5.92	403.71 ± 4.98
BOU	+0.95	+8.44	338.50 ± 5.35	372.89 ± 4.40
CAR	+1.20	+25.84	313.97 ± 3.78	311.25 ± 5.89
DAA	+0.26	+34.11	453.76 ± 4.76	550.99 ± 5.72
GAN	+0.09	+2.07	360.46 ± 3.71	364.15 ± 3.90
ILO	+0.25	-1.00	383.56 ± 3.81	386.42 ± 3.86
KIS	+0.93	+16.29	121.15 ± 4.39	139.59 ± 4.79
LER	+1.65	+103.06	166.70 ± 2.35	333.00 ± 6.73
LIN	+0.97	+25.99	256.92 ± 3.37	320.58 ± 4.81
PAY	+1.38	+17.51	289.70 ± 3.62	359.57 ± 4.76
REG	+1.53	+19.66	305.08 ± 3.48	285.95 ± 5.61
SOV	-0.11	+3.16	488.04 ± 4.86	504.06 ± 5.26

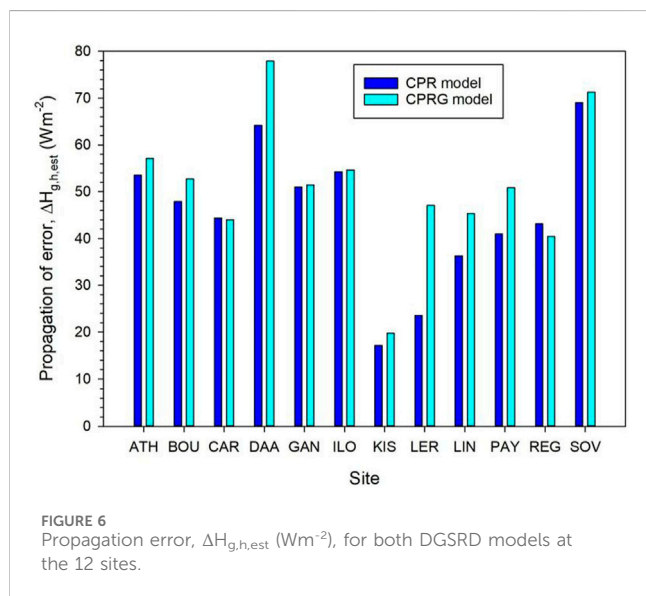


FIGURE 6 Propagation error, $\Delta H_{g,h,est} (Wm^{-2})$, for both DGSRD models at the 12 sites.

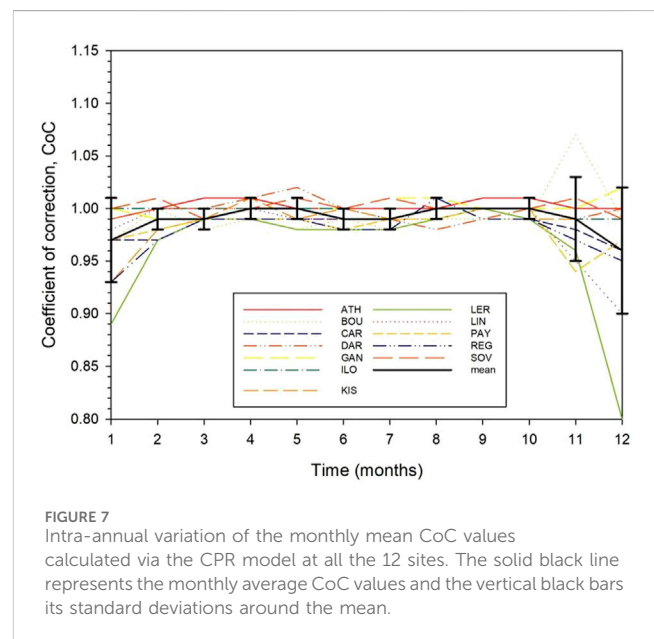


FIGURE 7 Intra-annual variation of the monthly mean CoC values calculated via the CPR model at all the 12 sites. The solid black line represents the monthly average CoC values and the vertical black bars its standard deviations around the mean.

