An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series

Larrañeta, M.[1],Moreno-Tejera, S.[2], Silva-Pérez,M.A.[2], Lillo-Bravo,I.[2]

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

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

Corresponding author:

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

Phone/Fax number: (+34)954487237/ (+34)954487233

E-mail: mlarraneta@gter.es

Abstract

Several studies have confirmed the relevant impact of the resolution and frequency distribution of solar radiation data on the results of detailed production models. Many of the available direct normal irradiance (DNI) databases generated from the satellite images have an hourly resolution. In the present work, we have proposed improvements to an existing model for the generation of 10-min synthetic DNI data from the hourly average DNI values. In the original model, the irradiance is divided into a deterministic and stochastic component, i.e., the contribution from the hourly mean and stochastic fluctuation obtained from the mean depending on the sky condition, respectively. We have implemented several improvements, and the most relevant is the consistency of the synthetic data with the state of the sky. The adaptation and application of the model to the location of Seville shows significant improvements over its predecessor as it achieved 7% rRMSD in hourly values and 1% rRMSD in daily values and presented a realistic frequency distribution in the 10-min resolution. In comparison to the original model, the application of the improved model showed significant performance improvements without any further adaptations to other locations with different climatological characteristics than Seville.

Keywords

DNI; high frequency; solar radiation models; cloud transients

1   Introduction

The direct normal irradiance (DNI) time series are the basic inputs for the simulation of solar thermal electricity (STE) plants. The simulation models have different requirements in terms of time resolution of the DNI time series depending on their use or application. Studies by Meyer et al. (2009) and Gall et al. (2010) have emphasized the need to use time series with a time step shorter than 1 hour for detailed performance simulations. Many performance models used in the commercial projects for assessing contractual performance require a one-year time series with a resolution time of 5–15 min, as a reasonable compromise between the computational cost and accuracy.

On the other hand, STE plant operators frequently use irradiance predictions based on the satellite estimates or meteorological and solar radiation models to operate in electricity markets or define subsequent operational strategies. Most of the DNI prediction models are based on the short-term weather forecasts, which hardly exceed an hourly time step (Vincent, 2013) even though most of the simulation tools require higher frequency.

The need for high-frequency DNI data is reflected in the latest publications of models for generating synthetic solar irradiance data focused on high temporal resolutions (Bright et al., 2015; Fernández-Peruchena et al., 2014, 2015; Grantham et al., 2013). Many of these models are based primarily on autoregressive moving average (ARMA) or Markov transition matrix (MTM) techniques, which provide time series with increased temporal resolutions. This leads to the introduction of improvements that helps in modeling the dynamic behavior of the solar radiation (Ngoko et al., 2014; Morf, 2013; Glasbey and Allcroft, 2008). Ngoko et al. (2014) proposed a second-order MTM model that included statistical characteristics associated with the atmospheric condition (clear, cloudy, overcast) leading to the improvement in the first-order MTM model employed by Richardson and Thomas (2011) for generating 1-min values. The use of wavelets and artificial neural network (ANN) techniques for the generation of solar radiation values is mainly focused on the forecast applications. However, in most of the cases, these models deal with daily or hourly global horizontal irradiation (GHI) data (Mellit et al., 2005; Linares-Rodriguez et al., 2011).

The solar thermal energy concentrating technologies exploit only the direct component. This component has unique statistical properties (Skartveit and Olseth, 1992), showing steeper gradients than the global radiation during the cloud transients. There exists a correlation between the DNI and GHI that helps in obtaining one from the other with acceptable results; this is supported by the model study conducted by Skartveit and Olseth (1992) on the synthetic generation of irradiance values at different time intervals. Several authors have generated DNI values synthetically at a high frequency from the global irradiance values, but only a few authors have focused their models on the generation of high resolution DNI series from low-resolution DNI values.

Morf (2013) generated sequences of instantaneous global solar irradiance values on a horizontal plane that can be split into beam and diffuse components. The model was divided into a stochastic and deterministic component related to the Angström-Prescott regression. The cloud cover was used as a stochastic driver for the generation of an on/off sequence of beam irradiance (Morf, 2011); the probability of beam irradiance is represented as the complement to one of the cloud covers. Bright (2015) generated minute irradiance time series on an arbitrary plane from the previous estimation of both direct and diffuse components. This methodology used weather observation data to generate cloud transients sequences using the Markov chains. Fernández-Peruchena proposed the generation of 1-minute resolution DNI series from the daily (Fernandez-Peruchena, 2014) and hourly (Fernández-Peruchena, 2015) means of DNI values. The method was based on the previous generation of a database of dimensionless high frequency daily curves or series of DNI values obtained from the observed data. The days were selected based on the closest Euclidean distance between the daily and hourly means of the generated and measured series. The results obtained with this model were satisfactory in the terms of variability and frequency distributions; however, the authors provided no results in

relation to the deviation observed between the synthetic and measured hourly and daily means, which is one of the major targets of the present work.

Polo et al. (2011) proposed a model that was relatively indistinct to GHI and DNI. The model generated 10-min data from the hourly values while maintaining the statistical characteristics of an observed data set. Conceptually, the DNI was divided into a deterministic and stochastic component: the contribution from the hourly mean and stochastic deviation from the mean, respectively. The main problems detected with this model were the difference between the daily and hourly cumulative values of the measured and synthetic datasets that reached 2–4% in daily totals and 15% for hourly values, and the mismatches in their frequency distribution in the 10-min resolution.

In the present work, we propose some improvements to the above-mentioned model, based on the knowledge of the unique features of the DNI behavior. The implementation of the improvements have been governed by two conditions: 1) the hourly values of the original series should be conserved reasonably in the synthetic one, and 2) the dynamics of the fluctuations of the DNI must be consistent with the state of the sky. Therefore, we propose several improvements oriented toward the adjustment of the model under different sky conditions, a parameter for the distinction of hazy days based on the hourly DNI values only, and an improvement to restrict the values of the stochastic component of the model whenever required.

As a result, we present an improved model for the synthetic generation of 10-min DNI values from the hourly DNI data. The model, which solves the weaknesses of its predecessor, has been validated in five locations with different climatic conditions, showing a satisfactory performance regardless of the location in which it was used. This result was obtained despite the fact that the model has only been trained with data from one of these locations.

