Methodology to synthetically downscale DNI time series from 1-hour to 1-min temporal resolution with geographic flexibility

Larrañeta, M.[1], Fernandez-Peruchena, C.[2], Silva-Pérez, M.A.[3], Lillo-Bravo, I.[3]

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

[2] National Renewable Energy Centre (CENER).

[3] 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

In this paper, we present two methods for the synthetic generation of 1-min Direct Normal solar Irradiance (DNI) data from hourly means that can be applied globally without any local adaptation, which are based in the modelling of the stochastic component of DNI, and in the adimensionalization of the daily profiles. The similitude between measured and generated DNI distributions has been evaluated through the Kolmogorov-Smirnov test Integral (KSI), and its performance on the thermal power produced by a parabolic trough (PT) plant has been calculated using the daily normalized root mean square deviations (NRMSD) with respect to site measurements. The generation methods provides, for an annual 1-min synthetic data set, KSI values of ~3.3 W/m^2 and ~12.9 W/m^2 (depending on the generation method used), and daily NRMSD of ~0.9% and 3.4%, respectively. Sites selected for validating these methods are located at different climates and latitudes, suggesting their global applicability.

Keywords

DNI; high frequency; solar radiation models; cloud transients

1 Introduction

Developers and operators of concentrated solar thermal (CST) plants require Direct Normal Irradiance (DNI) data with high resolution for the detailed performance simulations (Gall et al., 2010). However, high resolution DNI data are often limited in duration and location, and typically historical solar resource data are available at hourly scale (Fernández-Peruchena et al., 2010). DNI series can be calculated at 15-min time intervals from currently operating satellites, but even this resolution may not be sufficient when evaluating a CST system performance (Beyer et al., 2010). Moreover, satellite-derived long time historic DNI series often does not maintain the frequency distribution of the ground measured data (Hammer et al., 2009).

There have been attempts to generate high resolution solar irradiance data. An early example is given by Skartveit and Olseth, (1992) where the probability distribution of short-term irradiance data, normalized by transformation to clear sky index data together with the knowledge of the autocorrelation coefficient of these sets forms the bases for a scheme of data synthetization. Beyer et al. (2010) generated high frequency DNI series from their cumulative distribution functions. Morf (2013) generated sequences of instantaneous Global Horizontal solar Irradiance (GHI) values dividing the solar radiation into a deterministic and a stochastic component. The deterministic component was related to the Ångström–Prescott regression, while the stochastic component was derived from the cloud cover index. Polo et al. (2011) developed a model to generate synthetic 10-min DNI and GHI data by adding the deterministic contribution of the hourly mean values to the stochastic fluctuation from the mean. This model was improved by Larrañeta et al. (2015) for a more accurate DNI generation, and modified by Grantham et al. (2017) for generating matched pairs of 5-min GHI and DNI values from hourly means. Grantham et al. (2013) previously proposed the use of bootstrapping techniques for generating synthetic 5-min DNI series from hourly means. Ngoko et al. (2014) presented a second-order Markov Transition Matrix (MTM) to generate 1-min synthetic GHI from the daily clearness index. Bright et al. (2015) also used MTM to stochastically determine cloud cover to subsequently generate 1-min DNI, GHI and diffuse irradiance. However, the model requires other meteorological information such as cloud base height, wind speed or sea level pressure. The model was improved including the spatial dimension variation in the synthetic generation without the need of input irradiance data (Bright et al., 2017). Fernández-Peruchena proposed the generation of 1-minute resolution DNI series from the daily (Fernandez-Peruchena et al., 2014), 3-h (Fernández-Peruchena et al., 2017) and hourly (Fernández-Peruchena et al., 2015) DNI means. The method is based on the generation of a dimensionless high frequency database of daily DNI curves. The same concept was used by Fernández-Peruchena et al. (2016) to generate synthetic 1-min GHI from hourly means.

It is worth highlighting that these earlier procedures require high-frequency ground measurements for characterizing in the location under study for generating high-frequency solar irradiance series. The models presented in this work use measurements from one location to characterize the cloud transients and generate synthetic 1-min DNI data in any location where hourly DNI data is available, without any local adaptation. Also, they require different degrees of accuracy in the knowledge of local hourly DNI data, thereby facilitating their application:

- The SA (Stochastic Adaptation) method requires high-quality site hourly DNI series and consists on dividing the solar radiation into a deterministic and stochastic component (i.e., the contribution from the hourly mean and stochastic the fluctuation from the mean depending on the sky condition). This method is based on Polo et al. (2011) and Larrañeta et al. (2015).

- The ND (Non-Dimensional) method only requires the site intra-daily characterization of DNI variability and distribution, and thus does not require exact hour-to-hour local DNI series. This method is based on the adimensionalization of high frequency daily DNI profiles by a clear-sky envelope approach, and is based on Fernández-Peruchena et al. (2015).

The models have been applied in three locations with different climatic conditions. The paper is presented as follows: Section 2 presents measured database used in the work; Section 3 describes the methodologies proposed for generating 1-min DNI data from hourly means. Section 4 shows the results found and in Section 5 discussion, conclusions and future work are drawn.

2 Meteorological database

In this work, an extensive database is used for training the methods proposed (Table 1). This database is composed of 1-min averages values of DNI recorded during 14 consecutive years (2002–2015) for the location of Seville (Spain). The measurements were taken with a sampling and storage frequency of 0.2 Hz. A first class Eppley NIP pyrheliometer mounted on a sun tracker Kipp & Zonen 2AP measured the DNI. The devices are located at the meteorological station of the Group of Thermodynamics and Renewable Energy of the University of Seville and have been periodically calibrated, at least once every two years.

Table 1. Location selected for the training the methods.

	Latitude (°N)	Longitude (°W)	Altitude (m)	Climate	Period
Seville	37.4	6.0	12	Mediterranean	2002-2015

In addition, the models have been validated in three other locations belonging to different climates and latitudes (Table 2). We have selected these locations as a compromise solution between climate representativeness and availability of high quality 1-min DNI measured data. DNI data of Payerne (Vuilleumier et al., 2014) measured with a first class pyrheliometer Kipp & Zonen CHP1 pyrheliometer, and have been provided by the Baseline Surface Radiation Network (BSRN) (Ohmura et al., 1998); DNI data from Pretoria has been accessed from the Southern African Universities Radiometric Network (SAURAN) (Brooks et al., 2015), and have been measured with a Kipp & Zonen CHP1 pyrheliometer; DNI data from Almeria belongs to CIEMAT and DLR meteorological station at the Plataforma Solar de Almeria (PSA), and have been measured with a Kipp & Zonen CHP1 pyrheliometer. Data used in this work have been subjected to quality-control procedures following the BSRN recommendations (McArthur, 2004).

Table 2. Locations selected for validating the methods.

	Latitude	Longitude	Altitude (m)	Climate	Period	Radiometric Network
Almería	37.1 °N	2.3 °E	500	Semi-arid	2013	CIEMAT-DLR
Pretoria	-25.7 °N	28.2 °W	1410	Sub-Tropical	2016	SAURAN
Payerne	46.8 °N	6.9 °W	491	Continental	2014	BSRN

data in the 1-min

3 Methodology

In the next sections, we briefly describe the two methodologies implemented for the synthetic generation of high frequency synthetic DNI data from hourly means.

3.1 Stochastic adaptation (SA) model

The most recent scheme of this model is based on a methodology proposed by Larraneta et al. (2015) which in turn improved the model proposed by Polo et al. (2011). In this model, the DNI is divided into a deterministic and stochastic component:

- The deterministic component is generated by the cubic interpolation of the hourly means calculated every 4 hours in the high-resolution time scale (Ibn_{i3}^{i}).
- The stochastic component (Ibn_{stoc}^{i}) is dynamically reproduced by using random numbers from the beta distribution curve whose characteristic parameters were fitted for each sky condition, introducing a random sign for the fluctuation.