the contrary, their MBE values do not follow the high values of the RMSE and MAE; this means that the models at these sites overestimate the actual data so many times as they underestimate them; the result is that the overestimations are cancelled out by the underestimations. Nevertheless, for all the 12 stations $d \approx 1$, which indicates perfect agreement between the estimated and the observed hourly values; this may seem awkward in relation to the previous statistical indicators, but it can be taken as a close follow-up of the estimated to the observed hourly values, i.e., very similar patterns in the solar radiation variability. In conclusion, the CPR methodology seems to be quite successful for the ATH, DAA, LIN, and SOV sites at the hourly level and less

successful, but acceptable, at the other 8 locations. This is a differentiation to the conclusion drawn by Yao et al. (2015) who found that the CPR model is accurate for just one location in China (i.e., Jiading). To prove the truth of the above statistical results, Figure 2 shows the variation of the estimated versus the observed hourly $H_{g,h}$ values for all the sites. The linear regression equations for the scatter plots in Figure 2 are given in Table 3. These findings corroborate the overall significance of the regression models. In general, the CPR models seem to have a good performance at 11 sites and an inferior one for 1 site (BOU).

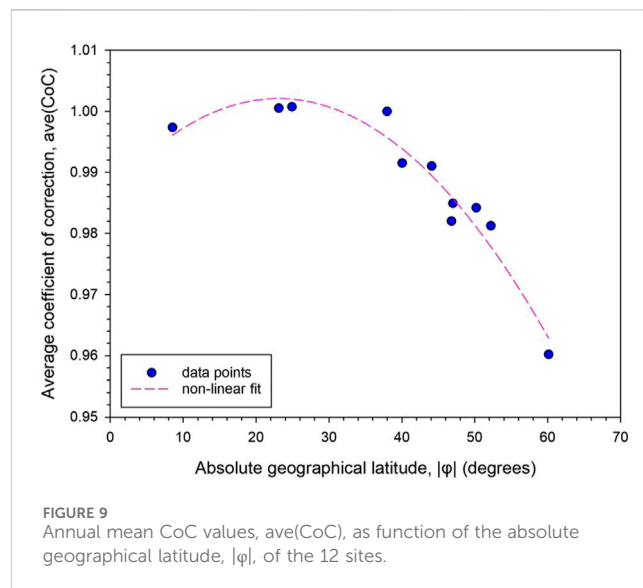
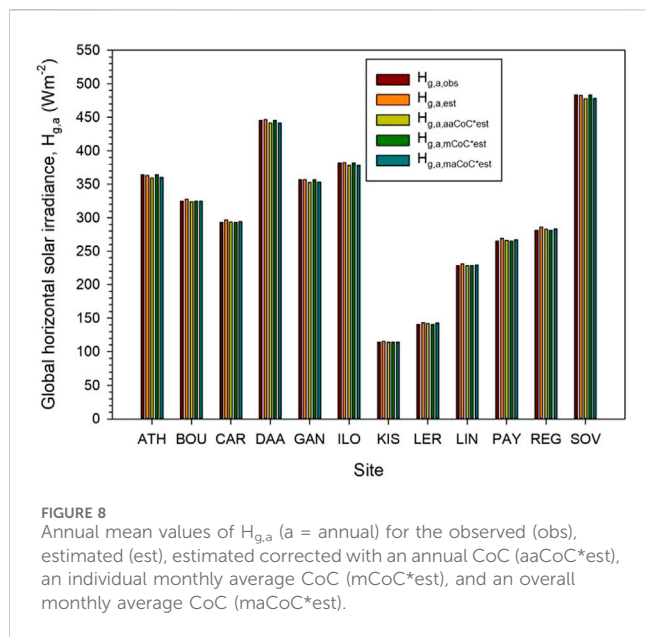


TABLE 6 Annual differences in Wm^{-2} and % between the $H_{g,a,obs}$ values and the $H_{g,a,est}$, $H_{g,a,aaCoC*est}$, $H_{g,a,mCoC*est}$ and $H_{g,a,maCoC*est}$ ones. All numbers have been rounded to the second decimal digit.

