SIMPLIFIED MODEL TO CORRECT THERMOPILE PYRANOMETER SOLAR RADIATION MEASUREMENTS FOR PHOTOVOLTAIC MODULE YIELD ESTIMATION

Lillo-Bravo I[1], Larrañeta M[2], Núñez-Ortega E[1], González-Galván R[1]

[1] Department of Energy Engineering, University of Seville.

[2] Andalusian Association for Research and Industrial Cooperation (AICIA).

Corresponding author:

Miguel Larrañeta, Andalusian Association for Research and Industrial Cooperation, Camino de los Descubrimientos, s/n. 41092, Seville, Spain.

Phone: (+34)954487237

E-mail: mlarraneta@gter.es

Abstract

This article presents and evaluates the performance of a simplified model to generate 10-min global horizontal synthetic solar radiation data that would correspond to the measurements of a calibrated photovoltaic monocrystalline cell. The model, which only requires global horizontal solar radiation data measured with a thermopile pyranometer as input, is based on the characterization of the relation between the data measured with a thermopile pyranometer and a calibrated cell as a function of the sky condition and the solar elevation. We have used an extensive solar radiation database for the location of Seville (Spain) for the training of the model that has been tested in Seville and Lancaster (USA), showing satisfactory results and suggesting a global applicability with no local adaptation or calibration requirement.

The model shows the best results for high levels of solar radiation and solar elevations and decreases its performance on days with high levels of diffuse irradiation and for very low solar elevation angles. We obtain a daily RMSD between measured and synthetic data of 1.9% in Seville and 5.2% in Lancaster. The frequency distribution of the synthetic datasets shows a KSI of 3.7 W/m^2 in Seville and 8.6 W/m^2 in Lancaster. We also evaluate the ramp rates of measured and synthetic sets through the KSI of the measured and synthetic ramp rates sets, obtaining 0.11 W/m^2•min in Seville and 0.20 W/m^2•min in Lancaster.

Keywords

Photovoltaic yield, calibrated solar cell, pyranometer.

1 Introduction

Photovoltaic (PV) modules have a practically identical spectral response, time, temperature coefficient and angular response to a calibrated solar cell of the same technology and encapsulation. For this reason, if radiation measurements are obtained by means of a calibrated photovoltaic cell of the same technology and encapsulation as the photovoltaic modules of the installation, most of the uncertainties in the determination of the electrical production of the photovoltaic installation are already included. Meydbray et al., [1-2] compare the use of solar reference cell and thermopile for PV applications, and point out that solar reference cells are especially useful for precise characterization of the PV performance, allowing a better detection of changes in PV system performance with time and shorter time assessment of the PV operating efficiency. Table 1 summarises the comparison of the use of solar reference cell and thermopile pyranometers for PV applications proposed by Meydbray et al [2].

Table 1. Cell-pyranometer comparison made by Meydbray et al [2] for PV applications.

	Reference Cell	**Thermopile pyranometer**
Spectral Response	Can be made to closely match solar panel	Broadband response needs to be corrected
Angle of incidence	Can be made to closely match solar panel	Response to all angles
Temperature response	Temperature response is similar to PV system	Are designed to minimize sensitivity to temperature. Not corrected for temperature
Time response	< milliseconds; matched to PV response	Up to 30 seconds, can be problematic for measuring PV performance
Other issues		Emission to cold sky and transients in ambient temperature affects output
International standards	IEC 60904	ISO 9847, ISO 9845, ISO 9846

For PV applications, Dunn et al., [3] analysed the expanded uncertainties in the measurements of the irradiance made in the plane of array throughout the course of a representative day with a thermopile pyranometer and a PV reference device. They conclude that during most of the day, uncertainties are in the order of ± 5% for a pyranometer, and ±2.4% for a calibrated cell, both stated with 95% confidence intervals. So calibrated cells provide superior irradiance measurements for PV power plant monitoring applications. Haeberlin et al., [4] compare the use of a pyranometer and a calibrated silicon solar cell and reach similar conclusions: it is preferable to use a calibrated Si-cell (of the same type as the cells in the PV array) as a reference device of the irradiance measurements for PV plant monitoring than a pyranometer.

The PV module performance is generally evaluated by manufacturers under standard test conditions (STC), which refer to 25°C module temperature, 1000 W/m^2 of incident solar irradiance on the PV module plane with 1.5 solar spectrum air mass (AM) as reference spectrum (IEC 60904-3, 2016) [5]. Nevertheless, in real conditions the PV module temperature, solar irradiance with incident angle and solar spectrum differ from STC.

Incident solar global irradiance that reaches the PV module is affected by the distribution of its solar spectrum. This solar spectrum depends on cloud cover, AM, precipitable water (PW) and aerosol optical depth (AOD) [6]. The short-circuit current is the main parameter affected by the distribution of the solar spectrum, but in some technologiesefficiency, maximum power and fill factor are also influenced. Polo et al, [7] analyse the use of reference modules as irradiance sensor for monitoring and modelling rooftop PV systems for different PV technologies with different spectral responses. If such affection is neglected and only the broadband irradiance is used for PV-module performance modelling, then errors of up to 10% can be introduced in extreme conditions [8].

Therefore, a spectral correction shall be used in order to adjust the incident solar irradiance and, hence, improve the PV module performance modelling [9]. This spectral correction depends on the PV module technology. Polo et al., [10] give the spectral factor for seven photovoltaic technologies and 124 sites. They found that the annual spectral factor for crystalline silicon technologies is rather homogenous worldwide with maxima spectral losses and gains of ≈3% and ≈1%, respectively. On the contrary, the spectral factor for thin film devices displays a higher spatial variability. Nuñez et al., [11] proposed a spectral matching ratio for multijunction cells within a concentrating photovoltaic module. The Sandia Array Performance Model applies a fourth-order polynomial correction based on AM [12]. M. Lee et al., [13] included in their model the influence of AM and PW in a similar function. Theristis et al., [14] developed a model including AM, PW and AOD to improve concentrating photovoltaic system performance modelling. Another option is to use direct measurements of spectral radiation, but the use of spectroradiometers is not yet widespread and an adequate validation is required [15].

Moreover, in many cases the PV module structure does not have two-axis tracking. Consequently, the incident solar global irradiance that reaches the PV module is affected by the solar angle of incidence (AOI). Angular losses in PV modules can introduce differences in the short circuit current under STC of up to 3.5% under global normal irradiance conditions with horizontal orientation compared with a device without losses due to angular effects [16]. King et al., [17] suggest the use of a 5th order polynomial function to represent the angular optical losses on the short circuit current. Martin et al., [18] formulated similar functions. Instead, given the importance of these effects on the performance of the photovoltaic module, manufacturers test their PV modules according to IEC standard 61853-2:2016 [19] where the effect of the angle of incidence and the spectral responsivity on the output power of the photovoltaic module is measured. The most used software for the design of photovoltaic installations has implemented a model to take into account the angular losses. PVsyst software [20] has implemented the model developed by Souka and Safat and adopted by the American Society of Heating, Refrigeration, and Air Conditioning (ASHRAE) [17]. Plag F et al., [21] related the spectral and angular effect showing that the spectral effect depends on the incident angle too.

An alternative to these methods to link the silicon photovoltaic calibrated cell measurement with the measurements made by a thermopile pyranometer are specific simplified models to relate the global solar radiation measured by both devices. J. J. Michalsky et al., [22] propose a table of correction factors depending on the clearness index and the brightness index. It requires the knowledge of diffuse and direct radiation from the site. King & Myers [23] correct the response of the PV device as a function of the solar spectrum, AOI and temperature. This corrected response allows us to obtain an improvement in the estimation of the total irradiance. The method also requires diffuse and direct data among the calculation of functions related to the absolute air mass, the AOI and a temperature coefficient. The main drawback in the implementation of the described methods is the requirement of rarely available measurements such as AOD, PW, diffuse and direct irradiance for the location under study.

The purpose of this work is the development of a simplified model for estimating the solar global radiation that a calibrated monocrystalline silicon cell would measure using only thermopile pyranometer measurements as input. This model could be used as a correction factor that would allow us to characterize the PV module performance from thermopile pyranometer radiation data.

Simplicity is the main advantage of this model because it depends only on the global irradiation data measured by a thermopile pyranometer avoiding the need to use more complex variables which are generally more difficult to obtain. The model has been developed for one location (Seville) and tested in the same location (Seville) and a different one (Lancaster) without any local adaptation showing satisfactory results.