The synthetic data for the instant i (Ibn_{synth}^{i}) is calculated as the combination of the deterministic plus the stochastic component (Eq. 1):

$$Ibn_{synth}^{i} = Ibn_{i3}^{i} + Ibn_{stoc}^{i} \quad (1)$$

We aim to generate synthetic data at 1-min resolution for facilitating detailed modelling of Concentrating Solar Power (CSP) performance (Ramirez et al., 2017) , even if the original model was designed to generate synthetic 10-min data. This decision implies a new DNI fluctuation characterization and the reassessment of the dynamics of the stochastic component reproduction.

3.1.1 Fluctuations in clear sky equivalent DNI hours.

In the previous model (Larraneta et al., 2015), the decision of whether or not to include fluctuations in the synthetic data was led by a daily index. In many cases, we added fluctuations in clear sky periods because a daily index may not be appropriate to characterize the intra-daily performance of the solar radiation. To solve this weakness, we implemented an algorithm to identify clear sky equivalent DNI periods from hourly DNI means (Larrañeta et al., 2017 (a)). The approach is performed comparing the differences between the means, the slopes and the lengths of the measured and theoretical clear sky curves. In periods identified as clear, we would not add fluctuations (Ibn_{stoc}^{i} = 0 W/m^2). Fig. 1 shows a daily shape of the measured DNI ($Ibn_{measured}$, dotted blue line) and the periods identified as clear sky equivalent DNI ($Ibn_{clear-eq}$, black line) with the mentioned algorithm.

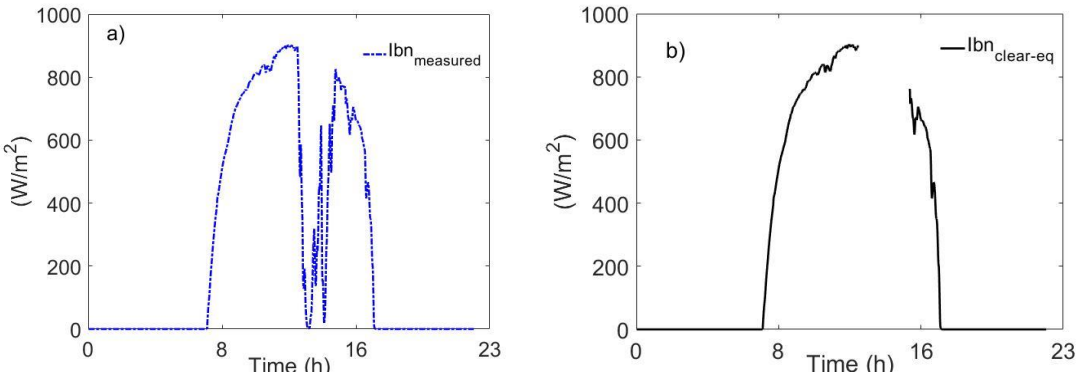


Figure 1. Daily example of the clear sky equivalent DNI identification algorithm output. Measured DNI (a), and clear sky equivalent DNI periods identified by the algorithm (b).

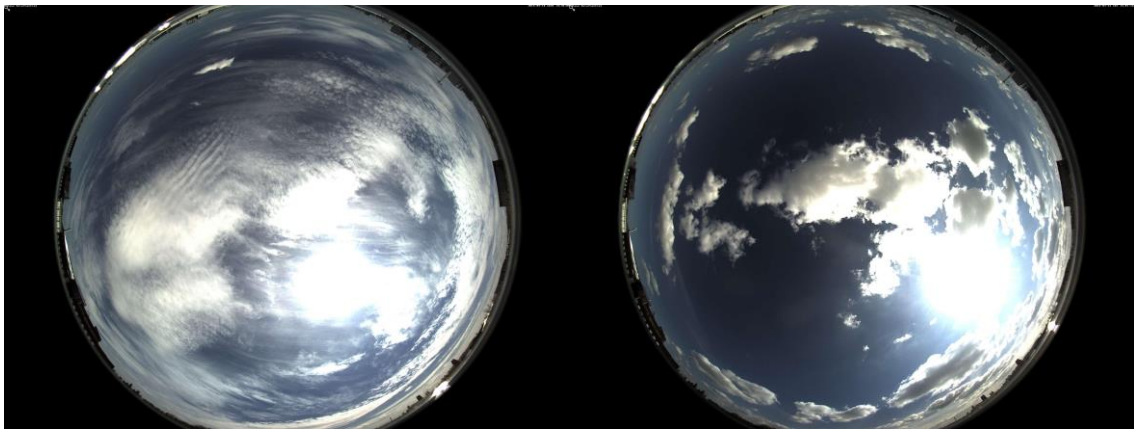
3.1.2 Characterization of the Fluctuations

The fluctuations of the instantaneous solar irradiance from their hourly means were randomly reproduced in the previous model from beta distributions fitted to different sky conditions. Each sky condition was defined as an interval of 0.1 points of the direct fraction index, k_b (Skartveit and Olseth, 1992).

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

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

Each group of data should represent the behaviour of the instantaneous DNI under different sky conditions. However, we found that different sky condition situations were lumped together in the same group. A group defined solely by means of an interval 0.1 points of the direct fraction index and even more in an hourly basis would include situations with different types of passing clouds; therefore, it may not be enough to characterize the fluctuations of the instantaneous solar irradiance. In Figure 2, we show two images of the sky with similar direct fraction index while the passing clouds are different. The left picture shows cirrostratus and the right picture shows broken clouds.



(a) (b)

Figure 2. Sky camera images taken with similar hourly direct fraction index. (a) k_b= 0.44 obtained with a cirrostratus. (b) k_b= 0.40 obtained with a set of broken clouds.

For a more accurate generation procedure, we apply a Machine Learning based classification method to perform a greater discretization of the 14-years database used for training the model. The aim of a clustering algorithm based on a partitioning method is to classify a set of data with the same features into groups or clusters. In this work, we use the *k-Medoid* algorithm (Han and Kamber, 2001) since it is the most appropriate algorithm for this type of data (Al-Shammari et al., 2016). The *k-medoid* algorithm estimates a reference point or "medoid" which is the most centrally located object or point in the cluster and then minimizes the sum of the metric distance between every object of a group and its medoid. The *k-medoid* requires input information to classify each hour *h*. Based on Perez et al. (2011), we calculate two statistics intending to define the sky condition by characterizing the DNI instant fluctuations from the mean.

1. The standard deviation of the differences between the 1-min DNI data to the cubic interpolation of the hourly values in the 1-min scale (σ^{h}_{1-min}).
2. The maximum value of the differences between the 1-min DNI data to the cubic interpolation of the hourly values in the 1-min scale ($\max^{h}_{1-min}$).

The first metric provides a measure of the distribution of the 1-min data within an hourly interval, while the second metric quantifies the highest fluctuation to be expected within an hour. The algorithm requires the specification of the number of clusters. We use three clusters for each interval of k_b in order to have enough data within each of them to be capable to define the cloud transients. This number of clusters is also justified by the average silhouette value of 0.72, which is indicative of a strong structure (Kaufman and Rousseeuw, 1990). The silhouette index (Rousseeuw, 1987) qualifies every point by considering its position with respect to the other points of the cluster to which it belongs and its position with respect to points from other clusters.

$$s(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}} \quad (3)$$

a(i) is the average dissimilarity between observation *i* and all other points of the cluster to which *i* belongs, while *b(i)* is the mean dissimilarity between *i* and its neighbouring cluster, i.e. the nearest one to which *i* does not belong. Observations with large s(i) are well clustered. The average silhouette value can be used to evaluate the quality of the classification.

In Fig. 3 we present the silhouette plot and the cluster groups together with their centroids of the interval 0.4 < k_b ≤ 0.5, respectively. In the silhouette plot, we present the silhouette value of each point. We also present the average silhouette value and the number of points of each cluster, and for the entire set.