2   Meteorological database

The data set used for training of the model consists of the 10-min and 1-h averages of DNI measurements registered every 5 seconds during 13 years (2000–2012) at the meteorological station of the Group of Thermodynamics and Renewable Energy of the University of Seville. The model improvements have been checked with the corresponding values for the year 2013 and later validated in five different locations in Spain covering latitudes from 37° N to 43° N. The selected sites are presented in Table 1.

All the data used in this work have been subjected to quality-control procedures following the BSRN recommendations (Moreno-Tejera et al., 2015).

3   Methodology

The study followed the methodology proposed by Polo et al. (2011) for the generation of 10-min synthetic irradiance values from a given hourly time set. The DNI was divided into a deterministic and stochastic component.

The dynamics of the DNI vary considerably depending on the atmospheric conditions (clouds, aerosols, etc.). Therefore, in the first step, it is necessary to perform a clustering of the 10-min

data available as a function of the atmospheric conditions. The clustering is performed by calculating the normalized clearness index, k'_t (Perez et al., 1990), and grouping the datasets into four k'_t classes or intervals.

The second step is the calculation of the standard deviation of the 10-min DNI values with respect to the hourly mean. The results are normalized to the maximum value of the complete dataset. This helps to generate the probability density function of the hourly time sets of normalized standard deviations values for each sky condition and fit it to a beta distribution curve.

The procedure for the generation of synthetic DNI values divides the solar radiation into a deterministic and stochastic component. The first is generated by the cubic interpolation of the hourly means calculated every 4 hours in the 10-min time scale. The stochastic component is dynamically reproduced by using random numbers from the beta distribution curve whose characteristic parameters have been fitted for each sky condition, introducing a random sign for the fluctuation. The procedure is roughly described below:

i.   Calculation of the cubic interpolation of the hourly values in a 10-minutes scale to generate the shape where the fluctuations will be added ($I^i_{10m\_i3}$).

ii. Generation of random numbers from an uniform distribution curve [0,1] and determination of the inverse beta value corresponding to that probability and sky condition. This value is multiplied by the maximum standard deviation to generate the amplitude of the fluctuation (A).

iii. Generation of random numbers from a normal distribution curve with zero mean and unit standard deviation to add or subtract the amplitude to or from the mean value (r).

Finally, to estimate the 10 minute-value for the instant, i, the next operation is performed by the following equation:

$$I^i_{10m} = I^i_{10m\_i3} + sign(r) \cdot A \quad (1)$$

Where, the subscript $10m$ represents the time scale and $i3$ represents the cubic interpolated value, i represents the time instant, r is the random number from the beta distribution, and A is the amplitude of the fluctuation.

4   Improvements

In this work, we propose the following improvements to the original model by Polo et al. (2011):

-   Classification of sky condition based on the use of k_b instead of k'_t.
-   Normalization of the deviations for each proposed k_b range to the maximum value in the interval, instead of using the maximum value of the complete data set.
-   Establishment of a minimum threshold of hourly DNI for the consideration of fluctuations.
-   Use of a perturbation coefficient to model hazy days.
-   Iterative procedure to match the daily irradiation of the measured and synthetic data.

4.1 Sky condition classification

Polo et al. (2011) used k'_t for the classification of the sky condition. The clouds transient do not have the same impact on the components of the solar radiation. Disturbances in DNI were found to be steeper than those of GHI for certain types of sky conditions. k'_t is based on GHI and does not properly define the sky conditions for this application. Instead, we propose the use of k_b (Skartveit and Olseth, 1992) divided in intervals of 0.1 as described in Table 2 instead of the original division into four intervals.

$$k_b = I_{bn}/I_{bn_{cs}} \qquad (2)$$

Where, I_{bn} is the observed direct normal irradiance and $I_{bn_{cs}}$ is the clear-sky DNI.

The clear-sky DNI is calculated based on the model AB proposed by Silva-Perez (2002):

$$I_{bn_{cs}} = I_{cs} \cdot E_0 \cdot \frac{A}{1+B \cdot m_R} \qquad (3)$$

Where, m_R is the relative air mass determined according to the expression of Kasten and Young (1989), I_{cs} is the solar constant, and E_0 is the correction due to Earth-Sun distance. *A* and *B* are empirical parameters intended to model the state of transparency or turbidity of the atmosphere.

4.2 Standard deviation normalization

In the analysis of the variability of the irradiance data of the original model, the standard deviations of the 10-min sets are normalized by dividing it by the maximum deviation of the complete dataset. This procedure assumes that the amplitude of fluctuations remains similar under all sky conditions. However, this assumption may not hold true in case of DNI. Therefore, we suggest the normalization of the deviations for each proposed k_b range to the maximum value in the interval.

The largest fluctuations occur during the passage of clusters of low clouds with high density like cumulus or stratocumulus. The maximum deviations are found in the central intervals of the clearness index as shown in Table 3.

The highest differences are found for low values of the clearness index because of the lower DNI fluctuations that occur due to the presence of denser and compact clouds during the mostly-covered sky conditions.

4.3 Minimum value for fluctuations

In case of absence of clouds or overcast sky with DNI hourly values lower than 90 W/m², no fluctuations are observed in the DNI values. Although, the latter case is not relevant for the performance of a STE plant, this case is still considered important for the development of an accurate model. The calculation of the synthetic DNI is implemented by neglecting the fluctuations for hourly values below 90 W/m². Hence, the synthetic DNI is equal to the cubic interpolation of the hourly values (deterministic component).

$$I_{10m}^i = I_{10m\_i3}^i \qquad (4)$$

Where, I_{10m}^i is the generated synthetic irradiance and $I_{10m\_i3}^i$ is the cubic interpolation of the hourly values.

Figures 1, 2, and 3 compare the results of the improvements in the daily DNI.

The fluctuations of the improved model are more consistent with the performance of the 10-min DNI measured data than the original model.

4.4 Perturbation coefficient

DNI modeling in hazy days is a task of high relevance (Gueymard, 2005) and has been only partially solved. While hourly and daily k_b values suggest partly cloudy days ($k_b \approx 0.55$), the DNI may show negligible fluctuations from the cubic interpolation. At this point, we have defined a new coefficient "P2" for determining the fluctuations that occur due to DNI.