The paper is presented as follows: Section 2 presents the database used for the training and testing of the model. In Section 3 we explain step by step the development of the model. In Section 4 we perform the result analysis of the model in two locations. Conclusions are then made in Section 5.

2 Meteorological database

In this work, an extensive database is used for training the method proposed (Table 2). This database is composed of 10-min averages values of GHI recorded during 42 consecutive months from July 2012 to December 2015 for the location of Seville (Spain). The measurements were taken with a sampling and storage time resolution of 5 s. A secondary standard pyranometer Kipp & Zonen CMP21 and an Atersa calibrated monocrystalline silicon PV cell measured the GHI. The devices used for the design and testing of the methodology are located at the meteorological station of the Group of Thermodynamics and Renewable Energy of the University of Seville and have been periodically calibrated, according to the recommendations from the instrument manufacturers. Data used in this work have been subjected to quality-control procedures [24] following the BSRN recommendations [25]. Only 1% of the data has been filled for the entire database. Data recorded at sun altitude lower than 2º have not been used in this study.

Table 2. Location selected for the method training.

	Latitude (°N)	Longitude (°W)	Altitude (m)	Köppen-Geiger Climate	Period
Seville	37.4	6.0	12	Csa	2012-2015

In addition, the model has been tested in two locations. We use data for one entire year for the location of Seville and for the location of Lancaster, USA (Table 3). GHI data from Lancaster were measured with a secondary standard pyranometer Kipp & Zonen CMP21 and an Atersa calibrated monocrystalline silicon PV cell, and has been subjected to quality control procedures

[24] following the BSRN recommendations [25]. According to the Köppen-Geiger climate classification system [26], Seville's climate is classified as Mediterranean with hot summers and Lancaster's climate is classified as Mediterranean with mild summers. When performing the quality control procedure of Lancaster database, we identify a period of 12 consecutive days with irregularities in the measurements in October. This period is not computed in the performance evaluation of the algorithm.

We use 10-min averaged data because high-quality data for the testing of the method was available in that time step, but the method can be used in any time resolution greater than 1-min.

Table 3. Location selected for testing the method.

	Latitude (°N)	Longitude (°W)	Altitude (m)	Köppen-Geiger Climate	Period
Seville	37.4	6.0	12	Csa	2016
Lancaster	34.6	118.3	812	Csb	2016

3 Methodology

In this section we describe the methodology implemented for the synthetic generation of reference PV cell solar radiation data (G_{cell}^{synth}) from pyranometer measurements (G_{pyr}). The model has been developed for 10-min time resolution and its only input is the G_{pyr}. The rest of the used parameters are theoretically estimated from the geographic position of the location under study (latitude and longitude) and the record time of the measurements. The method can be described in a sequence of 5 steps.

3.1 Step 1: Data clustering

In the first step, we cluster the data from pyranometer and PV cell measurements as a function of the clearness index k_t and the solar elevation, α. To characterize the type of day, other indices such as the Perez clearness index (ε), Perraudeau's brightness (I_N), sky ratio index (SR) [27] or illuminance fluctuation frequency index [28] could be used, but they require the knowledge of more variables such as diffuse and/or beam normal irradiance. Previous approaches have shown that using only one radiometric variable, different sky conditions can be identified [29]. We propose to cluster the data into 9 groups in terms of the clearness index, dividing the datasets into intervals of 0.1 from 0 to 0.8. Values greater than 0.8 have been clustered into the same group. For the solar elevation we chose 7 groups dividing the dataset into intervals of 5° from 20° to 40° and intervals of 10° from 40° to 60°. Data for solar elevations lower than 20° are clustered in the same group as well as data for solar elevations greater than 60°. We group the solar data into a total of 63 clusters.

3.2 Step 2: First regression

In the second step, we use a least squares procedure to perform a linear regression fit to the cloud point obtained, when comparing the solar radiation values measured with the pyranometer (G_{pyr}) with the solar radiation values measured with the calibrated solar cell (G_{cell}^{meas}), obtaining a function dependant on the solar elevation (α) and sky condition defined by the clearness index (k_t). See equation 1. In figure 1 we present an example of the cloud points and the linear fit for the cluster interval corresponding to solar elevations lower than 30° and greater than 25°, and clearness index lower than 0.6 and greater than 0.5. In table 4 we present the first regression coefficient RC (α, kt) obtained for each cluster from the training dataset.

$$RC(\alpha, k_t) = {G_{cell}^{meas}}/{G_{pyr}} \qquad (1)$$

We can relate the first regression coefficient (*RC*) to the spectral and angular losses related to the measurements of solar radiation with silicon monocrystalline cells in comparison to solar radiation data measured with a pyranometer, since those losses show a strong dependency on the solar elevation and sky condition.

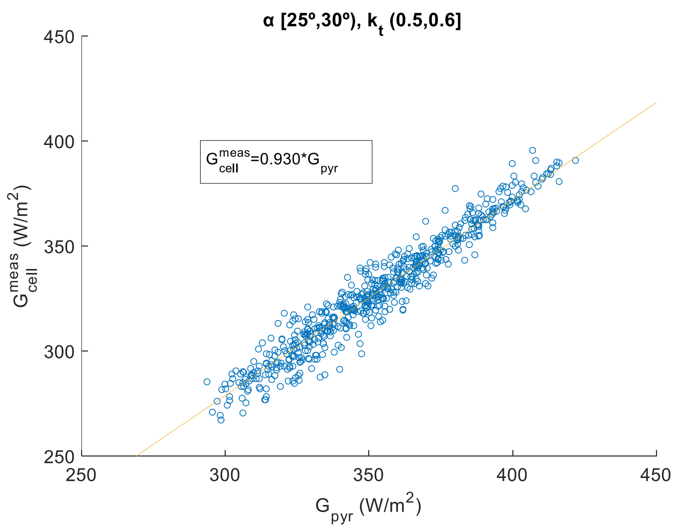


Figure 1. Cloud points and first linear regression fit for solar elevations between 25° and 30° and clearness indexes between 0.5 and 0.6 for the location of Seville.

Table 4. First regression coefficient (*RC*) for each cluster for the location of Seville. We shade the result for the data exposed in figure 1.

1st Regression coefficient *RC*(α, k_t)	α < 25°	25° ≤ α < 30°	30° ≤ α < 35°	35° ≤ α < 40°	40° ≤ α < 50°	50° ≤ α < 60°	α ≥ 60°
$k_t \leq 0.1$	1.050	1.052	1.065	1.059	1.062	1.058	1.049
$0.1 < k_t \leq 0.2$	1.000	1.008	1.017	1.022	1.022	1.025	1.016
$0.2 < k_t \leq 0.3$	0.966	0.976	0.981	0.981	0.987	0.990	0.984
$0.3 < k_t \leq 0.4$	0.938	0.957	0.961	0.967	0.967	0.970	0.970
$0.4 < k_t \leq 0.5$	0.891	0.944	0.950	0.954	0.960	0.963	0.964
$0.5 < k_t \leq 0.6$	0.854	0.930	0.944	0.951	0.956	0.960	0.960
$0.6 < k_t \leq 0.7$	0.854	0.921	0.936	0.946	0.953	0.958	0.959
$0.7 < k_t \leq 0.8$	0.862	0.925	0.939	0.949	0.956	0.960	0.960
$k_t > 0.8$	0.882	0.932	0.944	0.950	0.954	0.956	0.955

In figure 2 we present the first regression coefficient (*RC*) for all the sky conditions as a function of the solar elevation selected intervals (a) and the first regression coefficient (*RC*) vs the solar elevations as a function of the sky condition (b)