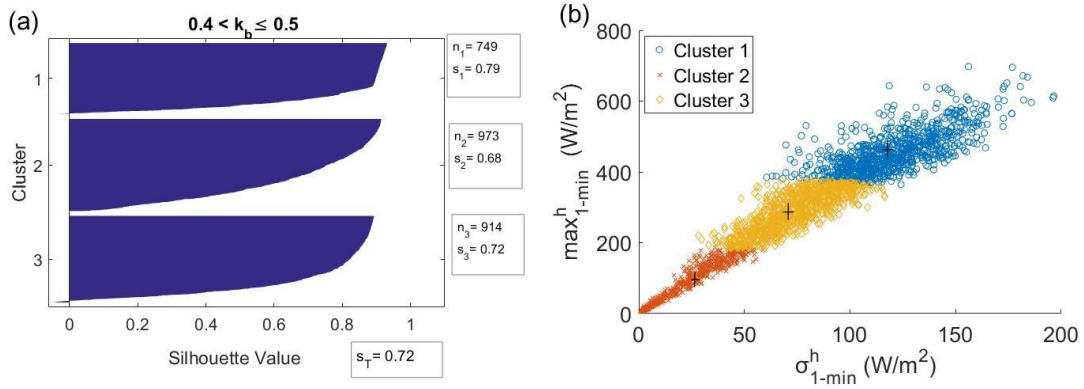


Figure 3. (a) Silhouette plot of the classification obtained for the interval 0.4 < k_b ≤ 0.5. (b) Clusters and centroids for the interval 0.4 < k_b ≤ 0.5.

Twenty-four groups are constructed using this cluster methodology: Three for each k_b bins of width 0.1. Table 3 shows the centroids obtained for each k_b interval. In addition, we present the probability of occurrence of each cluster within each k_b interval in the 14 years of measurements for the location of Seville (Spain).

Table 3. σ_{1-min}^{h} and $\max_{1-min}^{h}$ of the centroids obtained for each label and k_b interval and probability of occurrence of each cluster.

Interval	Label	σ_{1-min}^{h} (W/m^2)	$\max_{1-min}^{h}$ (W/m^2)	Probability
$0 < k_b \leq 0.1$	1	6	23	0.68
	2	44	174	0.25
	3	85	397	0.07
$0.1 < k_b \leq 0.2$	1	72	301	0.37
	2	30	116	0.39
	3	120	523	0.24
$0.2 < k_b \leq 0.3$	1	127	520	0.26
	2	77	303	0.4
	3	27	106	0.34
$0.3 < k_b \leq 0.4$	1	76	292	0.41
	2	124	477	0.29
	3	23	89	0.31
$0.4 < k_b \leq 0.5$	1	122	456	0.35
	2	20	83	0.28
	3	73	289	0.37
$0.5 < k_b \leq 0.6$	1	74	302	0.34
	2	20	79	0.34
	3	124	495	0.31
$0.6 < k_b \leq 0.7$	1	124	538	0.27
	2	15	59	0.41
	3	67	280	0.32
$k_b > 0.7$	1	133	599	0.23
	2	67	291	0.29
	3	18	70	0.48

3.1.3 Stochastic component reproduction

In the previous model (Larraneta et al., 2015), the stochastic component was reproduced by means of beta distributions fitted to the normalized standard deviations of the instant values to the hourly means, being the sign of the fluctuation randomly added from a normal distribution. We found that, in some cases, we might reproduce values that never happened because of the combination of the beta fit and the assumption of a normal distribution of the fluctuation sign.

To solve this weakness, we calculate the stochastic component based on the common practice for the dynamic generation of synthetic data from Markov models (Ngoko et al, 2014). To this end, we calculate the Empirical Cumulative Distribution Function (ECDF) of the differences between the 1-min DNI data to the cubic interpolation of the hourly values in the 1-min scale (dif_{1-min}) of each cluster. The procedure that has been already used for the synthetic generation of hourly DNI datasets (Larrañeta et al., 2017 (b)), is described below:

i. Calculate the ECDF of the dif_{1-min} for each cluster (Section 3.1.2).

ii. Select the group from which fluctuations are going to be reproduced: To this end, we generate a random number *Y* from a uniform distribution between 0 and 1 for each hour *h*, and locate the cluster whose probability of occurrence is the same as the generated with the random number in the training dataset.

iii. Estimate the stochastic component of the 1-min synthetic data: We generate random numbers *R* from a uniform distribution between 0 and 1 for each instant *i*, and locate the value, within the previously selected cluster (step ii), whose cumulative probability is the same as the generated with the random number *R* obtaining thus the stochastic component (Ibn^{i}_{stoc}). Fig. 4 shows a graphical explanation of the step iii.

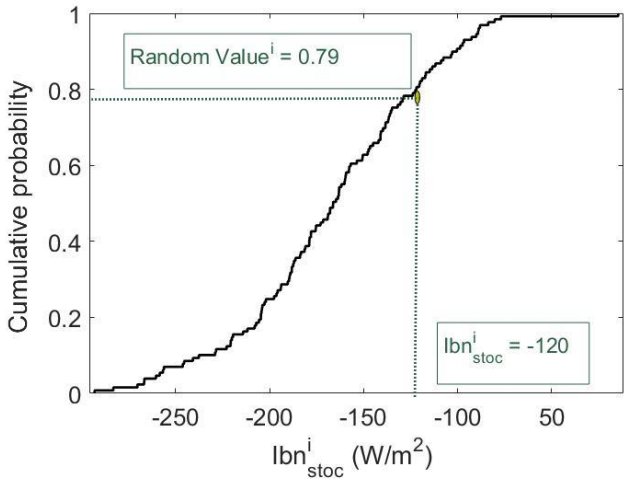


Figure 4. Graphical reproduction of step iii of the stochastic component. For a randomly generated probability of 0.79 we obtain a stochastic component of -120 W/m2.

In the new scheme of the model used in this work, the ECDFs of 14 years of measured data already include positive and negative fluctuations of the instant data from the mean, and thus it is not necessary to generate a random sign for the fluctuation.

As most of the cloud transients at 1-min time resolution are gradual (Larrañeta et al., 2017 (c)), we have limited the random number *R* from the uniform distribution curve (step iii) of two consecutive instants to a maximum difference of ±0.3.

3.2 Non.Dimensional (ND) model

The original method consists in the adimensionalization of the daily DNI curve by the clear-sky envelope approach, creating daily Dynamic Paths from observed hourly DNI data (Fernández-Peruchena et al., 2015). The method transforms each daily 1-min DNI curve into a dimensionless curve where the time scale and the DNI scale go from 0 to 1. In Fig. 5, we show the dimensionless daily shape for the same curve as presented in Fig. 1. For the synthetic generation of 1-min data, days were generated following the next steps:

1. Calculate the clear sky DNI envelopes. To this end, they adjust the two parameters defining the ASHRAE exponential model (MacPhee, 1972) for each day.

2. Generate a database of dimensionless daily curves of the location under study. Normalize the measured data in terms of time and energy. At least one year of 1-min measured solar radiation data was required.
3. Generation of synthetic 1-min DNI series on a given day. Combine envelopes and dimensionless daily DNI curves until the closest Euclidean distance between the daily and hourly means of the synthetic and measured series is found.

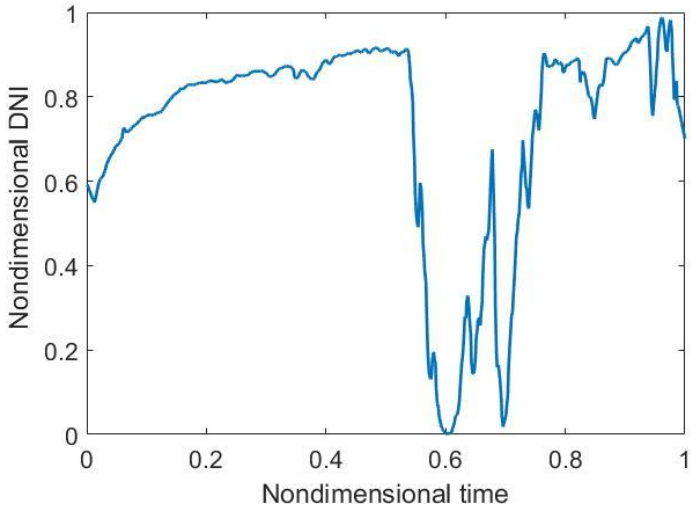


Fig 5. Dimensionless daily shape example of the same curve as presented in Fig. 1.