Site	$H_{g,a,obs} - H_{g,a,est}$ Wm^{-2} (%)	$H_{g,a,obs} - H_{g,a,aaCoC*est}$ Wm^{-2} (%)	$H_{g,a,obs} - H_{g,a,mCoC*est}$ Wm^{-2} (%)	$H_{g,a,obs} - H_{g,a,maCoC*est}$ Wm^{-2} (%)
ATH	0.95 (0.26)	5.01 (1.38)	0 (0)	4.12 (1.13)
BOU	-2.68 (-0.82)	0.99 (0.31)	0 (0)	0.41 (0.12)
CAR	-3.73 (-1.27)	-0.41 (-0.14)	0 (0)	-1.17 (-0.40)
DAA	-0.98 (-0.22)	4.02 (0.90)	0 (0)	4.09 (0.92)
GAN	-0.12 (-0.03)	3.87 (1.09)	0 (0)	3.53 (0.99)
ILO	-0.90 (-0.24)	3.38 (0.89)	0 (0)	3.36 (0.88)
KIS	-1.16 (-1.02)	0.13 (0.11)	0 (0)	-0.23 (-0.20)
LER	-2.69 (-1.91)	-1.09 (-0.77)	0 (0)	-1.71 (-1.22)
LIN	-2.39 (-1.05)	0.19 (0.08)	0 (0)	-0.77 (-0.34)
PAY	-3.89 (-1.47)	-0.88 (-0.33)	0 (0)	-1.72 (-0.65)
REG	-4.72 (-1.68)	-1.52 (-0.54)	0 (0)	-2.25 (-0.80)
SOV	0.57 (0.12)	5.98 (1.24)	0 (0)	5.17 (1.07)

The CPRG models (Table 4) at only 2 (namely DAA, and SOV) out of the 12 sites provide $R^2 > 0.90$. Though a non-linear regression has been applied to the estimated-observed hourly pairs, this does

not prove so successful for the majority of the sites as in the case of the CPR model. The values of the RMSE, MBE, MAE, and R^2 statistical indices are now worse than those for the same sites

and the CPR models. Nevertheless, $d \approx 1$ at all the sites (except for LER with $d = 0.53$), a result that is interpreted in the same way as for the CPR models. The lowest accuracy is shown at the LER site ($R^2 = 0.14$, Table 3). In conclusion, the CPRG methodology seems to be less successful for all the sites at the hourly level. This is again a contradiction to the conclusion drawn by Yao et al. (2015) who found that the CPRG model is quite accurate for Jiading, China.

To provide a visual inspection of the statistical results, Figure 3 shows the variation of the estimated versus the observed hourly $H_{g,h}$ values. The linear regression equations (CPR models) for the scatter plots of Figure 2 are given in Table 3, and the non-linear regression expressions (CPRG models) of Figure 3 in Table 3. One should note that in many sites the quadratic expressions for the CPRG models miss some coefficients; this comes from the regression analysis, which revealed their p values to be greater than 0.001, thereby non-significant coefficients; nevertheless, after removing these coefficients, the regression expressions satisfy the significance test. In general, the CPRG models seem not to have as a good performance as the CPR models at all the sites at the hourly level.

3.2 Evaluation of the CPR and CPRG models at the daily scale

After examining the efficiency of the CPR and CPRG models at the hourly level, it would be interesting to see whether the models can reproduce the daily $H_{g,d}$ values. To do this, the observed and estimated hourly mean $H_{g,h}$ values at all the sites were converted into daily mean ones. Figure 4 provides the results for the CPR and CPRG models together for space saving. The statistical indicators used in Section 3.1 have been omitted here, because the graphs in Figure 4 speak for themselves. One can easily conclude that the CPR models perform almost ideally, since the estimated-observed daily pairs coincide with the 1:1 line and their R^2 values are practically equal to 1. The CPRG models fail to reproduce the daily solar radiation values, especially in the case of the BOU and LER sites.

3.3 Evaluation of the CPR and CPRG models at the monthly scale

The last evaluation of the selected DGSRD models concerns the monthly level. As in Section 3.2, the hourly mean $H_{g,h}$ values at all the sites were converted into monthly mean ones in both models. The results are shown graphically in Figure 5 for the CPR and CPRG models together for space saving. It is evident to conclude that the CPR models are superior to the CPRG ones, because they lie along the 1:1 lines at all the sites.

3.4 Error analysis

This Section is devoted to the errors derived from the modelled data in relation to the observed one. The statistics used here refer to the PE and SE described in Section 2.2.4. Now, $PE (\%) = \frac{H_{g,h,est} - H_{g,h,obs}}{H_{g,h,obs}} \times 100$ (excluding the zero $H_{g,h}$ values in the calculation of the averages), and $SE = \frac{\sigma_{H_{g,h,est}}}{\sqrt{N}}$ (N is the number of the non-zero $H_{g,h,est}$ hourly data). The overbars indicate the average

value of the measured (obs) and modelled (est) hourly solar radiation values. Table 5 gives the PE and SE values for both CPR and CPRG models and all the 12 sites. From the Table it is easily seen that the PE values for the CPR models lie in the range $[-0.24\%, +1.65\%]$, which is fully acceptable as being $< \pm 5\%$. The CPRG models show PE values in the range $[-1.00\%, +103.06\%]$; the only acceptable PEs are those corresponding to the sites of GAN (+2.07%), ILO (-1.00%), and SOV (+3.16%). The - and + signs indicate respective lower and higher $H_{g,h,est}$ averages than the $H_{g,h,obs}$ ones. The greater PE values in the CPRG models imply that the $H_{g,h,est}$ hourly values are far more off the $H_{g,h,obs}$ ones in comparison to the CPR models; in other words, the errors associated with the CPRG models are higher than those related to the CPR ones.