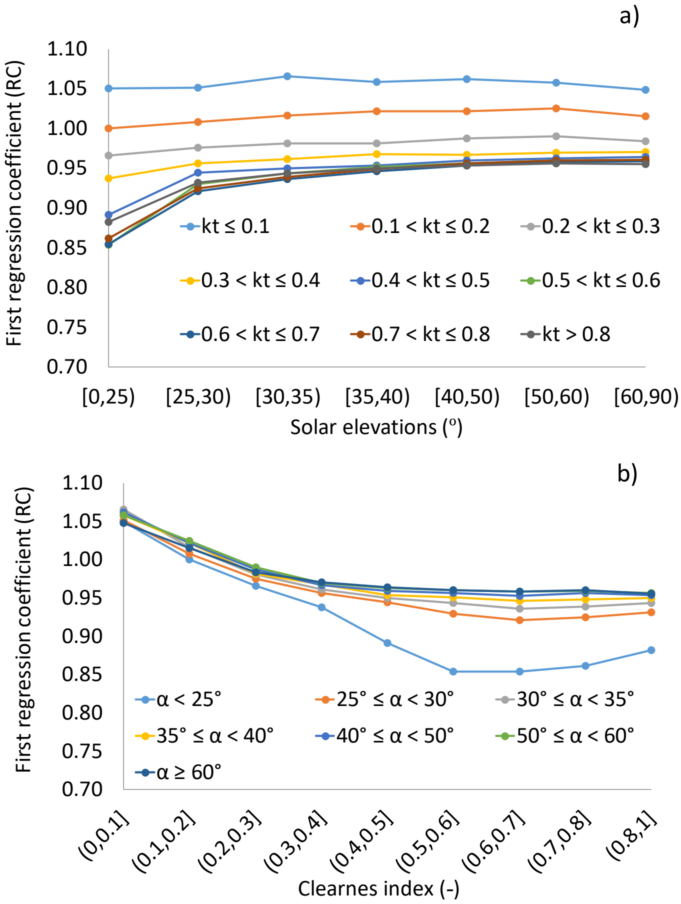


Figure 2. First regression coefficient (*RC*) versus all the selected intervals of solar elevations as a function of the clearness indexes (a) for the location of Seville. First regression coefficient (*RC*) versus the clearness indexes for all the selected intervals of solar elevations as a function of all the selected intervals of solar elevations (b) for the location of Seville.

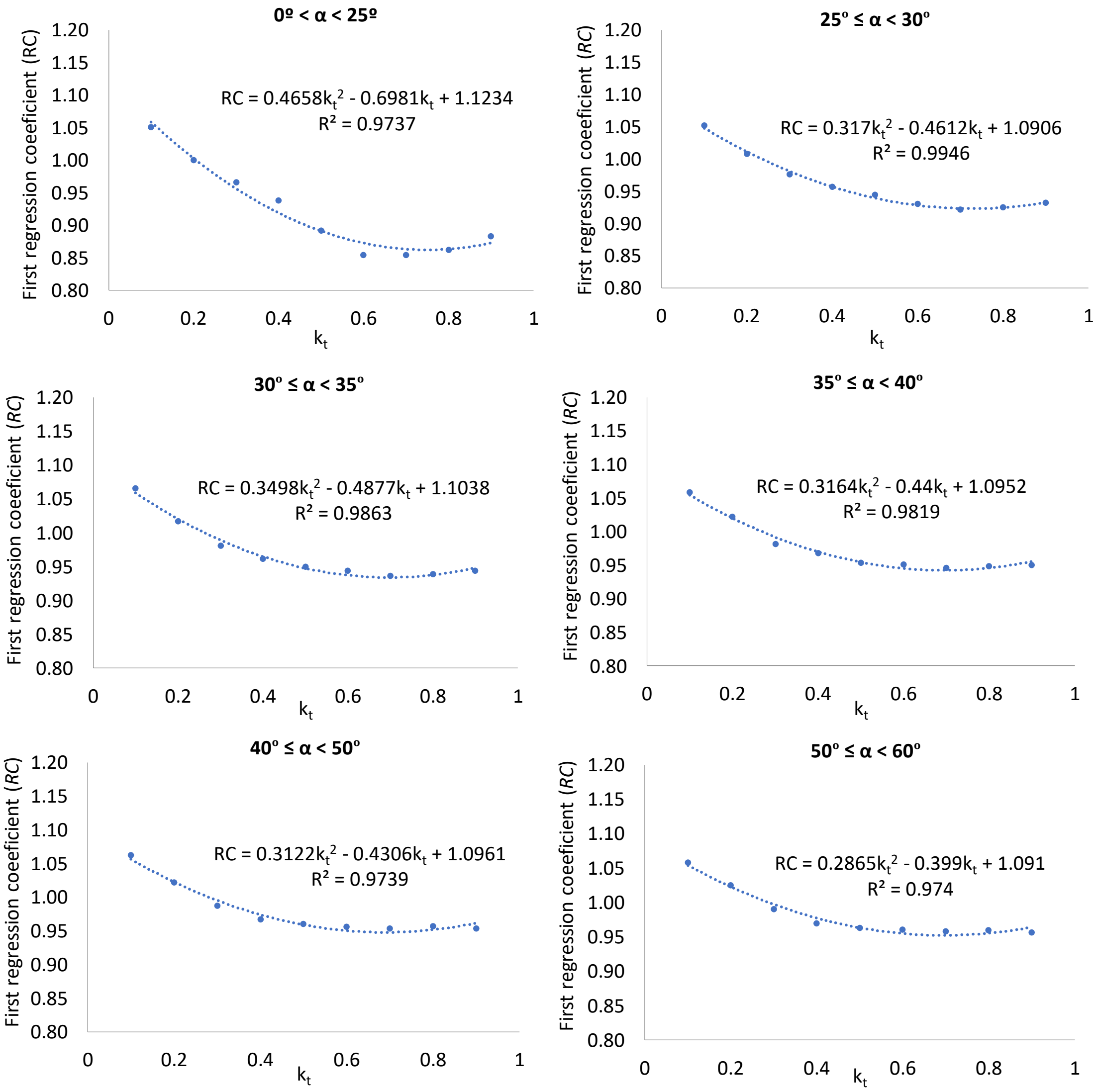
From table 4 and figure 2, we can observe that the first regression coefficient (RC) shows a different tendency for solar elevations lower than 25° and clearness indexes greater than 0.4. For k_t values lower than 0.4 and solar elevations lower than 25°, we can observe in both figures a, and b, that RC shows a change in its performance with respect to the rest of the k_t and α values.

This can be explained by the greater impact of the refraction effects for low solar elevations of the solar radiation on the PV reference cell in comparison to the pyranometer. Reference cell errors are particularly large at sunset and sunrise (up to 50 W/m^2) according to the Hukseflux Thermal Sensors report [30]. The greater the clearness index, the greater the direct component amount on the total global radiation which in turn is strongly affected by the refraction effects due to low solar elevations and horizon distortions. Moreover, the PV reference cells present a greater angular dependence.

3.3 Step 3: Second regression

In the third step, we run again a least squares procedure but in this case we perform a second order polynomic fit of the first regression coefficient (*RC*) as a function of the clearness index (k_t). The calculation is performed separately for each solar elevation selected interval following equation 2. From this step, we obtain the three functions $(b(\alpha), c(\alpha), d(\alpha))$ dependant on the solar elevation. In figure 3 we represent the polynomial fit of the first regression coefficient (*RC*) as a function of the clearness index (k_t) for each solar elevation interval.

$$RC(\alpha, k_t) = b(\alpha) \cdot {k_t}^2 + c(\alpha) \cdot k_t + d(\alpha) \qquad (2)$$



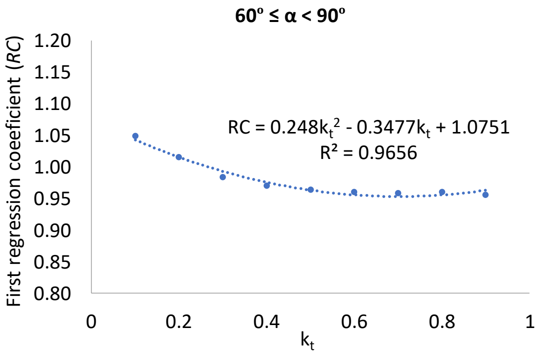


Figure 3. Polynomial fit of the first regression coefficient (*RC*) as a function of the clearness index (k_t) for each solar elevation interval for the location of Seville

The coefficient of determination ranges between 0.974 and 0.995 in the solar elevation selected intervals showing a high accuracy on the fit. In table 5 we present the functions obtained applying equation 2 for each solar elevation interval

Table 5. Second regression coefficients, $(b(\alpha), c(\alpha), d(\alpha))$ for each solar elevation interval for the location of Seville.

α (°)	b(α)	c(α)	d(α)
< 25	0.466	-0.698	1.123
25 - 30	0.317	-0.461	1.091
30 - 35	0.350	-0.488	1.104
35 - 40	0.316	-0.440	1.095
40 - 50	0.312	-0.431	1.096
50 - 60	0.287	-0.399	1.091
> 60	0.248	-0.348	1.075

3.4 Step 4: Third regression

In this step, we run again a least squares procedure to perform a linear regression fit, but in this case we fit the second regression coefficients, $(b(\alpha), c(\alpha), d(\alpha))$ obtained in step 3 (Table 5) as a function of the solar elevation defined intervals following equations 3-5.

$$b(\alpha) = b_1 \cdot \alpha + b_2 \quad (3)$$

$$c(\alpha) = c_1 \cdot \alpha + c_2 \quad (4)$$

$$d(\alpha) = d_1 \cdot \alpha + d_2 \quad (5)$$

We run the third regression fit only for solar elevations greater than 25°. In figure 3 we present the third regression linear fit of the second regression coefficients, $(b(\alpha), c(\alpha), d(\alpha))$ obtained in step 3 as a function of the solar elevation (greater than 25°).