The coefficient is based on the concept of tortuosity, commonly used in the diffusion of porous media (Epstein, 1989) that can be defined in a simplified manner as the relationship between the length of a curve, L, and a straight segment (chord) that joins its ends, X.

$$\tau = {}^L/_X \qquad (5)$$

Figure 4 represents the diffusion in a porous medium where the tortuosity is larger in case B than in case A because the chord is smaller even for the same length.

Many mathematical equations have been proposed over the years for the accurate estimation of the value of tortuosity for a certain curve f(t). Patasius et al. (2005) proposed estimation as the integral of the squared derivative of the curve divided by the curve length, L.

$$\tau = \frac{\int_{t1}^{t2}(f'(t))^2 dt}{L} \qquad (6)$$

With regard to solar radiation, Muselli et al. (2000) proposed a coefficient to estimate the perturbation state of the hourly clearness index curve during the day from the integral of the second derivative.

$$S2 = \sum_h \{k_{t_{h+2}} - (2 \cdot k_{t_{h+1}}) + k_{t_h}\}^2 \quad (7)$$

In this study, we propose a coefficient to estimate the perturbation state of the hourly DNI to estimate days without fluctuations in the 10-min scale similar to the S2 coefficient by using the direct normal irradiance profile instead of the clearness index profile.

$$P2 = \sum_h \{Ibn_{h+2} - (2 \cdot Ibn_{h+1}) + Ibn_h\}^2 \qquad (8)$$

In an experimental approach, we have observed that fluctuations with high amplitude and frequency occurred under variable sky conditions. For daily k_b index lower than 0.3, regular fluctuations were observed on the DNI. Thus, we have calculated the perturbation coefficient for daily values of $k_b > 0.3$ only. Observing the daily graphs of the DNI, we have identified a value of P2=$215 \cdot 10^3 W^2/m^4$ that defines a boundary between days, with and without the fluctuations, with an 82% of success.

Figure 5 illustrates the effect of this improvement by comparing the results of the original model, modified by using k_b instead of k'_t, for the data classification (original model + improvement 1, left) with the improved model (right) for a hazy day.

4.5 Similarity in the daily sums

The application of the original model in the generation of synthetic series often results in significant differences between the cumulative daily values of the original hourly and synthetic series. This problem can be solved by means of an iterative procedure where the daily synthetic series are recalculated until both cumulative daily values differ in less than 2%. A daily uncertainly of a 2% is accepted since that represents the uncertainty of most of the first class pyrheliometers.

Figure 6 shows the block diagram of the implemented model.

5   Results

The common practices were followed for benchmarking of the modeled irradiance datasets (Beyer et.al, 2008). We used the root mean squared difference (RMSD) as the main statistical method for comparison of these observations and generate data synthetically. In the analysis, only daylight hours were considered.

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I^i_{meas} - I^i_{synth}\right)^2} \quad (9)$$

Where, N is the number of data pairs, I_{synth} is the synthetic DNI and I_{meas} is the measured DNI (year 2013 for the location of Seville).

The corresponding relative differences are calculated as follows:

$$rRMSD = RMSD \Big/ \overline{I_{meas}} \qquad\qquad (10)$$

The normalized root mean squared deviation (NRMSD) is calculated as follows:

$$NRMSD = RMSD \Big/ (I_{max} - I_{\min}) \qquad\qquad (11)$$

Where, I_{max} and I_{min} are the maximum and minimum values of the observed dataset, respectively.

The analysis has been made in two time scales, daily and hourly. In order to quantify the effect of each implemented improvement, we have generated the following intermediate models:

• Original: This is the original model that uses the k'_t index for the classification of the sky conditions.

• Original_k_b: Includes the sky classification using the k_b index.

• M1: Expands the range of the k_b index to 0.1 for each sky condition.

• M2: Implements the normalization of the standard deviation (par. 4.2).

• M3: Includes a DNI threshold below which fluctuations are neglected (par. 4.3).

• M4: Includes the parameter P2 for identification of hazy days (par. 4.4).

• M5: Implements an iterative procedure to restrict the differences in daily cumulative values (par. 4.5).

• I3: Does not include any improvement, but only the cubic interpolation of the hourly values at a 10-min scale.

Each MX model includes the improvements of the previous ones. Thus, model M5 includes all the improvements. We have also included model I3, which implements only the deterministic component with no stochastic contribution.

The results of the benchmarking in the daily scale are presented in Table 4 and Figure 7; Table 5 and Figure 8 show the results in the hourly scale.

It is remarkable that the use of the clearness index, k_b led to a reduction of the daily and hourly errors of more than 50% compared to the original model. The improvements M1, M2, and M3 do not significantly reduce these errors because they mainly affect the low irradiance values that have less weight in the global computation. The effect of improvement 4 are only noticeable in hazy days, which is an infrequent condition and cannot be quantified on these scales, while improvement 5 reduces the error by less than 0.8% and 7% in a daily and hourly scale, respectively.

6   Discussion and validation

The cubic interpolation of the hourly values at a 10-min scale (I3) and the model that includes all the improvements (M5) show a similar performance in terms of RMSD, but the first one is much simpler. To compare if the respective synthetic datasets are statistically representative, we analyzed the probability density function (PDF) and cumulative distribution function (CDF) at the 10-min resolution, which are considered to have a significant impact on the performance of the STE plants. The use of data with unrealistic frequency distributions as input for STE simulation software leads to unrealistic energy yields (Silva - Pérez et al., 2014) reaching differences of up to 9% for sites with a similar annual DNI (Chhatbar and Meyer, 2011).

Therefore, it is interesting to compare the measured DNI against the DNI sets generated with the original model, with M5 and with I3, to identify the synthetic dataset that better resembles the statistical characteristics of the measured DNI. Figure 9 shows the PDF´s and CDF's of the addressed datasets depending on the sky condition.

In comparison to the frequency distribution of the original (measured) dataset, the cubic interpolation of the hourly values at a 10-min scaleset (I3) exhibits a completely different frequency distribution, especially for partly cloudy sky conditions where the greatest fluctuations take place. The CDF of the improved model (M5) is much closer and improves the performance of the original model.