We found that the third step demands a high computational cost and requires a high-resolution solar radiation database, of at least one year, in the location under study. The modifications proposed here have been implemented in order to speed up the generation procedure. We propose a simpler clear sky envelope method and the use of dimensionless curves of the database of Seville, to generate synthetic high temporal resolution DNI data in any location.

3.2.1 Clear sky envelope

For the calculation of the nondimensionlized daily data packs, we use the envelope clear sky concept (Gómez Camacho and Blanco Muriel, 1990) to determine the maximum daily envelope clear sky irradiance ($\mathrm{I_{bn_{cs}}}$). We implement the clear sky A-B model (Larrañeta et al. 2017 (a))

$$\mathrm{I_{bn_{cs}}} = \mathrm{I_{sc}} \cdot \mathrm{E_0} \cdot \frac{\mathrm{A}}{1 + \mathrm{B} \cdot \mathrm{m_R}}, \tag{3}$$

where $\mathrm{m_R}$ is the relative air mass determined according to the expression of Kasten and Young (1989), I_{sc} is the solar constant, E_0 the correction due to Earth-Sun distance and A and B are empirical parameters intended to model the turbidity of the atmosphere.

We intend to characterize a unique set of parameters (A, B) for each year in a given location. These parameters are annually estimated by adjusting them to the maximum observed values of the relation I_{bn}/E_0 for every value of solar elevation higher than 5°. We use the relationship I_{bn}/E_0 to become A and B parameters independent of the day of the year.

A daily clear sky curve in the hourly resolution is lower than in the 1-min resolution as it is shown in Fig. 6 because the hourly resolution trends to smooth the amplitude of the cloud transients. In Table 4, we present the annual A-B parameters in the 1-min and the hourly resolution for the 14 years used for the training of the model in Seville.

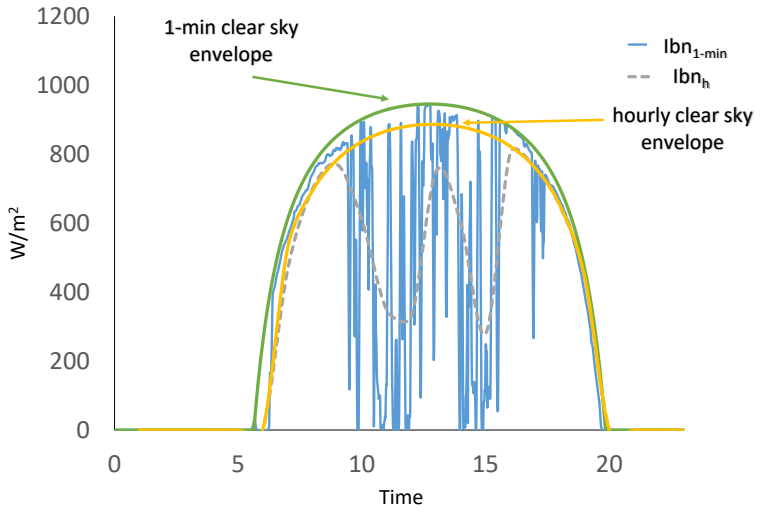


Fig 6. Daily clear sky curve fitted to measured 1-min and hourly DNI data.

Table 4. Annual 1-min and hourly A-B clear sky parameters for the years 2002-2016 at Seville.

	1-min resolution		1-h resolution	
Year	A	B	A	B
2002	0.829	0.119	0.864	0.178
2003	0.812	0.109	0.825	0.145
2004	0.837	0.113	0.859	0.165
2005	0.848	0.115	0.865	0.155
2006	0.844	0.115	0.843	0.155
2007	0.836	0.129	0.839	0.157
2008	0.851	0.130	0.860	0.166
2009	0.835	0.121	0.849	0.164
2010	0.830	0.104	0.861	0.164
2011	0.838	0.143	0.847	0.178
2012	0.824	0.117	0.835	0.144
2013	0.822	0.109	0.833	0.135
2014	0.814	0.116	0.857	0.168
2015	0.848	0.130	0.848	0.155

To generate a given day, we use a random iterative procedure that modifies the A and B parameters until the cumulative value of the available 1-h annual set and the synthetically generated 1-min annual set differ in less than 0.2%. The iterative process requires an initial A-B pair of values. We calculate them using the envelope clear sky method for the 1-h annual dataset used as input of the model.

3.2.2 Most similar day selection

In the original model (Fernández-Peruchena et al., 2015), the most similar day selection was carried out taking into account the similarity of the cumulative values between a given day and the database used for the training. Based on Moreno et al. (2017), we characterize the daily DNI curve shapes by means of the energy, variability and distribution. We use the daily direct fraction index to characterize the daily energy of a given day following next equation:

$$k_b = \frac{H_{bn}^d}{H_{cs}^d} \ , \tag{4}$$

where H_{bn}^d is the daily DNI and H_{cs}^d is the daily DNI under clear sky conditions. For the characterization of the variability, we use the Variability Index (VI) (Stein et al., 2012) defined as the ratio between the length of the DNI curve and the length of the maximum enveloping clear day curve calculated in Section 3.2.1.

$$VI = \frac{\sum_{k=2}^{n} \sqrt{\left(I_{bn_k} - I_{bn_{k-1}}\right)^2 + \Delta t^2}}{\sum_{k=2}^{n} \sqrt{\left(I_{cs_k} - I_{cs_{k-1}}\right)^2 + \Delta t^2}} \tag{5}$$

I_{cs} is the hourly enveloping clear sky direct normal irradiance, Δt refers to an interval of one hour, and n is the number of 1-hour intervals of the considered day. For the characterization of the temporal distribution, we use the morning fraction index F_m defined as the ratio between the accumulated DNI in the first half of the day and the accumulated DNI for the whole day.

$$F_m = \frac{H_{bn\_m}^d}{H_{bn}^d} \tag{6}$$

$H_{bn\_m}^d$ is the DNI recorded from the sunshine to the solar noon and H_{bn}^d is the daily DNI.

For the selection of the most similar day, we use the K Nearest Neighbour (kNN) classification algorithm of a supervised Learning Machine (Fix and Hodges, 1951). kNN categorizes objects based on the classes of their nearest neighbours in the dataset. kNN predictions assume that objects near each other are similar. We train the algorithm with the database of Seville. We use the daily triples of k_b, VI and F_m as predictors with the same weight and we assign one class label to each day. In a second step, we calculate these daily triples of indexes for each curve of the input dataset and then we predict the most similar day with the trained learning machine. The output of the algorithm would be the class label found to be the most similar day of the database of Seville to the input day.

3.3 Statistical indicators

To evaluate the distribution we calculate the KSI (Kolmogorov-Smirnov test integral) index that is defined as the integrated differences between the CDFs of the two data sets (Espinar et al., 2009). The unit of this index is the same for the corresponding magnitude

$$KSI = \int_{x_{min}}^{x_{max}} D_n dx, \tag{7}$$

where, x_{max} and $x_{min}$are the extreme values of the independent variable, and D_n are the differences between the CDFs of the measured and synthetic datasets.

The KSI in W/m^2 shows comparable results regardless of the time resolution of the synthetic data. The higher the KSI values, the worse the model fit.

It is a common practice to evaluate a relative value of KSI (%), obtained by normalizing the KSI (W/m^2) to the critical area $a_{critical}$.

$$KSI(\%) = 100 \cdot \frac{\int_{x_{min}}^{x_{max}} D_n dx}{a_{critical}} \quad (8)$$

$$a_{critical} = V_c \cdot (x_{max} - x_{min}) \quad (9)$$

The critical value V_c depends on the population size N and is calculated for a 99% level of confidence as:

$$V_c = {}^{1.63}/_{\sqrt{N}} \quad N \geq 35 \quad (10)$$

This relative metric should be used with care: When evaluating long time high-resolution datasets, the KSI (%) will result in high values even for the same model performance. The more extensive the population size, the lower the critical area (in inverse proportion to $\sqrt{N}$) and consequently the greater the KSI (%) for the same model performance. Therefore, the KSI (%) should only be used to compare datasets of the same length and resolution.