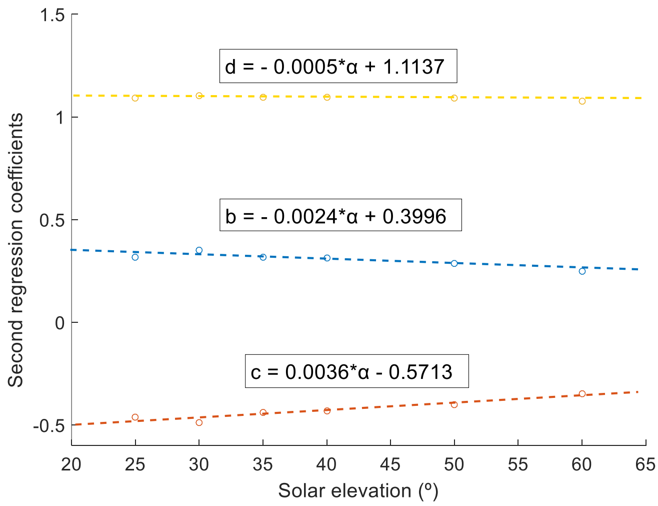


Figure 4. Linear fit of the second regression coefficients $(b(\alpha), c(\alpha), d(\alpha))$ as a function of the solar elevation for the location of Seville.

3.5 Step 5: Hyperbolic tangent fit

When substituting $(b(\alpha), c(\alpha), d(\alpha))$ from equation 3-5 into equation 2 separating calculations for solar elevations lower and greater than 25°, we obtain equations 6 and 7 respectively.

$$RC = 0.466 \cdot {k_t}^2 - 0.698 \cdot k_t + 1.123 \qquad if\ \alpha < 25° \tag{6}$$

$$RC = (-0.0024 \cdot \alpha + 0.3996) \cdot {k_t}^2 + (0.0036 \cdot \alpha - 0.5713) \cdot k_t + (-0.0005 \cdot \alpha + 1.1137) \qquad if\ \alpha \geq 25° \tag{7}$$

The model could be used from its corresponding equation (6 -7) according to the solar elevation, but it would lead to a discontinuity at the solar elevation α=25°. We use the hyperbolic tangent function to join equations 6 -7 in a simple equation (see equation 8) obtaining thus a continuous function.

$$RC(\alpha, k_t) = \Big(0.5 \cdot \big\{\big(0.466 \cdot {k_t}^2 - 0.698 \cdot k_t + 1.123\big) \cdot \big(1 - tanh\big(\beta \cdot (\alpha - 25)\big)\big) + \big((-0.0024 \cdot \alpha + 0.3996) \cdot {k_t}^2 + (0.0036 \cdot \alpha - 0.5713) \cdot k_t + (-0.0005 \cdot \alpha + 1.1137)\big) \cdot \big(1 + tanh\big(\beta \cdot (\alpha - 25)\big)\big)\big\}\Big) \tag{8}$$

β is a parameter that modifies the connection between equations 6 and 7. For a high β value, the connection is more abrupt. For a low β value, the connection is smoother, but the data for solar elevations different to 25º change significantly. We select a β=0.18 by minimizing the deviations between the reference PV cell measured and synthetic solar radiation sets from the Seville training database. Any value between 0.1 and 0.2 can be applied and no significant variation would be obtained. We recommend using β=0.18 for any location where the model may be applied.

By substituting equation 8 in equation 1 we can obtain the synthetic reference PV cell solar radiation data G_{cell}^{synth} from thermopile pyranometer measurements G_{pyr} following equation 9

$$G_{cell}^{synth} = \Big(0.5 \cdot \big\{\big(0.466 \cdot {k_t}^2 - 0.698 \cdot k_t + 1.123\big) \cdot \big(1 - tanh\big(0.18 \cdot (\alpha - 25)\big)\big) + ((-0.0024 \cdot \alpha + 0.3996) \cdot {k_t}^2 + (0.0036 \cdot \alpha - 0.5713) \cdot k_t + (-0.0005 \cdot \alpha + 1.1137)) \cdot \big(1 + tanh\big(0.18 \cdot (\alpha - 25)\big)\big)\big\}\Big) \cdot G_{pyr} \tag{9}$$

4 Results and discussion

To assess the performance of the model, we evaluate the mean, distribution and autocorrelation of the synthetically generated time series G_{cell}^{synth} obtained from pyranometer measurements G_{pyr} in comparison to the data measured with the reference cell G_{cell}^{meas} in each of the test datasets (Table 3).

In figure 5 we represent the pyranometer measured solar radiation in continuous orange, the reference PV cell measured solar radiation in discontinuous blue and the synthetic solar radiation in dotted green, of four selected days throughout the year in different sky conditions for the location of Seville on the figures on the left. We also present the deviation between the pyranometer ($Error_{pyr}$) and the PV calibrated cell measurements in continuous red and the deviation between the measured and synthetic calibrated cell data in dotted green on the right ($Error_{cell}^{synth}$). The deviations are calculated following equations 10 and 11.

$$Error_{pyr} = 100 \cdot \left({G_{pyr} - G_{cell}^{meas}} \middle/ {G_{cell}^{meas}} \right) \tag{10}$$

$$Error_{cell}^{synth} = 100 \cdot \left({G_{cell}^{synth} - G_{cell}^{meas}} \middle/ {G_{cell}^{meas}} \right) \tag{11}$$