For quantification of this statement, we have calculated the Finkelstein-Schafer (FS) statistic (Finkelstein and Schafer, 1971) for each dataset. This statistic takes into account the differences between the CDF of the measured and synthetic datasets, and permits the comparison with the results of other models regardless of the time resolution and analysis period.

$$FS = \frac{1}{n} \sum_{i=1}^{n} \delta_i \quad (12)$$

Where, δ is the absolute difference between the measured and synthetic CDF at each point, i, and n represents the number of readings.

Table 6 presents the FS of the original model, improved model, and cubic interpolation of the hourly values at a 10-min scale. The improved model significantly reduces the FS in comparison with the original and I3 model. The synthetic dataset generated with the improved model is similar to the measured dataset regardless of the sky condition, as shown in Figure 10.

The robustness of the model, trained with data from Seville only, can be observed when applying it to other sites with different climatic conditions and no previous adaptation (the parameters of the beta distribution used to generate the stochastic component are those obtained for Seville). We tested it on five sites in Spain at different latitudes in the Mediterranean, Atlantic, and Continental climates. The results are summarized in Table 7.

The improved model provides the best results when applied to Seville and slightly reduces its performance for the other five selected sites, but still the results are significantly better than those of the original model. The declining performance of the model in the locations selected for the validation is mainly caused because the stochastic component of the model has been trained with data from Seville.

7   Conclusions

In this paper, we propose improvements to the model of Polo et al. (2011) for the generation of synthetic DNI data in a 10-min resolution taking hourly mean values as input. The main weaknesses of the original model were found while comparing the differences between the hourly and daily sums of the generated and measured dataset due to the stochastic nature of the methodology. The application of the proposed improvements, based on the knowledge of the distinctive features of the DNI, kept the daily cumulative values and consistency of the fluctuations of the synthetic DNI with the observed sky condition to reduce both the errors in more than 40% rRMSD bringing them to 1.3% and 7.9% in daily and hourly sums, respectively. The use of k_b instead of k'_t resulted in a significant reduction of the RMSD while the outcome of the rest of the improvements showed a better performance with respect to the frequency distribution of the synthetic data. In addition, the improved model has been applied without any adaptation to other locations with diverse climatic conditions, achieving excellent results compared to the original model, and demonstrating the robustness of the proposed methodology.

Acknowledgments

References

Beyer, H. G., Pozo-Martinez, J., Suri, M., Torres, J. L., Lorenz, E., Hoyer-Klick, C., Ineichen, P., 2008. D 1.1.1 Handbook on Benchmarking, Management and Exploitation of Solar Resource Knowledge, CA – Contract No. 038665.

Bright, J.M., Smith, C.J., Taylor, P.G., Crook, R., 2015. Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data. Solar Energy 115 229–242

Chhatbar, K., Meyer, R., 2011. The Influence of Meteorological Parameters on the Energy Yield of Solar Thermal Power Plants. SolarPACES Proceedings, Granada, Spain.

Epstein, N., 1989. On tortuosity and the tortuosity factor in flow and diffusion through porous media. Chem. Eng. Sci., 44(3), 777– 779.

Fernández-Peruchena, C.M., Blanco, M., Gastón, M., Bernardos, A., 2015. Increasing the temporal resolution of direct normal solar irradiance series in different climatic zones. Solar Energy 115, 255-263.

Fernández-Peruchena, C.M, Blanco, M., Bernardos, A., 2014. Generation of series of high frequency DNI years consistent with annual and monthly long-term averages using measured DNI data. Energy Procedia 49, 2321-2329.

Finkelstein, J.M., Schafer, R.E., 1971. Improved Goodness-of-Fit Tests. Biometrika, 58, 641-645.

Gall, J., Abel, D., Ahlbrink, N., Pitz-Paal, R., Anderson, J., Diehl, M., Teixeira Boura, M., Schmitz, M., Hoffschmidt, B., 2010. Simulation and control of thermal power plants. In: Proceedings of the International Conference on Renewable Energies and Power Quality (ICREPQ'10), Granada (Spain), March 23–25, 2010, pp. 294–298.

Glasbey, C. A., Allcroft D. J., 2008. A Spatiotemporal Autoregressive Moving Average Model for Solar Radiation. Applied Statistics, 57, 343-355.

Grantham, A.P., Pudney, P.J., Boland, J.W., Belusko, M., 2013. Synthetically interpolated five-minute direct normal irradiance. In: 20th International Congress on Modelling and Simulation. Adelaide, Australia.Gueymard, C.A., 2005. Importance of atmospheric turbidity and associated uncertainties in solar radiation and luminous efficacy, modeling. Energy 30,1603–1621.

Gueymard, C. A., 2005. Importance of atmospheric turbidity and associated uncertainties in solar radiation and luminous efficacy, modeling. Energy 30, 1603–1621

Kasten, F., Young A.T., 1989. Revised optical air mass tables and approximation formula. Applied Optics 28 (22) 4735-4738.

Linares Rodirguez, A., Ruiz-Arias, J.A., Pozo-Vázquez, D., Tovar-Pescador, J., 2011. Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks. Solar Energy 36, 5356-5365.

Mellit, A., Benghanem, M., Hadj Arab, A., Guessoum, A., 2005. A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach. Solar Energy 79, 469-482.

Meyer, R., Beyer, H.G., Fanslau, J., Geuder, N., Hammer, A., Hirsch, T., Hoyer-Click, C., Schmidt, N., Schwandt, M., 2009. Towards standardization of CSP yield assessments. Proceedings of the SolarPACES conference, Berlin (Germany).

Moreno-Tejera, S. Ramírez-Santigosa, L., Silva-Pérez, M.A., 2015. A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data. Solar Energy 76, 531-537.

Morf, H., 2013. A stochastic solar irradiance model adjusted on the Ångström–Prescott regression. Solar Energy 87, 1-21.

Morf, H., 2011. The stochastic two-state cloud cover model STSCCM. Sol. Energy 85 (5), 985–999.

Muselli, M., Poggi, P., Notton, G., 2000. Classification of typical meteorological days from global irradiation records and comparison between two Mediterranean coastal sites in Corsica Island. Energy Conversion & Management 41, 1043-1063.

Ngoko, B.O., Sugihara, H., Funaki, T., 2014. Synthetic generation of high temporal resolution solar radiation data using Markov models. Solar Energy 103, 160-170.