The KSI statistic evaluates the differences of the CDFs of solar radiation measured and synthetic datasets since these differences are assumed to have an impact in plant production. It means that the interest in the DNI high temporal resolution synthetic generation and therefore the simulation of the cloud transients relies not only on the DNI itself, but also mainly on its impact on plant production. In this paper, we propose to compare the synthetic data to the measured data also in terms of thermal power produced in a field of a Parabolic Trough (PT) plant. To this end, we have simulated in NREL´s SAM software (version 2017.1.17) (Blair et al., 2014) a PT plant with a similar configuration to the plant of Andasol 3 (NREL, 2013) currently in operation. The main characteristics are summarized in Table 5.

Table 5. Main technical data used in SAM to model a plant configuration similar to ANDASOL 3.

Parameter	Andasol 3
Net output at design (MWe)	50
Number of loops	156
Collectors per loop	4
Solar field aperture area (m^2)	510,120
HTF	Therminol VP-1
Storage capacity (hours)	7.5

We evaluate the deviations between the modelled thermal power produced in the field with measured and synthetic data using the $NRMSD$ defined as:

$$NRMSD\ (\%) = 100 \cdot \left({}^{RMSD}/_{P_{max} - P_{min}}\right), \quad (11)$$

where P_{max} and P_{min} are the maximum and minimum power production values of the observed dataset, respectively and RMSD is the Root Mean Squared Deviation.

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_m^i - P_s^i\right)^2}, \tag{12}$$

where N is the number of data pairs, P_m^i is the power produced when using the measured DNI as input and P_s^i is the produced when using the synthetic DNI as input in SAM. The evaluation is performed in the hourly and daily resolution. Only daylight hours are considered in this analysis.

For the evaluation of the autocorrelation, we calculate the ramp rates (RRs) as the difference between successive data points over one minute using the eq. 9:

$$RR = \left(\left(I_{bn_k} - I_{bn_{k-1}}\right) - \left(I_{cs_k} - I_{cs_{k-1}}\right)\right)\Big/_{\Delta t}, \tag{13}$$

where Δt refers to an interval of one minute. The units will be given in W/m^2·min. We calculate the absolute RRs values for the annual datasets taking into account only daytime observations for solar elevations above 5°.

The synthetic data generated with the SA model is expected to follow the shape of the measured data hour by hour, but the fluctuations of the solar radiation due to the cloud transients are synthetically reproduced. On the other hand, the ND model would reproduce fluctuations that have been measured but may not follow the hourly distribution of the measured data.

4 Results

Some examples of the daily profiles are presented in Fig. 7 where the goodness of the methods in reproducing the daily shapes of the 1-min DNI is qualitatively illustrated. We show four consecutive days measured at the location of Almeria (blue line) together with the corresponding synthetic data generated by each methodology: SA modelled (left side of Fig. 7, red dotted lines), and ND modelled data (right side of Fig. 7, green dotted line).

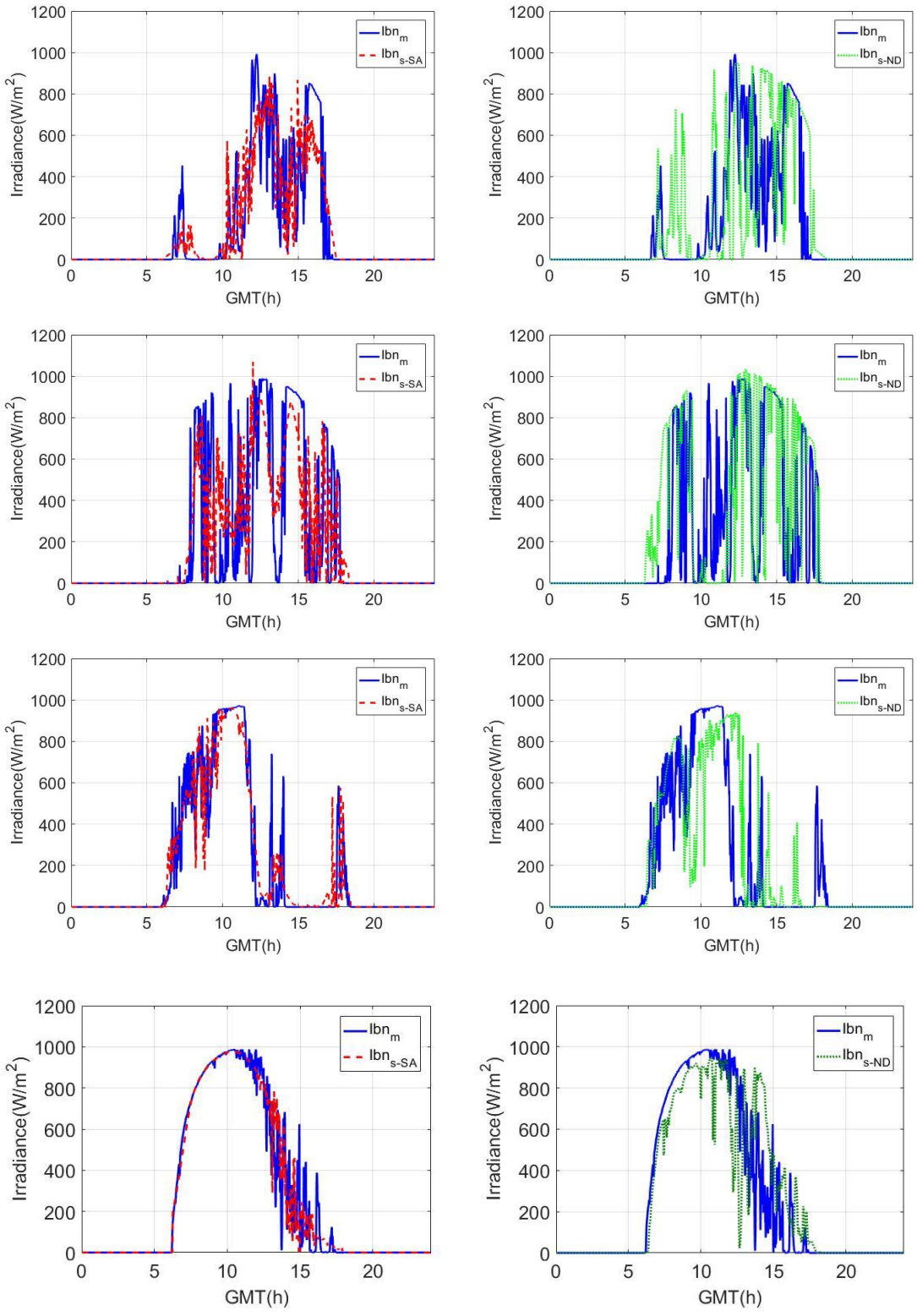


Figure 7. Illustrative examples of the results of the synthetic generation with the stochastic adaptation model ($Ibn_{s\text{-}SA}$) and adimensionalization model ($Ibn_{s\text{-}ND}$) compared to the measured dataset (Ibn_m).

To assess the performance of the models, we evaluate the mean, distribution and autocorrelation of the generated time series in comparison to the measured ones with modelled ones in each location.

The models maintain the mean by definition. The SA model includes an iterative procedure where the daily synthetic series are recalculated until both cumulative daily values, measured and synthetic, differ in less than 2%. The ND model uses an iterative procedure until the cumulative value of the available 1-h annual set and the synthetically generated 1-min annual set differ in less than 0.2%. Fig. 8 shows, the ECDFs of the measured (blue line) and synthetic datasets generated with each model (red and green dotted line) in one-year dataset for the three locations under study.