The SE was calculated by taking into account the entire $H_{g,h,est}$ time series at each site. The SE values were found to lie in the range $[1.81, 7.30]$ for the CPR and in the range $[1.19, 6.31]$ for the CPRG model. It is noticed that the SE values are contained in comparable ranges for both models; this is quite reasonable as this statistic refers to the representativeness of the $H_{g,h,est}$ time series to another repeated estimations for the same site and model. In other words, the mean value of any future $H_{g,h,est}$ data set will have a 68% probability to lie in the range $\overline{H_{g,h,est}} \pm SE$ for the same site and model, where $\overline{H_{g,h,est}}$ is the mean value of the estimated data set in the present study.

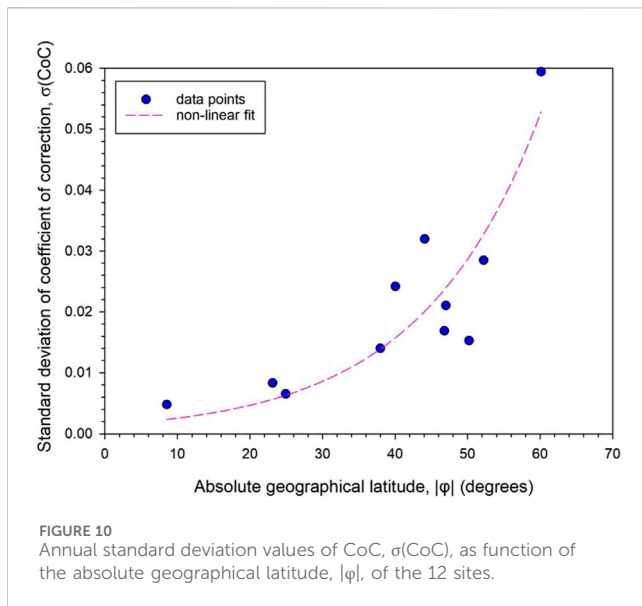
To find the propagation error in the $H_{g,h,est}$ data for each site and model, the error in the mean value was considered as follows. For each site and model the annual $H_{g,h,est}$ mean was taken into account from Tables 2, 4. New averages A and B were then derived as $\bar{A} = \overline{H_{g,h,est}} + 0.1 \times \overline{H_{g,h,est}}$, and $\bar{B} = \frac{\overline{H_{g,h,est}} - 0.1 \times \overline{H_{g,h,est}}}{\sqrt{(\bar{A} - \overline{H_{g,h,est}})^2 + (\bar{B} - \overline{H_{g,h,est}})^2}}$. Their combined error is $\Delta C = \sqrt{(\bar{A} - \overline{H_{g,h,est}})^2 + (\bar{B} - \overline{H_{g,h,est}})^2}$. Figure 6 provides these propagation errors for all the sites and both models. It is clearly seen that the propagation error in the CPRG models is higher than that for the CPR ones. This is interpreted as another drawback of the former models.

The possible source of extra errors in the CPRG models in comparison with that for the CPR ones may be the non-linear factor f_c introduced in the CPRG formulation (see Equation 7). Indeed, if f_c is taken equal to 1 for a moment, the CPRG expression coincides with the CPR one. Therefore, a probable improvement for the CPRG model's performance would be the re-definition or modification of the f_c factor.

3.5 Coefficient of correction

In Sections 3.1–3.3 the superiority of the CPR model to the CPRG one was shown. This fact comes from the linearity (along the 1:1 line) of the former and the quadratic behaviour (away from the 1:1 line) of the latter model. Therefore, the present Section deals with the CPR model only.