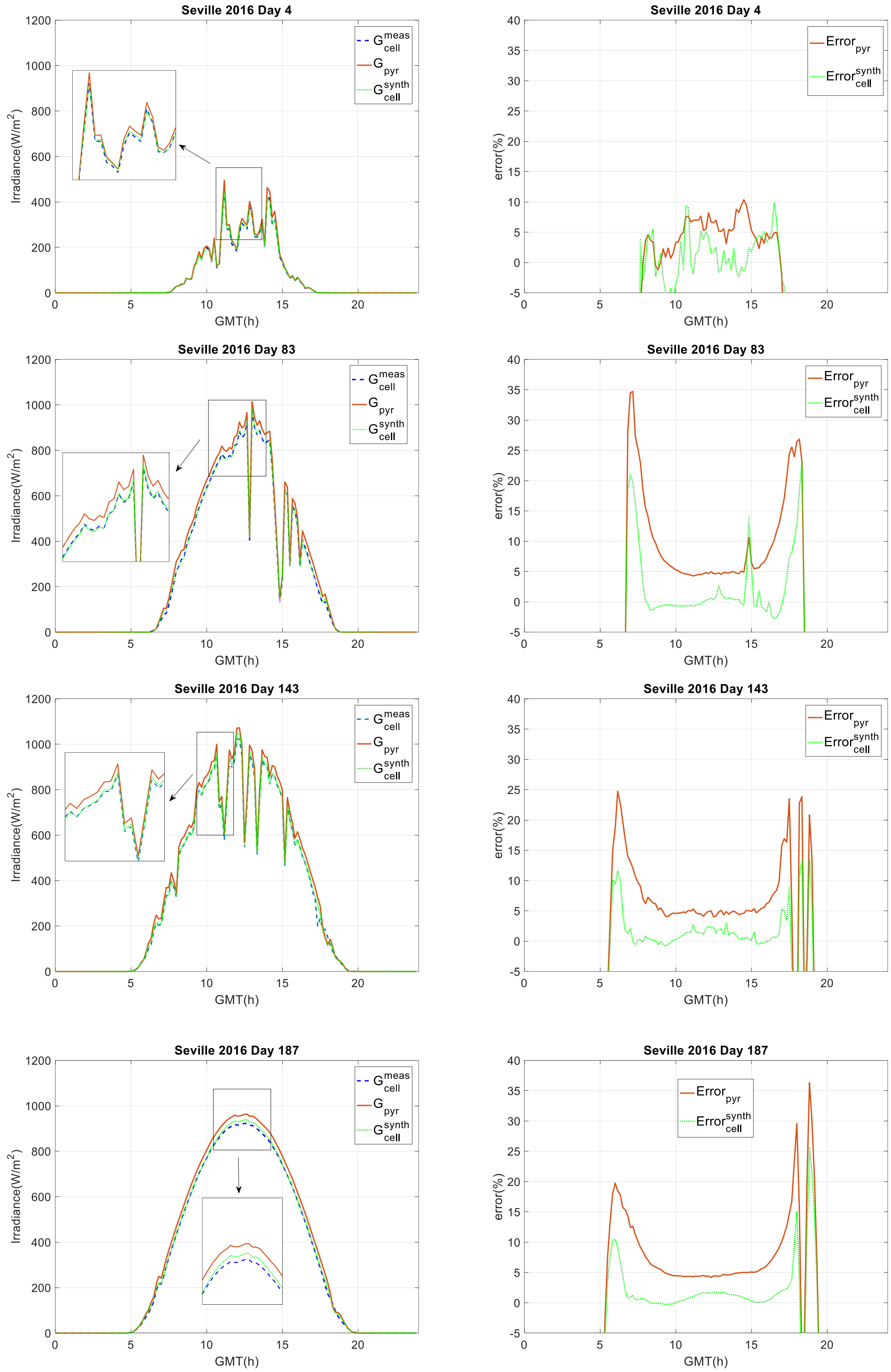
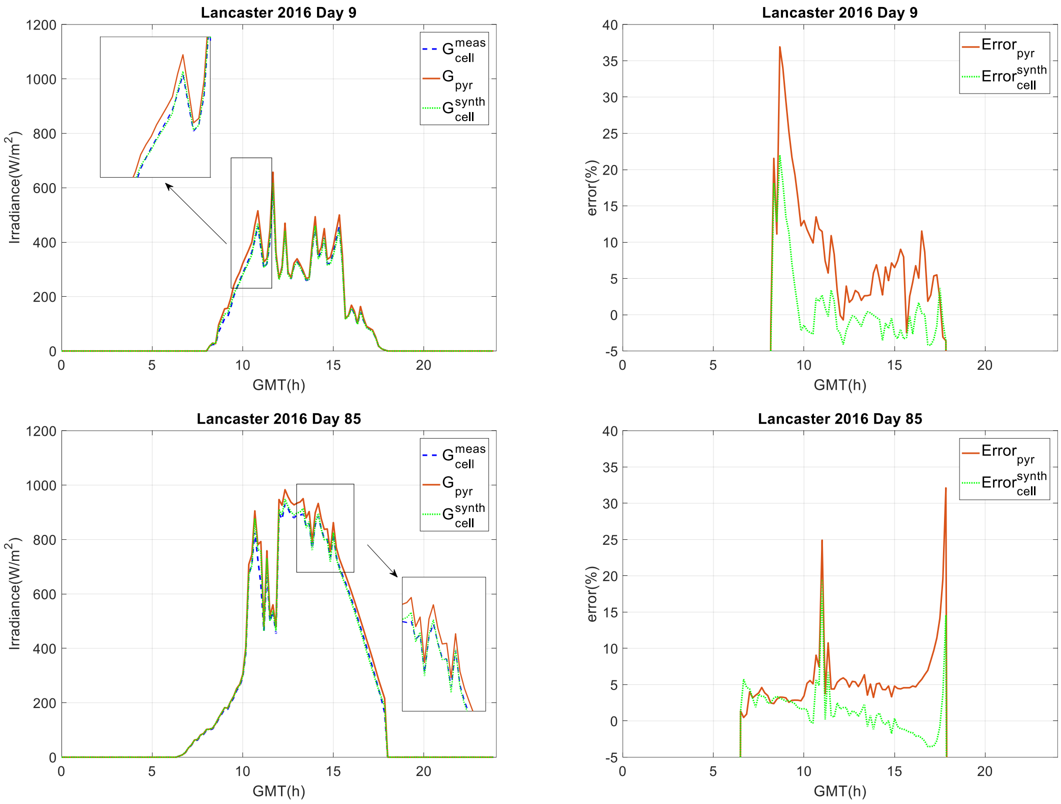


Figure 5. Four daily 10-min solar radiation profiles for the location of Seville. G_{pyr} is the solar radiation measured with the pyranometer, G_{cell}^{meas} is the solar radiation measured with the PV reference cell and G_{cell}^{synth} is the synthetically obtained PV reference cell data. On the left, the daily profiles, on the right the instant errors.

From day 187 of figure 5 we can observe that the greater error values are found for low solar elevations at clear sky conditions α<5° mainly caused by high incidence angles. It is noticeable that on cloudy days, such as the one presented in day 4 in figure 5, this tendency is not observed. From day 4 we can observe that on cloudy days, with a greater amount of DHI, the differences between both errors is lower than on sunny days, meaning that the model shows worse performance on cloudy days, possibly due to the spectral distribution of the solar radiation that may drop most of the energy in the accepted wavelengths range of the cell on cloudy days. According to Nann et al., [31] clouds act somewhat as aneutral density filter upper wavelength limit of the VIS region of the spectrum. At the upper wavelength limit of VIS and in the NIR region, clouds are strong absorbers of radiation in selected wavelength bands, due to increased water-vapour absorption and by liquid-water absorption.

In any case, synthetic data errors ($Error_{cell}^{synth}$) are lower than pyranometer data errors ($Error_{pyr}$) for all sky conditions and for a wide range of solar elevations and solar radiation levels.

In figure 6 we perform the same evaluation for Lancaster. The error of the synthetic data ($Error_{cell}^{synth}$) is lower than the error when comparing pyranometer to calibrated cell data ($Error_{pyr}$) for all the sky conditions. In any case, the approach is more accurate for clear sky conditions and greater solar elevations. We can only find errors in the synthetic data greater than 5% for solar elevations lower than 5°. It should be noted that solar radiation data at low elevations have negligible impact on the photovoltaic module yield.



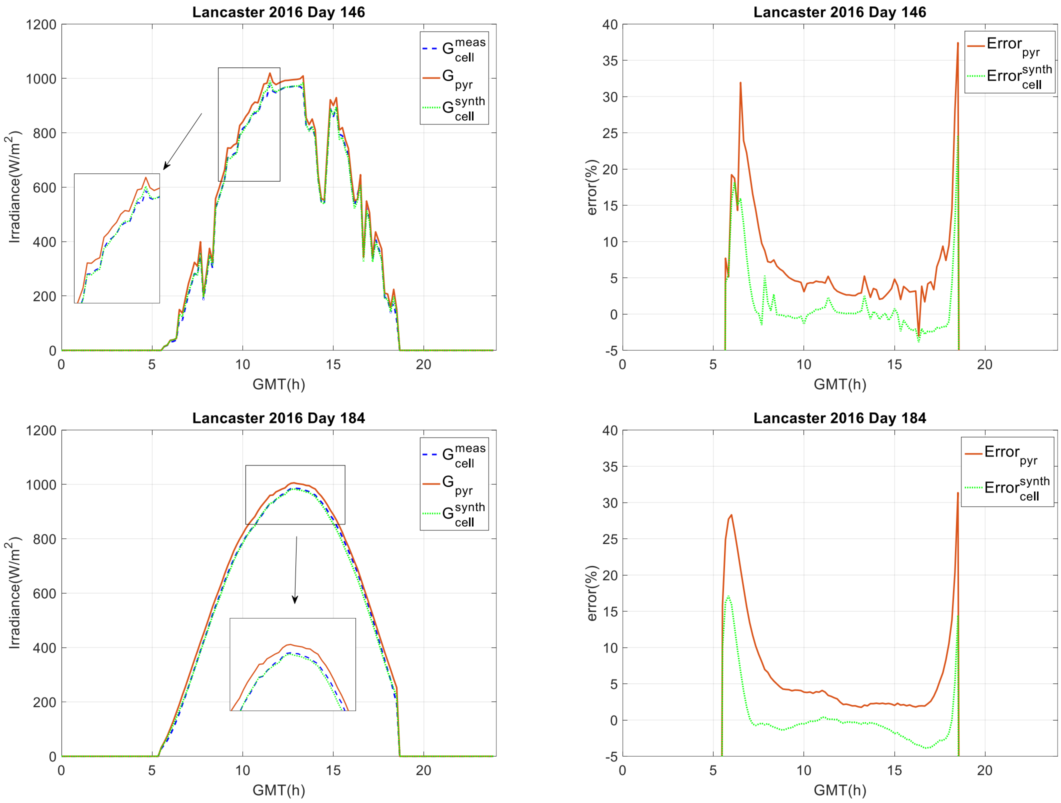


Figure 6. Four daily 10-min solar radiation profiles for the location of Lancaster. G_{pyr} is the solar radiation measured with the pyranometer, G_{cell}^{meas} is the solar radiation measured with the PV reference cell and G_{cell}^{synth} is the synthetically obtained PV reference cell data. On the left, the daily profiles, on the right the instant errors.