Patasius, M., Marozas, V., Lukosevicius, A., Jegelevicius, D., 2005. Evaluation of tortuosity of eye blood vessels using the integral of square of derivative of curvature. Proceedings of the 3rd IFMBE European Medical and Biological Engineering Conference, Prague. 11, 1-4.

Perez, R., Ineichen, P., Seals, R., Zelenka, A., 1990. Making full use of the clearness index for parameterizing hourly insolation conditions. Solar Energy 45, 111–114.

Polo, J., Zarzalejo, L.F., Marchante, R., Navarro, A.A., 2011. A simple approach to the synthetic generation of solar irradadiance time series with high temporal resolution. Solar Energy 85, 1164-1170.

Richardson, I., Thomson, M., 2011. Integrated simulation of photovoltaic micro-generation and domestic electricity demand: a one-min resolution open source model. In: Microgen II: 2nd International Conference on Microgeneration and Related Technologies, Glasgow (Scotland).

Silva-Pérez, M.A., 2002. Estimación del recurso solar para sistemas termosolares de concentración. University of Seville.

Silva-Pérez, M.A., Barea-García, J.M., Larrañeta, M., Moreno-Tejera, S., Lillo, I., 2014. Analysis of the distribution of measured and synthetic DNI databases and its effect on the expected production of a parabolic trough plant. Energy Procedia 49, 2512-2520.

Skartveit, A., Olseth, J.A., 1992. The probability density and autocorrelation of short-term global and beam irradiance. Solar Energy 49, 477-487.

Vincent, E.L., 2013. Chapter 12-Forecasting Solar Irradiance with Numerical Weather Prediction Models. Solar Energy Forecasting and Resource Assessment. Pages 229-318.
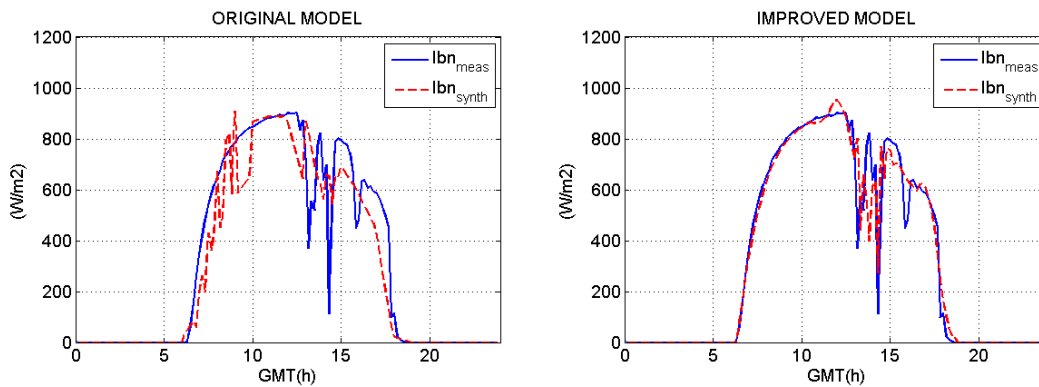
FIGURES

Fig 1. Comparison of the daily measured DNI with the synthetic data generated with the original and improved model (effect of the improvement 4.1).
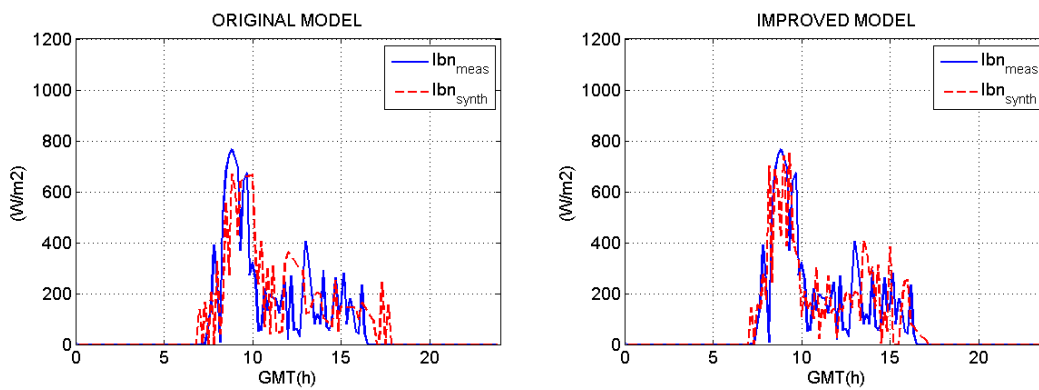

Fig 2. Comparison of the daily measured DNI with the synthetic data generated with the original and improved model (effect of the improvement 4.2).
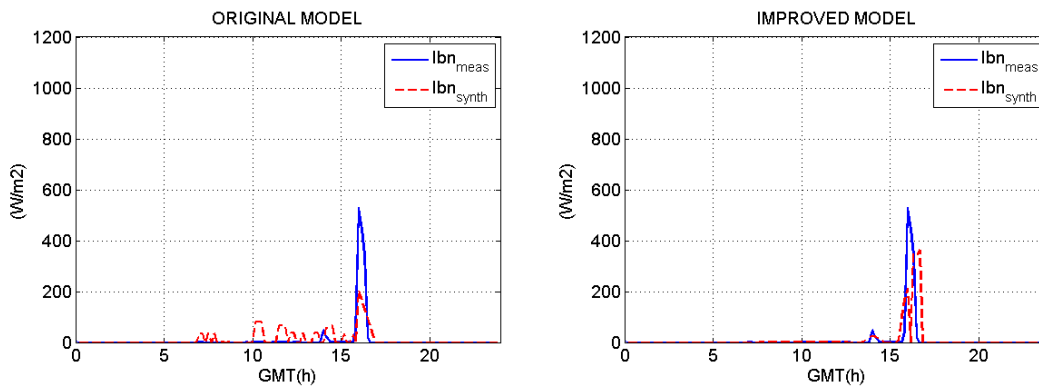