At this stage, a question may be raised about how to improve further the performance of the CPR model universally. For this reason, the ratios of the monthly averages of $H_{g,m,obs}$ to the $H_{g,m,est}$ values were considered for each site. These monthly ratios, named coefficients of correction (CoC), are shown in Figure 7 for all the 12 sites. It is seen that almost all CoC curves fall within the monthly standard deviations except for the BOU and LER sites, which were



found in the previous sections not to perform equally well to those of the other 10 sites; indeed, the CoC values of January, November and December for the BOU and LER sites fall outside of the standard deviation limits in these 3 months. Nevertheless, this may not mean a catastrophe, and, therefore, the universal CoCs can be adopted as are.

The CoC values for all the sites were afterwards used in three ways: (i) by multiplying the estimated hourly $H_{g,h,est}$ values by the annual average CoC = 0.99 calculated from all sites, (ii) by multiplying the estimated hourly $H_{g,h,est}$ values by the particular monthly averages of CoC (0.97, 0.99, 0.99, 1, 1, 0.99, 0.99, 1, 1, 1, 0.99, 0.96, respectively for January to December) calculated from all sites, and (iii) by multiplying the estimated hourly $H_{g,h,est}$ values by using the individual monthly averages of CoC for each site. New $H_{g,h,est}$ values were then derived. The outcome of this process is presented in Figure 8, which shows the annual mean $H_{g,a}$ (subscript a implies annual) values for the observed, the before-correction-estimated and the after-corrections-estimated values.

The differences in the annual means upon the various estimated values at the sites in comparing them to the observed values are seen to be almost negligible (see Figure 8). To quantify these differences, Table 6 shows them in both Wm^{-2} and in %. The highest precision in the estimated values is provided by the application of the individual monthly mean CoC values for each site. This gives estimated values equal to the observed ones (column 4, Table 6). Nevertheless, the calculation of the monthly CoC values for every site becomes a tedious work and, in many aspects, not so wise. An acceptable accuracy is given by the application of an annual universal value of CoC = 0.99 (column 3, Table 6), which gives very comparable results to the application of monthly mean universal CoC values (column 5, Table 6). This last conclusion is considered acceptable and may be used worldwide.

A last issue concerning the variation of the annual mean and standard deviation values of CoC for all the sites as function of the absolute geographical latitude (meaning that the – southern latitude of DAA becomes positive) are shown in Figures 9, 10, respectively. Non-linear fits to the data points in both plots have been introduced.

There is, though, a guess that these fits may become flat in the range $0^\circ < |\varphi| < 23.5^\circ$ if more sites are added in this zone. This conclusion is in line with a similar recent finding by Kambezidis et al. (2023) for the correction factor (CF), as function of $|\varphi|$; the CF has been defined as the ratio of energy received on an inclined plane by taking into account a near-real ground albedo to that under the constant albedo value of 0.2. It is interesting to note that the zone $[-23.5^\circ, +23.5^\circ]$ lies between the Tropic of Capricorn in the southern hemisphere and the Tropic of Cancer in the northern one. The average CoCs in Figure 9 show a decline after the latitude of $|\varphi| > 23.5^\circ$; this can be interpreted as a lesser need to correct the CPR-estimated solar radiation values as an increase in $|\varphi|$ is associated with fewer clear-sky (CS) days and, therefore, lower CS solar radiation levels. On the contrary, the decrease in CS events with an increase in $|\varphi|$ results in a higher dispersion of the CS and non-CS events, thereby increasing $\sigma(\text{CoC})$.

4 Discussion

The performance of two DGRSD models that can generate hourly global horizontal solar radiation ($H_{g,h}$) values at a site from recorded daily values under all-sky conditions was thoroughly investigated in this work. The CPR (Collares-Pereira and Rabl) and the CPRG (Collares-Pereira, Rabl and Gueymard) were the two models chosen. Twelve global locations were chosen for this purpose in order to demonstrate whether or not both models are applicable everywhere. Since the goal was to demonstrate the effectiveness of the models rather than to provide climatological results, specific years for each place were utilised rather than a range of years.

Several statistical indices, including RMSE, MBE, MAE, R^2 , and d , at the 99.9% CI, were employed to assess the models' performance. The hourly $H_{g,h}$ ratios were calculated by multiplying the hourly r_{CPR} and r_{CPRG} ratios (Equations 1, 6, respectively) by the daily $H_{g,d}$ values.