In figure 7 we present the daily cumulative values of G_{pyr}, G_{cell}^{meas} and G_{cell}^{synth} for the location of Seville and Lancaster. In Lancaster there is a period with no available data covering 12 days in October. In the correct data, we can observe that the synthetically obtained daily cumulative data is more similar to the PV reference measured solar radiation data than the pyranometer measurements daily cumulative data.

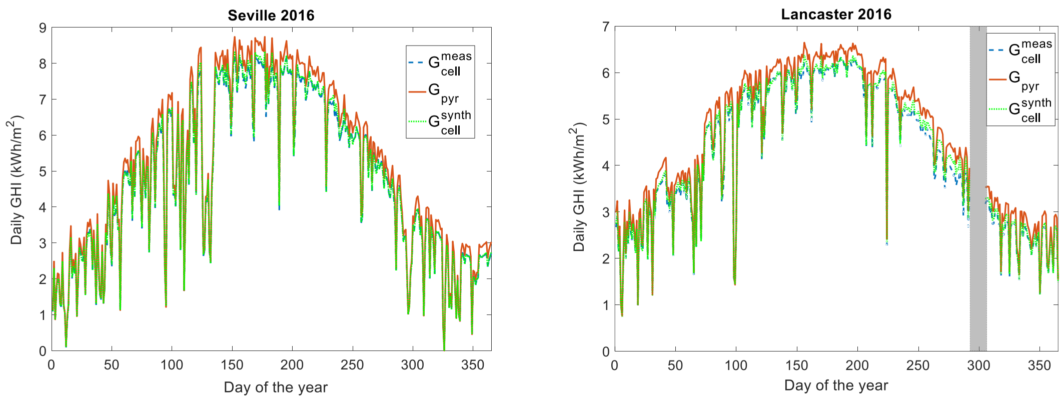


Figure 7. Daily cumulative solar radiation values for the location of Seville (left) and Lancaster (right). G_{pyr} is the solar radiation measured with the pyranometer, G_{cell}^{meas} is the solar radiation

measured with the PV reference cell and G_{cell}^{synth} is the synthetically obtained PV reference cell data.

In tables 6-7 we present the monthly and annual cumulative values of the measured and synthetic time series for one year for the location of Seville and Lancaster respectively with the percentage differences in comparison to the reference PV cell measured data. We identified continuous irregularities on twelve consecutive days in October in Lancaster. We have not included this month in the performance evaluation of the model. We can observe that the synthetically generated data show lower differences in the monthly cumulative values in comparison to the monthly cumulative values of the reference PV cell-measured dataset varying from -0.2% to 1.4% in Seville and from 0.3% to 3.3% in Lancaster, while the monthly cumulative values of the pyranometer measurements show differences that range from 5.3% to 9.5% in Seville and 4.4% to 11.1% in Lancaster. We can observe that the greater differences in the pyranometer monthly data in comparison to the cell-measured monthly data are found from September to March coinciding with the time of the year with the lower solar elevations.

Table 6. Monthly and annual cumulative values of the synthetic sets and their differences in comparison to the measured reference cell data for the location of Seville.

Dataset	Unit	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
G_{cell}^{meas}	kWh/m^2	62	87	154	159	195	229	229	210	165	114	77	70	1751
G_{pyr}	kWh/m^2	66	93	164	167	206	243	242	224	175	121	83	77	1861
	$Error_{pyr}$	7.0	7.1	6.5	5.3	5.7	6.1	5.8	6.5	5.7	6.2	7.5	9.5	6.3
G_{cell}^{synth}	kWh/m^2	62	87	154	160	197	232	231	212	165	114	76	70	1760
	$Error_{cell}^{synth}$	-0.2	0.4	0.2	0.4	1.1	1.4	0.9	1.0	-0.4	-0.5	-0.7	-0.1	0.5

Table 7. Monthly and annual cumulative values of the synthetic sets and their differences in comparison to the measured reference cell data for the location of Lancaster.

Dataset	Unit	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year(*)
G_{cell}^{meas}	kWh/m^2	74	95	122	150	174	180	185	166	136	-	80	71	1432
G_{pyr}	kWh/m^2	80	102	130	159	181	190	194	176	149	-	89	78	1527
	$Error_{pyr}$	7.8	8.0	6.1	5.6	4.4	5.6	4.7	6.4	9.4	-	11.1	9.3	6.6
G_{cell}^{synth}	kWh/m^2	75	96	123	152	174	182	186	169	141	-	82	72	1452
	$Error_{cell}^{synth}$	0.8	1.9	1.0	0.9	0.3	1.4	0.4	1.6	3.7	-	3.3	1.3	1.4

(*)October not included in the total sum

The increased differences in Lancaster may be attributed to the good training of the model for Seville (through the multi-step regression analysis) and the use of an empirical parameter developed for Seville.

4.1 Mean evaluation

We evaluate the deviations between the modelled and measured PV reference cell data using the Root Mean Squared Deviation ($RMSD$) calculated following equation 12. We also calculate the RMSD of the pyranometer data in comparison to the calibrated cell-measured data in order to show a base to compare (equation 13).

$$\mathrm{RMSD}_{cell} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(G_{cell}^{meas} - G_{cell}^{synth}\right)^2} \tag{12}$$

$$\mathrm{RMSD}_{pyr} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(G_{cell}^{meas} - G_{pyr}\right)^2} \tag{13}$$

where N is the number of data pairs, G_{cell}^{meas} is the global horizontal irradiance measured with the reference PV cell, G_{cell}^{synth} is the global horizontal irradiance synthetically obtained from pyranometer measurements and G_{pyr} is the solar radiation measured with the pyranometer. The evaluation is performed in the 10-min and daily resolution. Only data for solar elevations greater than 2° are considered in this analysis. In table 8 we present the RMSD_{cell} and in table 9 the RMSD_{pyr} in the 10-min and daily resolution for the test datasets. We also present the RMSD in % by dividing the RMSD by the average value (equations 14 and 15).

$$\mathrm{RMSD}_{cell}(\%) = 100 \cdot \left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(G_{cell}^{meas} - G_{cell}^{synth}\right)^2} \Big/ \overline{G_{cell}^{meas}} \right) \tag{14}$$

$$\mathrm{RMSD}_{pyr}(\%) = 100 \cdot \left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(G_{cell}^{meas} - G_{pyr}\right)^2} \Big/ \overline{G_{cell}^{meas}} \right) \tag{15}$$

For Seville, the average value ($\overline{G_{cell}^{meas}}$) in the 10-min resolution is 391 W/m^2 and in the daily resolution is 4.7 kWh/m^2. For Lancaster, the average value $\overline{(G_{cell}^{meas}}$) in the 10-min resolution is 371 W/m^2 and in the daily resolution is 3.8 kWh/m^2.

Table 8. RMSD_{cell} in the 10-min and daily resolution for Seville and Lancaster test datasets.

Station (year)	$\mathrm{RMSD}_{10\text{-}min}$		RMSD_{daily}	
	W/m^2	%	kWh/m^2	%
Seville (2016)	9.5	2.4	0.06	1.4
Lancaster (2016)	19.7	3.8	0.09	2.5

Table 9. RMSD_{pyr} in the 10-min and daily resolution for Seville and Lancaster test datasets.

Station (year)	$\mathrm{RMSD}_{10\text{-}min}$		RMSD_{daily}	
	W/m^2	%	kWh/m^2	%
Seville (2016)	29.6	7.5	0.33	6.9
Lancaster (2016)	41.7	8.1	0.29	7.7

The RMSD between measured data from the reference cell and pyranometer, mainly related to spectral and angular losses, reach 8% on a daily basis. Moreover, when applying the correction algorithm, the RMSD between measured data from the reference cell and synthetic data is reduced to 2%. In figure 8 we present the daily G_{cell}^{meas} vs the daily G_{pyr} together, with the daily G_{cell}^{synth} in Lancaster in the figure on the left and Seville in the figure on the right in a scatter plot. We can observe how the BIAS is reduced significantly from values around 6% to values lower than 2%. Reduction is even higher in Seville.