Fig 3. Comparison of the daily measured DNI with the synthetic data generated with the original and improved model (effect of the improvement 4.3).
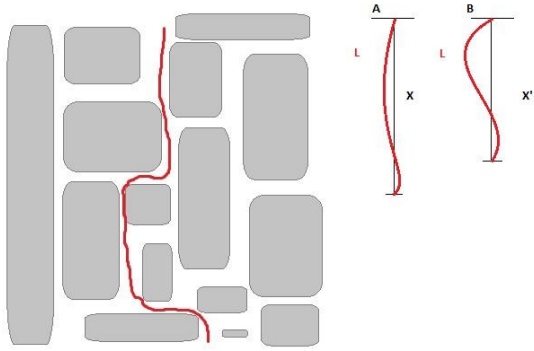
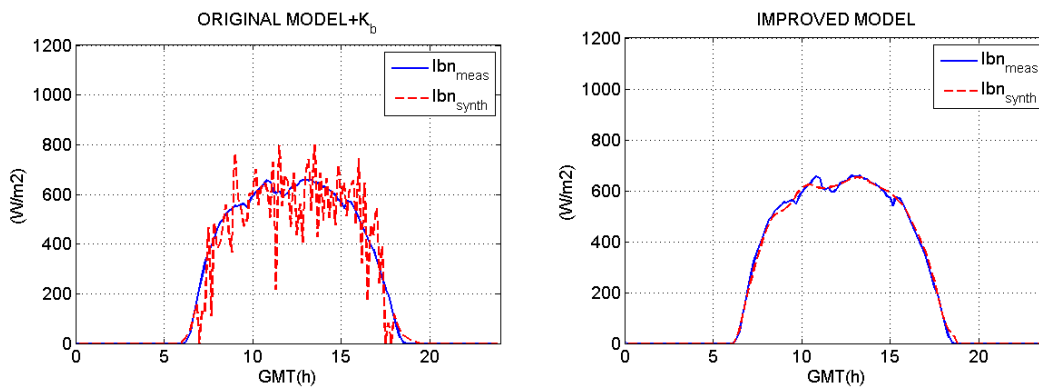Fig 4. Tortuosity in porous media.

Fig 5. Comparison of the measured DNI on a hazy day with the synthetic data generated with the original and improved model (effect of the improvement 4.3).
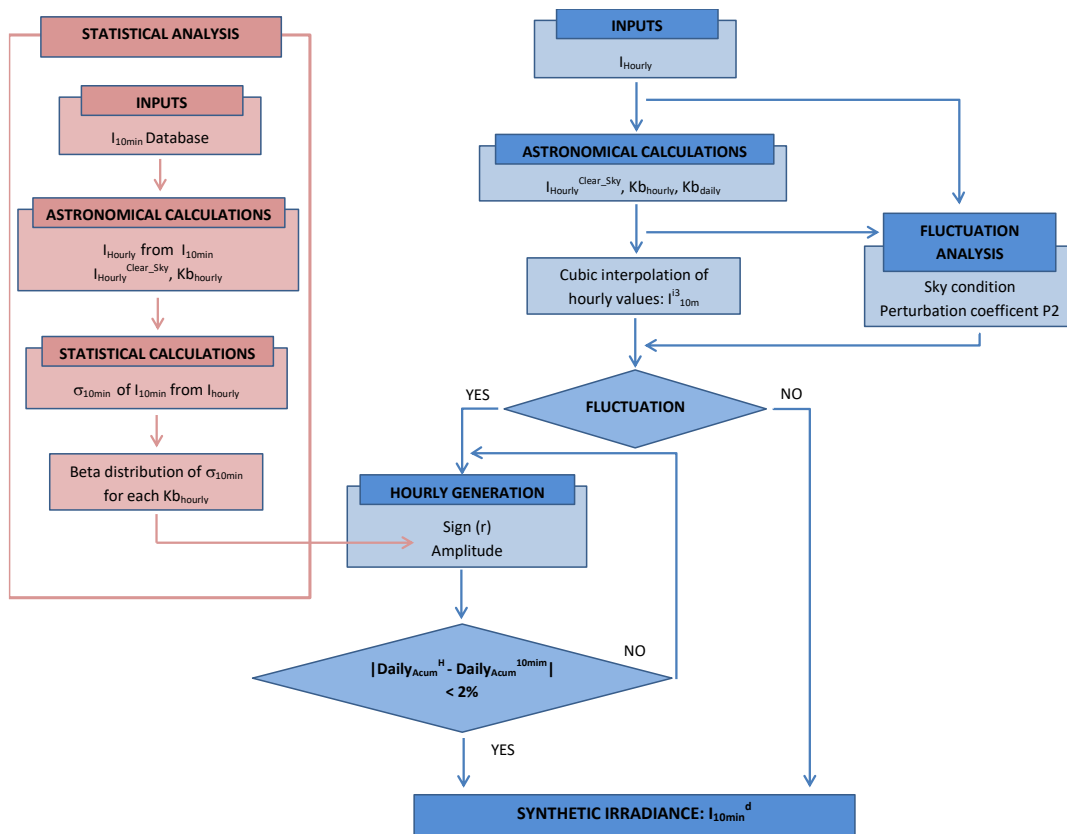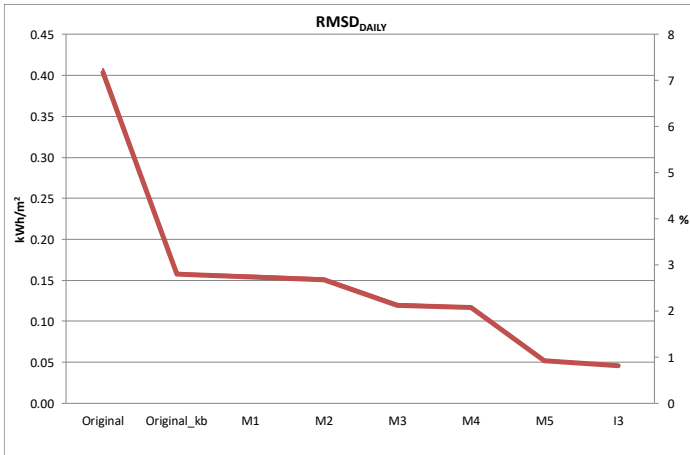

Fig 6. Block diagram of the model.