The observed and calculated hourly $H_{g,h}$ values were contrasted. The CPR models produced $R^2 < 0.90$ for 8 of the 12 sites (BOU, CAR, GAN, ILO, KIS, LER, PAY, and REG). However, while the MBEs of the 8 sites were not significantly higher than those of the other 4, the RMSE and MAE values of the aforementioned 8 models (particularly BOU, ILO, and REG) were higher. The BOU site had the lowest accuracy ($R^2 = 0.76$). However, the result of $d \approx 1$ for every site showed perfect agreement between the estimated and observed values. In summary, the CPR methodology appears to be quite effective for the ATH, DAA, LIN, and SOV locations at the hourly level. Overall, the CPR models appear to perform well across all locations, with BOU showing a lower performance. Out of the 12 sites, only 2 (DAA and REG) have CPRG models that yielded $R^2 > 0.90$. These two sites' RMSE, MBE, MAE, and R^2 statistical index values were found lower than those of the CPR models and the identical sites. However, $d \approx 1$ at every site, which is comparable to the CPR models' outcome. The LER site had the lowest accuracy ($R^2 = 0.14$). In summary, the CPR modelling appears to be more effective than the CPRG methodology at the hourly level for all sites.

At every site and in both models, the hourly $H_{g,h}$ values that were observed and calculated were transformed into daily $H_{g,d}$ values. For the 12 locations and both models, twelve scatter plots were created for the $H_{g,m,est} - H_{g,m,obs}$ pairs. For the CPR and CPRG models, linear

and quadratic fits were obtained, respectively. Only the data pairs pertaining to CPR were discovered to be along the 1:1 line, despite the fact that both regression fits had extremely high R^2 values. In summary, it was discovered that the CPR models performed nearly flawlessly; however, the CPRG models were unable to replicate the daily values, particularly for the BOU and LER sites.

The monthly level was the focus of the most recent assessment of the chosen DGSRD models. For both models, the hourly mean $H_{g,h}$ data at each site was transformed into monthly mean $H_{g,m}$ values. Regression fits were derived to the $H_{g,m,est} - H_{g,m,obs}$ data pairs, and scatter plots were created similarly to the daily-scale case. Finding that the data at every site falls along the 1:1 lines confirmed once more that the CPR model outperforms the CPRG one.

The following was shown by an error analysis of the $H_{g,h,est}$ hourly values at each site and in both DGSRD models. While the PE for the CPRG model was found to be higher, indicating a poorer performance compared to the CPR model, the percent error for the CPR model fell within the acceptable range of $\pm 5\%$ about the individual annual $H_{g,h,est}$ average. The standard error analysis of the $H_{g,h,est}$ data set revealed that, for the same location and model, the mean value of any subsequent $H_{g,h,est}$ data set will probably fall within the range $\overline{H_{g,h,est}} \pm SE$. At every site, a propagation-of-error study revealed a greater mistake in the CPRG modelling. A potential explanation for the aforementioned errors may be attributed to the extra f_c factor introduced in the expression of the CPRG model.

The coefficient of correction (CoC) was defined as the ratio between the observed to the estimated solar radiation values. This was carried out to look into the potential for further enhancing the CPR models' performance. CoCs were determined using 3 methods: (i) an overall annual average for all sites, (ii) monthly averages for each site, and (iii) overall monthly averages for all sites. For the 3 CoC methods, the differences between the yearly mean values of $H_{g,a,obs}$ and $H_{g,a,est}$ were computed in both Wm^{-2} and %. The differences were found negligible, thus concluding that the overall annual average of $CoC = 0.99$ can be used universally to bring the hourly CPR-derived values closer to the observed ones. In practical terms, the universal CoC value is useful in the transformation of hourly, daily or monthly global horizontal solar radiation data that can be obtained via satellite observations, solar modelling or data re-analysis at regions that do not have access to these types of measurements. Naturally, even though this CoC value is universal, it could change if other sites are added in a future study.

Ultimately, the current study showed that the CPR model is highly practical and applicable globally. However, in order to confirm its universality, the methodology created in this work should be applied to more locations worldwide in future research. Particularly, if new sites are chosen in the zone $0^\circ < |\phi| < 30^\circ$, this could improve the effectiveness of the CPR models and validate the conclusion made in Section 3.5 with a potential explanation of the significance of this occurrence using machine-learning techniques.

Data availability statement

Publicly available datasets were analysed in this study. This data can be found at: Baseline Solar Radiation Network (<https://bsrn.awi.de/data/data-retrieval-via-pangaea/>) and National Observatory of Athens (<https://www.iersd.noa.gr/en/services/parochi-klimatikon-dedomenon/>, upon request).

de/data/data-retrieval-via-pangaea/) and National Observatory of Athens (<https://www.iersd.noa.gr/en/services/parochi-klimatikon-dedomenon/>, upon request).