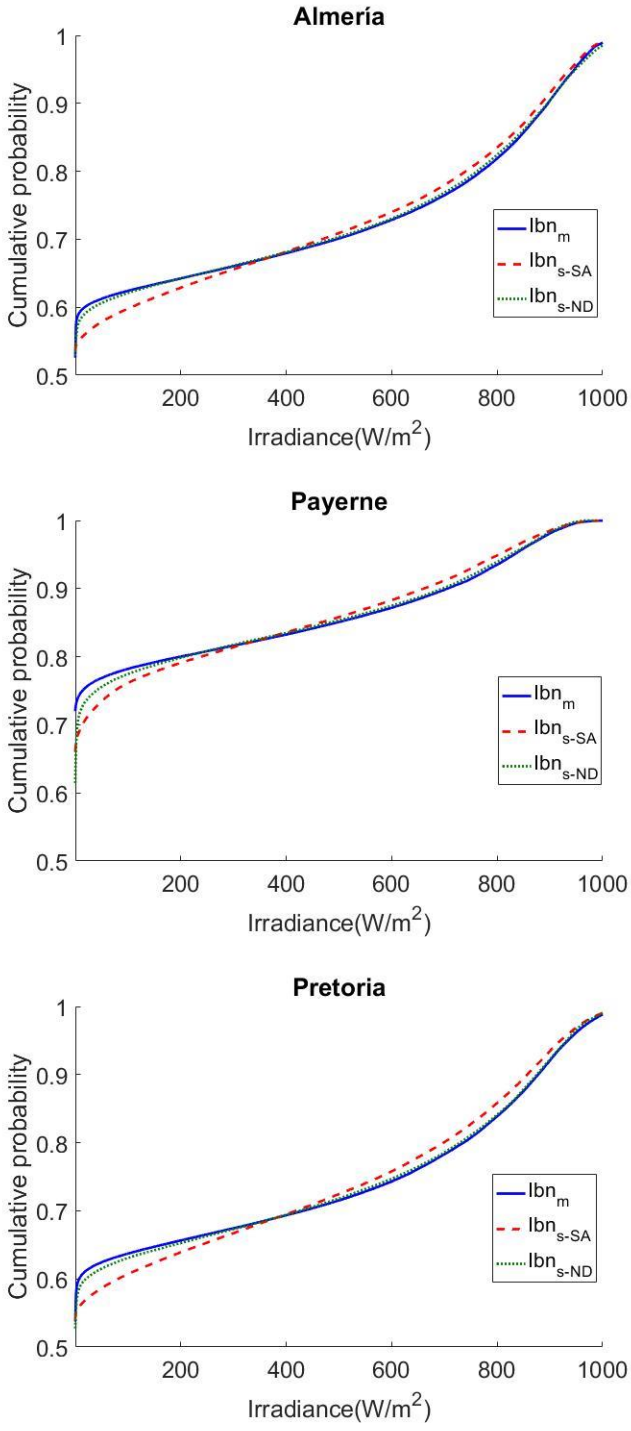


Figure 8. ECDFs of the measured (Ibn_m) and synthetic DNI datasets generated with the stochastic adaptation model ($Ibn_{s\text{-}SA}$) and the adimensionalization model ($Ibn_{s\text{-}ND}$).

Table 6 shows the KSI value of each model in the selected locations.

Table 6. KSI of the implemented models for the measured and synthetic DNI annual sets for each location

Parameter	Station (year)	SA Model	ND Model
KSI (W/m^2)	Almería (2013)	12.7	2.8
	Pretoria (2016)	15.4	3.4
	Payerne (2014)	10.6	3.6

Fig. 9 and Fig. 10 show a bar plot of the the monthly KSI (W/m^2) and KSI (%) of the DNI time series for the three locations under evaluation.

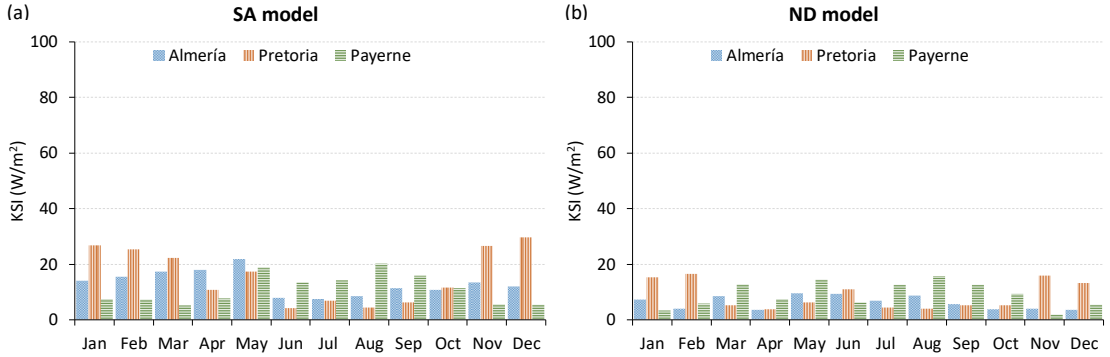


Figure 9. KSI (W/m^2) values obtained in the comparison of 1-min synthetic DNI data compared to the measured data for the locations of Almería, Pretoria and Payerne with the SA model (a) and the ND Model (b).

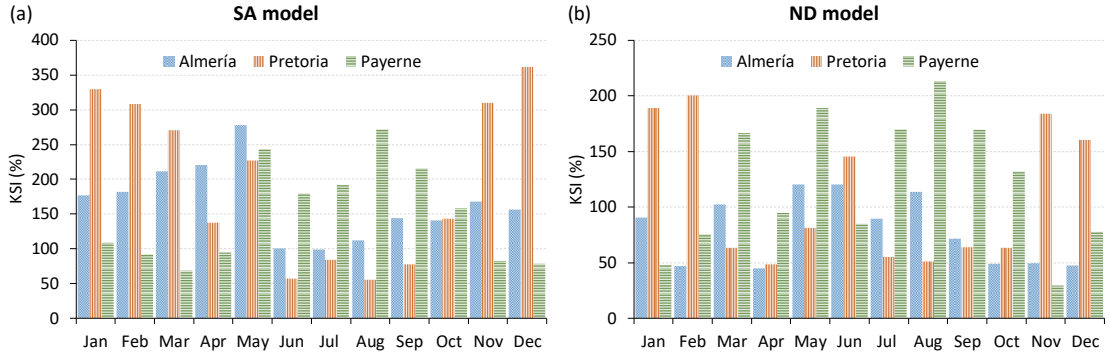


Figure 10. KSI (%) values obtained in the comparison of 1-min synthetic DNI data compared to the measured data for the locations of Almería, Pretoria and Payerne with the SA model (a) and the ND Model (b).

The SA method provides average monthly KSI values of 11.2 W/m^2 in Payerne, 16.0 W/m^2 in Pretoria and a maximum value of 29.6 W/m^2 in Pretoria. The ND model provides average monthly KSI values of 6.2 W/m^2 in Almería, 9.0 W/m^2 in Payerne and a maximum value of 16.5 W/m^2 in Pretoria.

The greater KSI values are found in those months having a large number of high variability days, suggesting that in the case of the ND model, the characterization of the intra-daily variability

through the VI index may not be sufficient. Moreover, the use of hourly values in the calculation of the variability index entails a loss of information. This is also the case of the SA model: the grater the intra-daily DNI variability, the worse the model performance. The stochastic component reproduction (section 3.1.3) may not be accurate enough since it is independent of the intra-daily variability. Future improvements of this method should focus on this issue.

Future works will model the relation of the variability index on different time resolutions and would characterize the variability of the solar radiation from two indexes, one giving information about the number of fluctuations and the other giving information about the amplitude of the fluctuations.

The results for the NRMSD of the thermal power produced in the three analysed locations are presented in Table 7.

Table 7. $NRMSD$ of the modelled thermal power produced with the measured and synthetic datasets in the proposed PT plant.