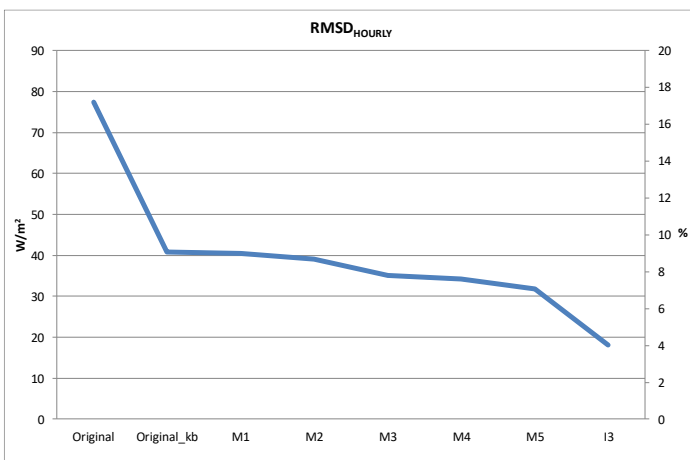Fig 7. Variation of the daily RMSD with the application of the proposed improvements.

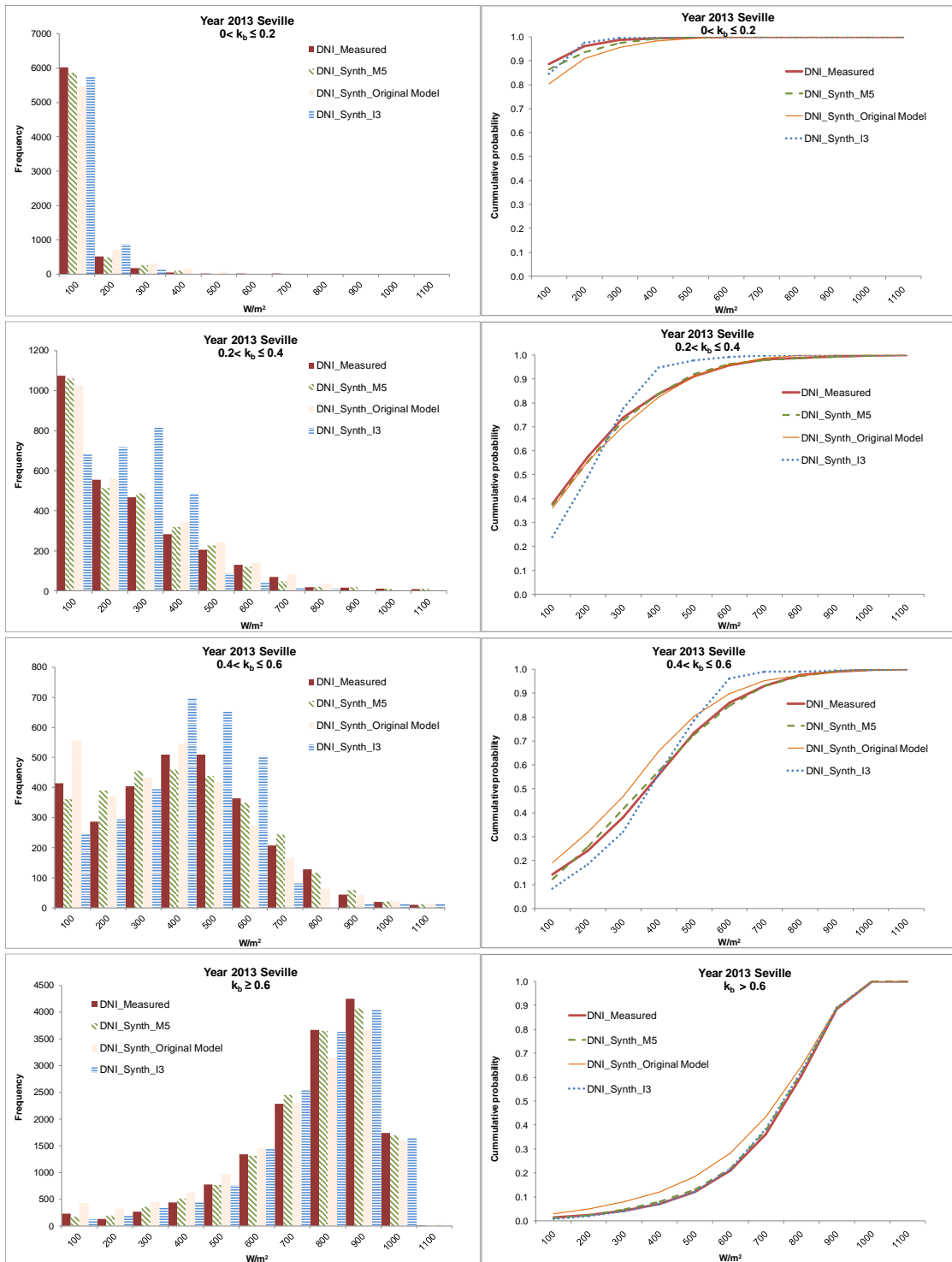Fig 8. Variation of the hourly RMSD with the application of the proposed improvements.