Author contributions

HDK: Conceptualisation, Data curation, Methodology, Validation, Visualisation, Writing–original draft, Writing–review and editing.

Funding

The author declares that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The author is especially grateful to the following scientists: (i) S. Wacker (LIN) and A. Aculinin (KIS) for providing formulated data from their BSRN stations for years 2018 and 2020, respectively; (ii) A. Driemel (World Radiation Monitoring Centre) for providing access to the PANGAEA website (<https://bsrn.awi.de/data-retrieval-via-pangaea>, accessed on 1 July 2021) and downloading data from 10 BSRN stations. Data from the ATH station has been possessed by the author as member of the Institute where the station belongs to and as the responsible scientist for the station's smooth operation until 2014. Also acknowledged are the following scientists who are/were responsible for their BSRN stations used in this study: S. Morris BOU; J.F. Morel CAR; J. Botai DAA; P. Kumar GAN; T.O. Aro ILO; G. Hodgetts LER; L. Vuilleumier PAY; D. Halliwell REG; M. Olefs SON; N. Al-Abadi SOV.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Baig, A., Akhter, P., and Mufti, A. (1991). A novel approach to estimate the clear day global radiation. *Renew. Energy* 1 (1), 119–123. doi:10.1016/0960-1481(91)90112-3
- Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorological Soc.* 95 (9), 1431–1443. doi:10.1175/BAMS-D-13-00047.1
- Boscaini, R., Robaina, A. D., Peiter, M. X., Bruning, J., Rodrigues, S. A., da Silva, J. G., et al. (2020). Performance of solar radiation models for obtaining reference evapotranspiration at Santa Maria-RS, Brazil. *Rev. Bras. Ciências Agrar.* 15 (1), 1–8. doi:10.5039/AGRARIA.V15I1A7661
- Christie, M. A., Glimm, J., Grove, J. W., Higdon, D. M., Sharp, D. H., and Wood-Schultz, M. M. (2005). Error analysis and simulations of complex phenomena. *Los Alamos Sci.* 29, 6–25.
- Collares-Pereira, M., and Rabl, A. (1979). The average distribution of solar radiation-correlations between diffuse and hemispherical and between daily and hourly insolation values. *Sol. Energy* 22 (2), 155–164. doi:10.1016/0038-092X(79)90100-2
- Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., et al. (2018). Baseline surface radiation network (BSRN): structure and data description (1992–2017). *Earth Syst. Sci. Data* 10 (3), 1491–1501. doi:10.5194/essd-10-1491-2018
- El shazly, S. M. (1996). Estimation of hourly and daily global solar radiation at clear days using an approach based on modified version of Gaussian distribution. *Adv. Atmos. Sci.* 13, 349–358. doi:10.1007/bf02656852
- Garg, H. P., and Garg, S. N. (1987). Improved correlation of daily and hourly diffuse radiation with global radiation for Indian stations. *Sol. Wind Technol.* 4 (2), 113–126. doi:10.1016/0741-983X(87)90037-3
- Giesen, R. H., van den Broeke, M. R., Oerlemans, J., and Andreassen, L. M. (2008). Surface energy balance in the ablation zone of Midtdalsbreen, a glacier in southern Norway: interannual variability and the effect of clouds. *J. Geophys. Res.* 113 (D21), D21111. doi:10.1029/2008JD010390
- Gueymard, C. (1986). Mean daily averages of beam radiation received by tilted surfaces as affected by the atmosphere. *Sol. Energy* 37 (4), 261–267. doi:10.1016/0038-092X(86)90043-5
- Gueymard, C. (2000). Prediction and performance assessment of mean hourly global radiation. *Sol. Energy* 68 (3), 285–303. doi:10.1016/S0038-092X(99)00070-5
- Jain, P. C. (1984). Comparison of techniques for the estimation of daily global irradiation and a new technique for the estimation of hourly global irradiation. *Sol. Wind Technol.* 1 (2), 123–134. doi:10.1016/0741-983X(84)90014-6
- Jain, P. C. (1988). Estimation of monthly average hourly global and diffuse irradiation. *Sol. Wind Technol.* 5 (1), 7–14. doi:10.1016/0741-983X(88)90085-9