Parameter	Station (year)	SA Model	ND Model
$NRMSD_{hourly}$ (%)	Almería (2013)	2.3	12.4
	Pretoria (2016)	2.7	12.8
	Payerne (2014)	2.3	14.7
$NRMSD_{daily}$ (%)	Almería (2013)	0.8	3.0
	Pretoria (2016)	1.2	3.1
	Payerne (2014)	0.8	4.1

The ND model synthetic data presents the best results in terms of frequency distribution for the location of Almería, while the SA model presents a higher KSI in all the locations. In Fig. 8, we can observe that the greatest differences are found for low levels of irradiance. When evaluating the NRMSD of the thermal energy produced in a PT plant with a common configuration, we observe that the SA model shows better performance. This occurs because it synthetizes the data hour by hour intending to follow the daily shape of the DNI while the ND model uses daily parameters to select the day to be generated. The hourly NRMSD is greater in the ND model than in the SA model in days with an uncoupled synthetic and measured variability, which leads to greater differences in plant performance.

Fig. 11 shows the ECDFs of the measured and synthetic absolute RR values generated with each model (left) and their differences (right) in one complete year for each location. Fig. 12 shows a bar plot of the the monthly KSI values for the absolute RR time series of each model in the selected locations. Both modelled 1-min DNI datasets show ECDFs similar to the measured one. For high RRs (> 300 W/m^{2}min), the differences in the measured and synthetic ECDFs values are almost negligible. The differences found at low RR indicate a lower variability of the measured dataset with respect to both modelled datasets.

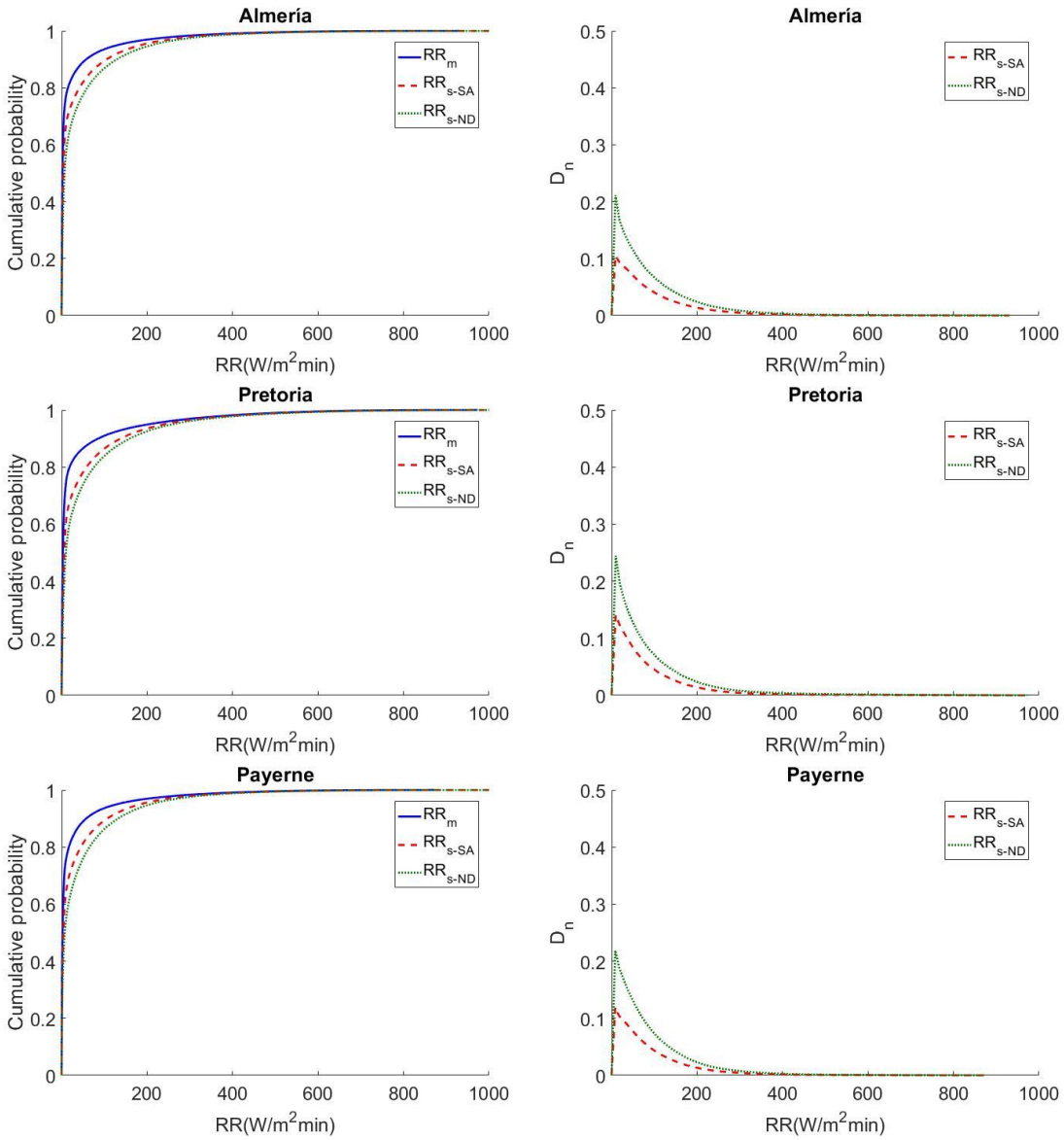


Figure 11. ECDFs of the measured (RR_m) and synthetic absolute RR datasets generated with the stochastic adaptation model ($RR_{s\text{-}SA}$) and adimensionalization model ($RR_{s\text{-}ND}$) (left) and their differences (right).

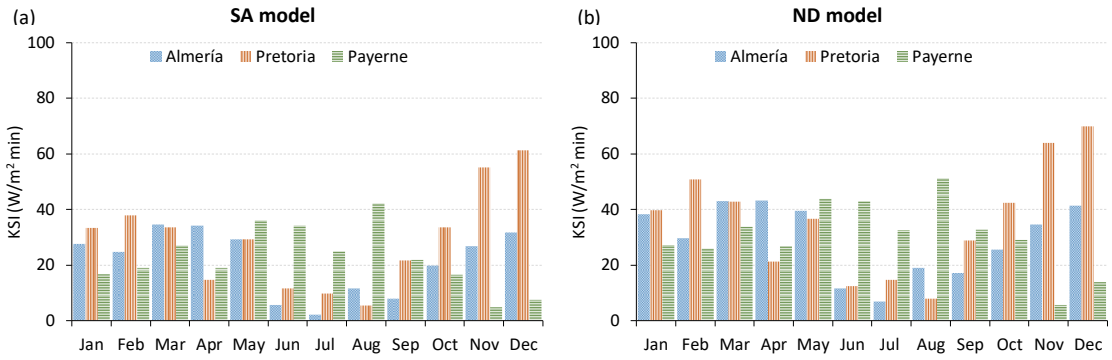


Fig 12. Monthly KSI values for the absolute RR time series of each model in the selected locations obtained with the SA model (a) and the ND Model (b).

5 Conclusions

In this paper, we present two methodologies for generating synthetic DNI data at 1-min time resolution in the absence of local high-frequency measured DNI series. Their global applicability is assured taken into account the flexibility of inputs: from accurate hourly DNI series (the SA method), to intra-daily characterization of DNI variability and distribution (the ND method), where exact hour-to-hour local DNI evolution is not required.

These methodologies are based on previous works (Larrañeta et al., 2015; Fernández-Peruchena et al., 2015)), where some modifications have been applied in order to speed up its application. Larrañeta et al. (2015) methodology modifications are focused on the characterization and reproduction of the stochastic component of the solar radiation (SA model); Fernández-Peruchena et al. (2015) methodology modifications are focused on the characterization of the daily profiles in terms of energy, variability and distribution (ND model).