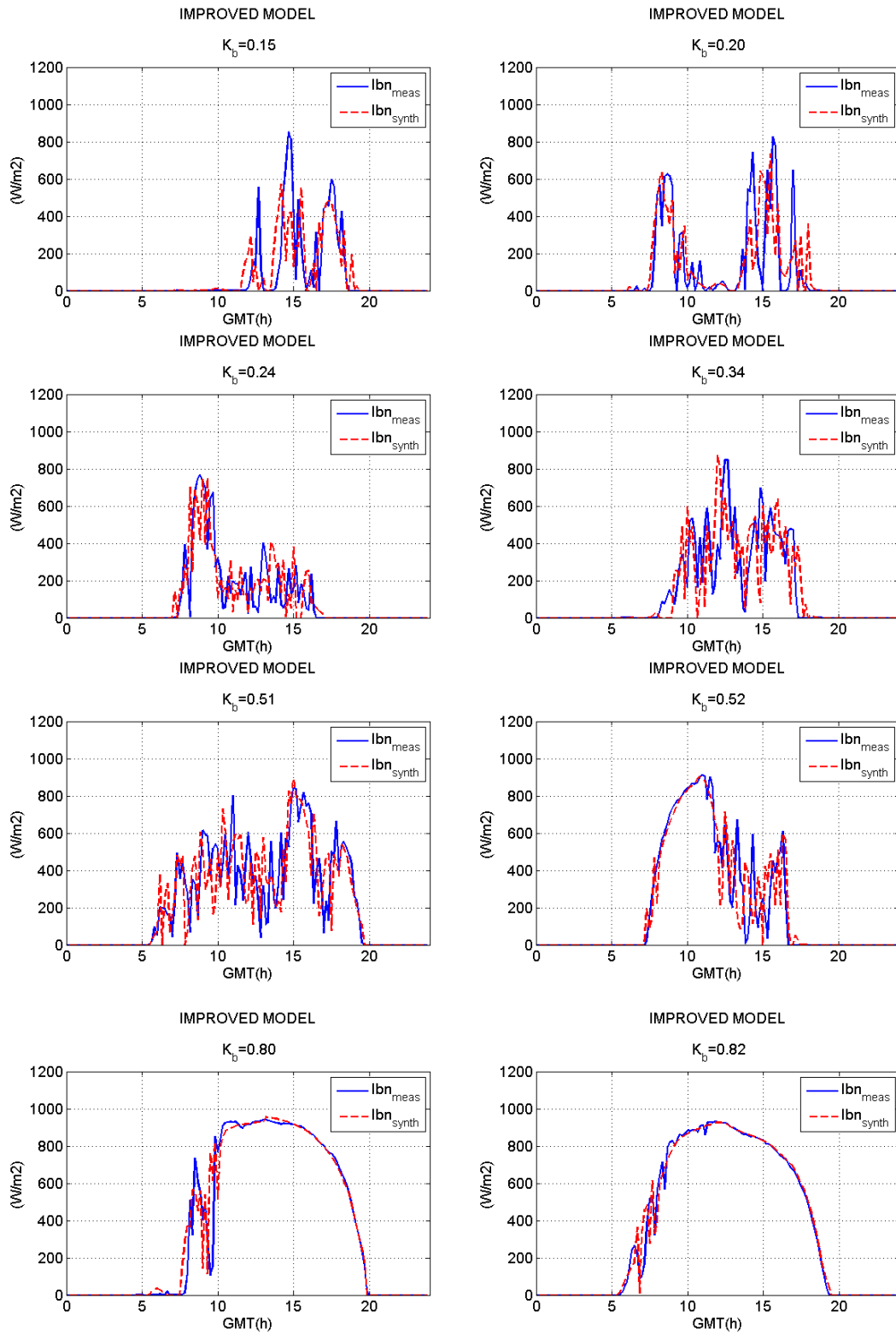Fig 9. PDF (left) and CDF (right) analysis of the synthetic and measured datasets.

Fig 10. Daily examples

TABLES

Table 1. Selected locations for the model validation.

	Latitude (°N)	Longitude (°W)	Altitude (m)	Climate	Year
Granada	37.1	−3.0	1100	Continental Mediterranean	2013
Badajoz	38.9	−6.4	200	Continental Mediterranean	2011
Alicante	38.6	−0.8	500	Continental	2009
Pamplona	42.8	−1.6	450	Atlantic	2010
Almeria	37.1	−2.3	500	Mediterranean	2012
Seville (trained)	37.4	−6.0	10	Mediterranean	2013

Table 2. Sky condition for each k_b interval.

	Sky condition	**Type of day**
$K_b \leq 0.1$	Totally covered	Thick clouds, no fluctuations
$0.1 < k_b \leq 0.2$	Totally covered	Thick clouds, some fluctuations
$0.2 < k_b \leq 0.3$	Mostly covered	Alternative clouds and clears, some fluctuations
$0.3 < k_b \leq 0.4$	Mostly covered	Alternative clouds and clears, great fluctuations
$0.4 < k_b \leq 0.5$	Partly covered	Thin clouds, greater fluctuations
$0.5 < k_b \leq 0.6$	Partly covered	Thin clouds, some fluctuations (and hazy days)
$0.6 < k_b \leq 0.67$	Mostly clear	No clouds but certain turbidity
$k_b > 0.67$	Totally clear	No clouds

Table 3. Maximum deviation of the 10-min direct normal irradiance with respect to the hourly mean for each k_b interval.

Maximum standard deviation analysis							
k_b	[0, 0.1]	[0.1, 0.2]	[0.2, 0.3]	[0.3, 0.4]	[0.4, 0.5]	[0.5, 0.6]	[0.6, 0.67]
σ_{max}^{improv} **(W/m²)**	215	345	401	431	442	435	430
σ_{max}^{orig} **(W/m²)**	442	442	442	442	442	442	442
Diff (W/m²)	227	97	41	11	0	7	12

Table 4. Effect of the improvements in the daily time scale.

Daily time scale								
	Original	Original_kb	M1	M2	M3	M4	M5	I3
RMSD (kWh/m²)	0.42	0.16	0.16	0.12	0.12	0.12	0.05	0.05
rRMSD (%)	7.4	2.9	2.8	2.8	2.2	2.1	0.9	0.8

Table 5. Effect of the improvements in the hourly time scale.

	Hourly time scale							
	Original	Original_kb	M1	M2	M3	M4	M5	I3
RMSD (W/m^2)	77.8	41.0	40.7	39.1	34.8	34.0	31.9	18.1
rRMSD (%)	17.2	9.1	9.0	8.7	7.8	7.6	7.1	4.0

Table 6. Finkelstein-Schafer (FS) analysis

	FS
Original model	0.0233
I3	0.0160
M5 model	0.0061

Table 7. Comparison of rRMSD, NRMSD, and FS when using the M5 in different locations.

	rRMSE$_{daily}$ (%)	NRMSE$_{daily}$ (%)	rRMSE$_{hourly}$ (%)	NRMSE$_{hourly}$ (%)	NRMSE$_{10-min}$ (%)	FS
Granada	1.5	0.8	10.4	5.5	12.3	0.0090
Badajoz	1.3	0.6	6.3	3.1	11.8	0.0092
Alicante	1.5	0.7	7.2	3.4	12.5	0.0101
Pamplona	1.9	0.7	9.6	3.6	13.0	0.0099
Almeria	0.7	0.4	8.4	4.2	10.5	0.0100
Seville	0.8	0.5	5.7	3.2	9.8	0.0061