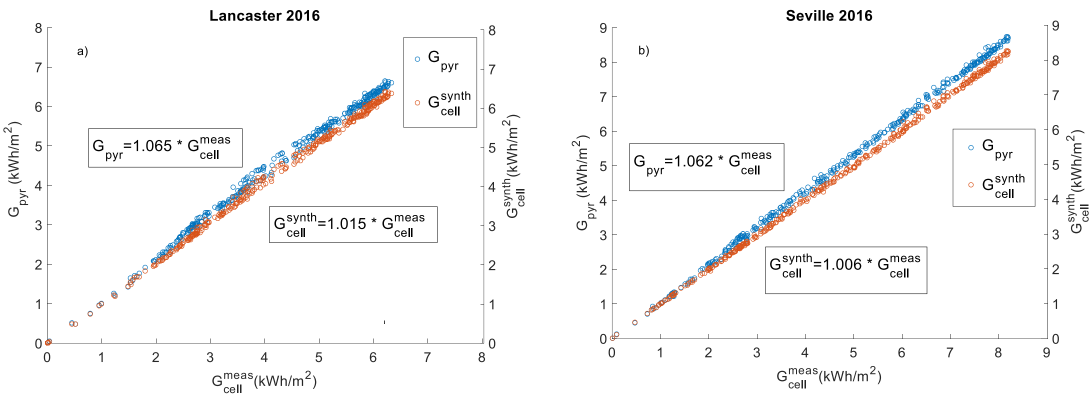


Figure 8. Scatter plot of the daily cumulative solar radiation values measured with the cell versus the daily solar radiation values measured with the pyranometer and the synthetically generated data for the location of Lancaster (a) and Seville (b).

4.2 Distribution evaluation

The distribution is evaluated since it is assumed that differences in these terms may lead to differences in PV plant production. To evaluate the distribution, we calculate the KSI (Kolmogorov-Smirnov test integral, equation 16) index defined as the integrated differences between the CDFs of two datasets. It is a widely used index to compare cumulative distributions of measured and synthetic solar radiation sets [32]. The unit of this index is the same for the corresponding magnitude. We only analyse data for solar elevations greater than 2º.

$$KSI = \int_{x_{min}}^{x_{max}} D_n dx, \tag{16}$$

where, x_{max} and x_{min} are the extreme values of the independent variable, and D_n are the differences between the CDFs of datasets evaluated in 100 points. The higher the KSI values, the greater the differences in the CDFs of the evaluated datasets. We calculate the KSI as a comparison of the CDFs of the solar radiation sets measured with the pyranometer to the solar radiation sets measured with the PV reference cell in one hand, and the synthetically obtained PV reference cell solar radiation sets to the solar radiation sets measured with the PV reference cell in the other hand. This calculation is performed in order to quantify the impact of the developed model when used as a correction factor for spectral and angular losses of silicon cells. In Figure 9 we represent the CDFs of the measured and synthetic sets for Seville and Lancaster 2016.

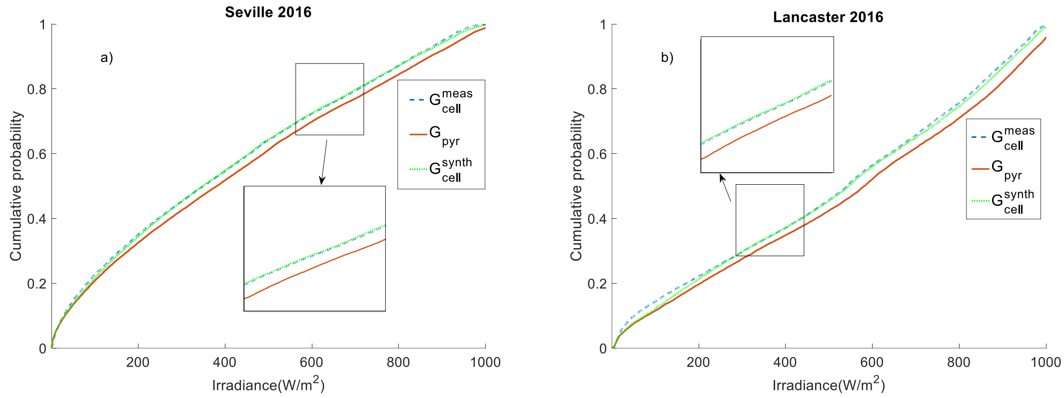


Figure 9. CDFs of the measured and synthetic sets for Seville 2016 (a) and Lancaster 2016 (b). G_{pyr} is the solar radiation measured with the pyranometer, G_{cell}^{meas} is the solar radiation measured with the PV reference cell and G_{cell}^{synth} is the synthetically obtained PV reference cell data.

In Table 10 we present the KSI value (W/m^2) of the evaluated datasets. The measurements with pyranometer and reference PV cell present different CDFs. However, the synthetically generated reference PV cell data present a significant improvement towards similarity to the measured PV cell CDF.

Table 10. KSI (W/m^2) for Seville and Lancaster test datasets.

Parameter	Station (year)	G_{pyr}	G_{cell}^{synth}
KSI (W/m^2)	Seville (2016)	24.6	3.7
	Lancaster (2016)	35.6	8.6

The similarity on the measured and synthetic reference PV cell data can be quantified with the KSI. It is reduced by 85% for the location of Seville and by 75% in Lancaster. The model has been trained with more than three years of measurements from the location of Seville, therefore its performance is slightly better for Seville than for Lancaster. In any case, it is worth highlighting that the model can be globally applied without any local adaptation. Obviously the errors are lower for the location where the model has been trained.

4.3 Autocorrelation evaluation

There is a relation between consecutive values that should be maintained when generating synthetic solar radiation data. To that end, we evaluate the autocorrelation qualitatively by calculating and representing the ramp rates (RRs), defined as the difference between successive data points over 10 minutes (equation 17).

$$RR = {(GHI_k - GHI_{k-1})}/{\Delta t} \tag{17}$$

where Δt refers to an interval of 10 minutes. The units will be given in W/m^2·min. We calculate the absolute RR values for the annual datasets taking into account only daytime observations. We qualitatively evaluate the RRs of each dataset through the CDFs (Figure 10) and quantitatively through the Kolmogorov Smirnov Index of the CDFs of the Ramp rates (KSI_{RR}) (Table 11).

Figure 10. CDFs of the RRs of the measured and synthetic sets for Seville 2016 (a) and Lancaster 2016 (b). G_{pyr} is the solar radiation measured with the pyranometer, G_{cell}^{meas} is the solar radiation measured with the PV reference cell and G_{cell}^{synth} is the synthetically obtained PV reference cell data.

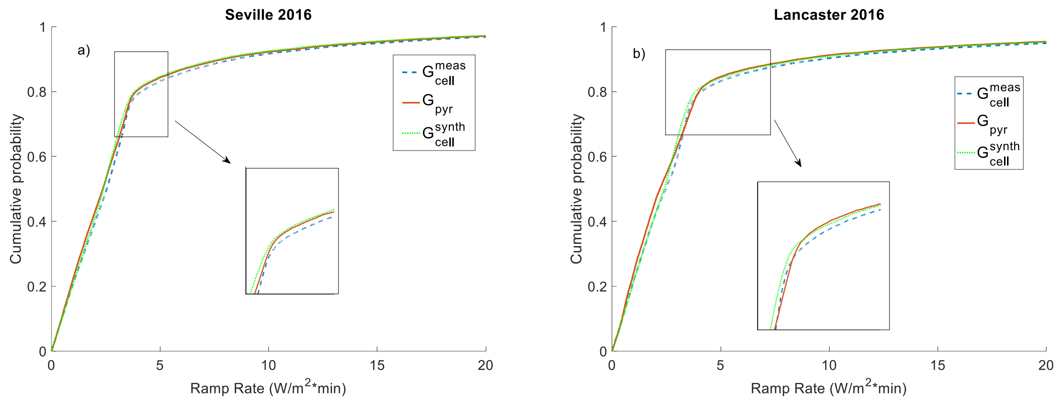


Table 11. KSI_{RR} (W/m^2·min) for Seville and Lancaster test datasets.

Parameter	Station (year)	G_{pyr}	G_{cell}^{synth}
KSI_{RR} (W/m^2·min)	Seville (2016)	0.22	0.11
	Lancaster (2016)	0.41	0.20

From the RRs evaluation we can observe a reduction in the KSI_{RR} of around 50% in both locations from G_{pyr} to G_{cell}^{synth}. The model, which has been trained with data measured in Seville, presents therefore a better performance in terms of autocorrelation for the location of Seville than for the location of Lancaster, but we can highlight its global applicability without any local adaptation. Nevertheless, more locations should be evaluated to assess the performance of the model in a wider range of climates.