SA and ND have been trained with 14 years of measured 1-min DNI data for the location of Seville (Spain), and require no previous knowledge of the high-frequency patterns of solar irradiance at these sites (i.e., no local measurements of high-frequency solar irradiance are required to apply them). To address their performance, we compare the outputs of the proposed methodologies in three locations with diverse climatic conditions. The comparison of the frequency distribution of the synthetic DNI data compared to the measured ones reveals that the ND methodology shows a better performance (KSI ~3.3 W/m^2) compared with the SA methodology (KSI ~12.9 W/m^2). We also propose a comparison in terms of the thermal power produced in a field of a PT plant with a common configuration. In this case, we observe that the SA methodology provides a lower hourly NRMSD (~0.9 %) compared with the ND methodology (~3.4 %). Consequently, we may not assume that a closer distribution of the synthetic solar radiation dataset in comparison with the measured dataset brings more accurate results in terms of the power produced by a CSP plant. These results suggest that the high-frequency DNI generation procedure should have into consideration the hourly time series at the site.

This article shows, for the first time, the generation of 1-min DNI series in different locations without any local adaptation or calibration. Sites selected for this generation are located at different climates and latitudes, suggesting a global applicability of the methodologies. Future works would include the extension of the methodologies for generating coupled pairs of both GHI and DNI components, to fully characterize the high-frequency solar resource at a given site and would perform a deeper characterization of the daily variability of the solar radiation.

Acknowledgments

The authors are grateful to S. Wilbert (DLR) and L. Ramirez (CIEMAT) for providing the data from Almería.

References

Aguiar, R., Collares-Pereira, M., 1992. Statistical properties of hourly global radiation. Solar Energy 48, 157–167.

Al-Shammari, E.T., Shamshirband, S., Petković, D., Zalnezhad, E., Yee, P.L., Taher, R.S., Ćojbašić, Ž. 2016. Comparative study of clustering methods for wake effect analysis in wind farm. Energy 95, 573–579.

Beyer, H.G., Fauter, M., Schumann, K., Schenk, H., Meyer, R. 2010. Synthesis of DNI time series with sub-hourly time resolution. In: 16th SolarPACES Conference. Perpignan (France).

Blair, N., Dobos, A.P., Freeman, J., Neises, T., Wagner, M., Ferguson, T., Gilman, P., Janzou, S., 2014. System Advisor Model, SAM, 2014.1.14: General Description. Technical Report, NREL/TP-6A20-61019, Golden, USA.

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.

Bright, J.M., Babacan, O., Kleissl, J Taylor, P.G., Crook, R., 2017. A synthetic, spatially decorrelating solar irradiance generator and application to a LV grid model with high PV penetration. Solar Energy 115 229–242.

Brooks, M. J., du Clou, S., van Niekerk, J. L., Gauche, M. J., Leonard, P., Mouzouris, C., Meyer, A. J., van der Westhuizen, E. E., van Dyk, N., Vorster, F. 2015 “SAURAN: A new resource for solar radiometric data in Southern Africa,” J. Energy South. Africa, vol. 26, pp. 2–10.

Espinar, B., Ramírez, L., Drews, A., Beyer, H.G., Zarzalejo, L.F., Polo, J., Martín, L., 2009. Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. Solar Energy 83 (1), 118–125.

Fernández-Peruchena, C. M. F., Gastón, M., Schroedter-Homscheidt, M., Kosmale, M., Marco, I. M., García-Moya, J. A., & Casado-Rubio, J. L. (2017). Dynamic Paths: Towards high frequency direct normal irradiance forecasts. Energy, 132, 315-323

Fernández-Peruchena, C.M., Gastón, M., 2016. A simple and efficient procedure for increasing the temporal resolution of global horizontal solar irradiance series n. Renew. Energy 86, 375–383.

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.

Fernandez-Peruchena, C., Ramírez, L., Blanco, M., Bernardos, A., 2010. Variability in global and direct irradiation series generation: scope and limitations. In: 16th SolarPACES Conference. Perpignan (France).

Fix, E., Hodges, and J.L. 1951. An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges International Statistical Review/Revue Internationale de Statistique Vol. 57, No. 3, pp. 233-238

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.

Gómez Camacho, C., Blanco Muriel, M., 1990. Estimación de la atmósfera estándar de radiación solar a partir del concepto de día claro envolvente. Aplicación a la Plataforma Solar de Almería. Era Solar 40, 11–14.

Gordon, J.M., Reddy, T.A., 1988. Time series analysis of hourly global horizontal solar radiation. Solar Energy 41 (5), 423–429.

Grantham, A.P., Pudney, P.J., Ward, L.A., Belusko, M., Boland, J.W., 2017. Generating synthetic five-minute solar irradiance values from hourly observations. Solar Energy 147, 209-221

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.

Graham, V.A., Hollands, K.G.T., 1990. A method to generate synthetic hourly solar radiation globally. Solar Energy 44 (6), 333–341.

Hammer, A., Lorenz, E., Kemper, A., Heinemann, D., Beyer, H.G., Schumann, K., Schwandt, M., 2009. Direct normal irradiance for CSP based on satellite images of Meteosat Second Generation. In: 15th SolarPACES Conference. Berlin, Germany.

Han, J., Kamber, M., 2001. Data Mining. Concepts and Techniques. Chapter 8. Cluster analysis, 1st Editio. Ed. Morgan Kaufmann Publishers.

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

Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. N.J.

Larrañeta, M., Reno, M.J., Lillo-Bravo, I., Silva-Pérez, M.A. 2017 (a). Identifying periods of clear sky direct normal irradiance. Renewable Energy, 113, pp. 756-763.

Larrañeta, M., Moreno-Tejera, S., Lillo-Bravo, I., Silva-Pérez, M.A. 2017 (b). A methodology for the stochastic generation of hourly synthetic direct normal irradiation time series. Theoretical and Applied Climatology, pp. 1-11. Article in Press.

Larrañeta, M., Moreno-Tejera, S., Lillo-Bravo, I., Silva-Pérez, M.A. 2017 (c). Cloud transient characterization in different time steps. AIP Conference Proceedings, 1850, art. no. 140016,

Larrañeta, M., Moreno-Tejera, Silva-Pérez, M.A., Lillo-Bravo, I. 2015. An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series. Solar Energy 122. 517–528.

MacPhee, C., 1972. ASHRAE Handbook of Fundamentals. R.a.C.E. American Society of Heating, NY.

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

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

Moreno-Tejera, S., Silva-Pérez, M.A., Ramírez, L., Lillo-Bravo, I., 2017. Classification of days according to DNI profiles using clustering techniques. Solar Energy 146. 319–333.

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.

NREL. 2013. Concentrated Solar Power Projects. Andasol-3 [WWW Document]. URL https://www.nrel.gov/csp/solarpaces/project_detail.cfm/projectID=117

Ohmura, A., Gilgen, H., Hegner, H., Müller, G., Wild, M., Dutton, E. G., ... & König-Langlo, G. (1998). Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research. Bulletin of the American Meteorological Society, 79(10), 2115-2136.

Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Hoff, T., 2011. Parameterization of site-specific short-term irradiance variability. Solar Energy 85, 1343–1353.

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.

Ramirez, Lourdes and Nielsen, K.P. and Vignola, Frank and Blanco, Manuel and Blanc, Philippe and Meyer, Richard and Wilbert, Stefan, 2017. Road Map for Creation of Advanced Meteorological Data Sets for CSP Performance Simulations. IEA SolarPACES report.

Ramírez, L., Pagh Nielsen, L., Vignola, F., Blanco, M., Blanc, P., Meyer, R., Wilbert, S., 2017. Road Map for Creation of Advanced Meteorological Data Sets for CSP Performance Simulations. IEA SolarPACES report.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.

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

Stein, J., Hansen, C., Reno, M., 2012. The variability index: a new and novel metric for quantifying irradiance and PV output variability. World Renew. Energy, 1–7.

Vuilleumier, L., Hauser, M., Félix, C., Vignola, F., Blanc, P., Kazantzidis, A., Calpini, B. 2014. “Accuracy of ground surface broadband shortwave radiation monitoring,” J. Geophsical Res. Atmos., vol. 119, pp. 1365–1382.