- Kambezidis, H. D. (2021). The solar radiation climate of Greece. *Climate* 9 (12), 183. doi:10.3390/cli9120183
- Kambezidis, H. D. (2022). “The solar resource.” *Compr. Renew. Energy* Editor T. M. Letcher, 26–117. doi:10.1016/B978-0-12-819727-1.00002-9
- Kambezidis, H. D., Kavadias, K. A., and Farahat, A. M. (2024). Solar energy received on flat-plate collectors fixed on 2-Axis trackers: effect of ground albedo and clouds. *Energies* 17, 3721–3727. doi:10.3390/en17153721
- Kambezidis, H. D., Mimidis, K., and Kavadias, K. A. (2022). Correction of the solar azimuth discontinuity at sunrise and sunset. *Sun Geosph.* 15 (1), 39–44. doi:10.31401/SunGeo.2022.01.04
- Kambezidis, H. D., Mimidis, K., and Kavadias, K. A. (2023). The solar energy potential of Greece for flat-plate solar panels mounted on double-Axis systems. *Energies* 16 (13), 5067. doi:10.3390/en16135067
- Kambezidis, H. D., and Tsangrassoulis, A. E. (1993). Solar position and right ascension. *Sol. Energy* 50 (5), 415–416. doi:10.1016/0038-092X(93)90062-S
- Kambezidis, H. D. D., and Papanikolaou, N. S. S. (1990). Solar position and atmospheric refraction. *Sol. Energy* 44 (3), 143–144. doi:10.1016/0038-092X(90)90076-O
- Khatib, T., Mohamed, A., and Sopian, K. (2012). A review of solar energy modeling techniques. *Renew. Sustain. Energy Rev.* 16 (5), 2864–2869. doi:10.1016/j.rser.2012.01.064
- Larsen, K. S., Ibrom, A., Beier, C., Jonasson, S., and Michelsen, A. (2007). Ecosystem respiration depends strongly on photosynthesis in a temperate heath. *Biogeochemistry* 85 (2), 201–213. doi:10.1007/s10533-007-9129-8
- Liu, B. Y. H., and Jordan, R. C. (1960). The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Sol. Energy* 4 (3), 1–19. doi:10.1016/0038-092X(60)90062-1
- Newell, T. A. (1983). Simple models for hourly to daily radiation ratio correlations. *Sol. Energy* 31 (3), 339–342. doi:10.1016/0038-092X(83)90024-5
- Nwokolo, S. C., Amadi, S. O., Obiwulu, A. U., Ogbulezie, J. C., and Eyibio, E. E. (2022). Prediction of global solar radiation potential for sustainable and cleaner energy generation using improved Angstrom-Prescott and Gumbel probabilistic models. *Clean. Eng. Technol.* 6 (May 2021), 100416. doi:10.1016/j.clet.2022.100416
- Taylor, J. R. (1982). *An introduction to error analysis*. 2nd edition, Vol. 101. Sausalito, CA: University Science Books.
- Walraven, R. (1978). Calculating the position of the sun. *Sol. Energy* 20 (5), 393–397. doi:10.1016/0038-092X(78)90155-X
- Whillier, A. (1956). The determination of hourly values of total solar radiation from daily summations. *Arch. Für Meteorol. Geophys. Und Bioklimatol. Ser. B* 7 (2), 197–204. doi:10.1007/BF02243322
- Yao, W., Li, Z., Xiu, T., Lu, Y., and Li, X. (2015). New decomposition models to estimate hourly global solar radiation from the daily value. *Sol. Energy* 120, 87–99. doi:10.1016/j.solener.2015.05.038

Nomenclature

Abbreviations

AI	artificial intelligence
amsl	above mean sea level
ANN	artificial neural network
ASNOA	actinometric station of the National Observatory of Athens
BSRN	baseline surface radiation network
CI	confidence interval
CoC	coefficient of correction
CPR model	Collares-Pereira and Rabl model
CPRG model	Collares-Pereira, Rabl and Gueymard model
CS	clear skies
DGSRD model	daily global solar radiation decomposition model
LST	local standard time
MAE	mean-absolute error
MBE	mean-bias error
PE	percent error
RMSE	root mean-square error

SE	standard error
UTC	universal time coordinated

Symbology

γ	solar altitude (degrees)
δ	solar declination (degrees)
λ	geographical longitude (degrees)
σ	standard deviation
ϕ	geographical latitude (degrees)
ψ	solar azimuth (degrees)
ω	solar hour angle (degrees)
d	index of agreement
H_g	global horizontal solar irradiance (Wm^{-2})
R^2	coefficient of determination
N	number of data points
z	altitude (m)