5 Conclusions

In this paper, we present a simple model to obtain reference cell 10-min solar radiation data from thermopile pyranometer measured data series with reasonable accuracy. Our model does not require any other input data besides the pyranometer measurements and can be globally applied with no local adaptation or calibration. We have trained the model in one location (Seville, Spain) and applied it in two locations, Seville and Lancaster (USA), obtaining satisfactory results. The model is developed from an initial cluster of the training data and several regressions, taking into account the main parameters affecting the angular and spectral losses; the solar elevation and the clearness index. At the time of using it, only thermopile pyranometer data and a few geometrical calculations are required. This way, the model shows great simplicity facilitating its applicability.

The model shows better performance on sunny days with high solar elevations, when most of the energy is obtained. We have evaluated the mean, distribution and autocorrelation of the synthetic sets in comparison to measured data with a reference cell using measured data with a pyranometer as input. Daily RMSD is reduced from 7-8% to 1-2% when evaluating a complete annual set. The frequency distribution of the synthetic sets is also improved. It has been quantified in an 80% reduction of the KSI. The autocorrelation is quantified through the KSI of the RRs obtaining also a reduction of 50%. The model shows better performance for the location of Seville, where it has been trained, since solar radiation depends on local phenomena. Future works will focus on the improvement of the model by using data from more different climates for its training, and tests will focus on the evaluation of its global applicability.

Acknowledgments

The authors are grateful to Quintas Energy, S.A for providing the data from Lancaster.

References

[1] J. Meydbray, K. Emery. Pyranometers and reference cells, what's the difference? NREL/JA-5200-54498, March 2012.

[2] J. Meydbray, R. Riley, L. Dunn, K. Emery. Pyranometers and reference cells: part 2: what makes the most sense for PV power plants. NREL/JA-5200-56718, October 2012.

[3] L. Dunn, M. Gostein, and K. Emery, "Comparison of Pyranometers vs. Reference Cells for Evaluation of PV Array Performance," Proceedings of the 38th IEEE Photovoltaic Specialists Conference (PVSC), Austin, TX, June 3-8, 2012. 2012 IEEE.

[4] H. Haeberlin and Ch. Beutler. Comparison of pyranometer and si-reference cell solar irradiation data in long term PV plant monitoring. 13th EU PV Conference on Photovoltaic Solar Energy Conversion, Nice, France, 1995. Pages 1-4.

[5] IEC 60904-3: 2016. Photovoltaic devices - Part 3: Measurement principles for terrestrial photovoltaic (PV) solar devices with reference spectral irradiance data

[6] B. Marion, "Influence of atmospheric variations on photovoltaic performance and modelling their effects for days with clear skies," in Proc. 38th IEEE Photovoltaic Spec. Conf., 2012, pp. 3402–3407

[7] Polo, W.G. Fernandez-Neira, M.C. Alonso-García. On the use of reference modules as irradiance sensor for monitoring and modelling rooftop PV systems, Renewable Energy, Volume 106, 2017, Pages 186-191, ISSN 0960-1481

[8] P. Faine, S. R. Kurtz, C. Riordan, and J. M. Olson, "The influence of spectral solar irradiance variations on the performance of selected single-junction and multijunction solar-cells," Sol. Cells, vol. 31, pp. 259–278, Jun. 1991

[9]. Alonso-Abella, M., Chenlo, F., Nofuentes, G., Torres-Ramírez, M., 2014. Analysis of spectral effects on the energy yield of different PV (photovoltaic) technologies: the case of four specific sites. Energy 67, 435–443.

[10] Jesús Polo, Miguel Alonso-Abellá, Jose A. Ruiz-Arias, José L. Balenzategui. Worldwide analysis of spectral factors for seven photovoltaic technologies, Solar Energy, Volume 142, 2017, Pages 194-203.

[11] Núñez Júdez, Rubén; Domínguez Domínguez, César; Askins, Stephen; Victoria Pérez, Marta; Herrero Martin, Rebeca; Antón Hernández, Ignacio y Sala Pano, Gabriel 2015. Determination of spectral variations by means of component cells useful for CPV rating and design. "Progress in Photovoltaics: Research and Applications"; ISSN 1099-159X.

[12] Bruce H. King, Clifford W. Hansen, Daniel Riley, Charles D. Robinson and Larry Pratt Procedure to Determine Coefficients for the Sandia Array Performance Model (SAPM). SANDIA REPORT. SAND2016-5284. June 2016.

[13] M. Lee and A. Panchula, “Combined air mass and precipitable water spectral correction for PV modelling,” in Proc. 4th PV Perform. Model. Monit. Workshop, Cologne, Germany, 2015.

[14] Marios Theristis, Eduardo F. Fernández, Florencia Almonacid, and Pedro Pérez-Higueras. Spectral Corrections Based on Air Mass, Aerosol Optical Depth, and Precipitable Water for CPV Performance Modeling. IEEE JOURNAL OF PHOTOVOLTAICS, VOL. 6, Nº. 6, November 2016. Pp. 1598-1604.

[15] V. Tatsiankou, K. Hinzer, H. Schriemer, S. Kazadzis, N. Kouremeti, J. Gröbner, R. Beal. Extensive validation of solar spectral irradiance meters at the World Radiation Center, Solar Energy, Volume 166. 2018, Pages 80-89.

[16] Shimokawa R, Miyake Y, Nakanishi Y, Kuwano Y, Hamakawa Y. Possible errors due to deviation from the cosine response in the reference cell calibration under global irradiance. Jpn J Appl Phys. 1986;25(2): L102-L104.

[17] King, D. L., E. E. Boyson and J. A. Kratochvil (2004). Photovoltaic Array Performance Model. Albuquerque, NM, Sandia National Laboratories, SAND2004-3535.

[18] Martin Chivelet, Nuria & M. Ruiz, J. (2002). A new model for PV modules angular losses under field conditions. International Journal of Solar Energy. 22. 19-31. 10.1080/01425910212852.

[19] IEC 61853-2, 2016. Photovoltaic (PV) module performance testing and energy rating – Part 2: Spectral Responsivity, Incidence Angle and Module Operating Temperature Measurements.

[20] PVSyst. S.A. PVsyst 6.7.0 Sofware. www.pvsyst.com

[21] Plag F, Kröger I, Fey T, Witt F, Winter, S. Angular-dependent spectral responsivity—Traceable measurements on optical losses in PV devices. Prog. Photovolt. Res. Appl. 2017;1–14.

[22] Michalsky, J.J. et al., 1991. Spectral and temperature correction of silicon photovoltaic solar radiation detectors. Solar Energy, 47(4), pp.299–305.

[23] D. L. King and D. R. Myers, “Silicon-Photodiode Pyranometers: Operational Characteristics, Historical Experiences, and New Calibration Procedures,” Conf. Rec. Twenty Sixth IEEE Photovolt. Spec. Conf. - 1997, no. September, pp. 1285–1288, 1997.

[24] Moreno-Tejera, Sara; Silva-Pérez, Manuel Antonio; Lillo-Bravo, Isidoro; Ramírez-Santigosa, Lourdes. 2016. Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions. Solar Energy. 132: 430-441.

[25] McArthur, L. B. J. 2004 "Baseline Surface Radiation Network (BSRN): Operations Manual (Version 2.1),"

[26] M.C. Peel, B.L. Finlayson, T.A. Mcmahon. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. Discuss., 4 (2007), pp. 439-473

[27] A.H. Fakra, H. Boyer, F. Miranville, D. Bigot, A simple evaluation of global and diffuse luminous efficacy for all sky conditions in tropical and humid climate, Renewable Energy, Volume 36, Issue 1,2011,Pages 298-306

[28] Siwei Lou, Danny.H.W. Li, Wenqiang Chen. Identifying overcast, partly cloudy and clear skies by illuminance fluctuations. Renewable Energy. Volume 138. 2019. Pages 198-211.

[29] M. Larrañeta, M.J. Reno, I. Lillo-Bravo, M.A. Silva-Pérez. Identifying periods of clear sky direct normal irradiance. Renewable Energy. Volume 113. 2017. Pages 756-763.

[30] Outdoor PV performance monitoring: Pyranometers versus reference cells. 2011 https://www.hukseflux.com

[31] Nann, S. and C. Riordan, 1991: Solar Spectral Irradiance under Clear and Cloudy Skies: Measurements and a Semiempirical Model. J. Appl. Meteor., 30, 447–462,

[32] Larrañeta M., Fernandez-Peruchena C., Silva-Pérez M.A., Lillo-Bravo I. 2018. Methodology to synthetically downscale DNI time series from 1-h to 1-min temporal resolution with geographic flexibility. Solar Energy 162, 573-584.