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

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

A methodology for the stochastic generation of hourly synthetic direct normal irradiation time series

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

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

Many of the available solar radiation databases only provide global horizontal irradiance (GHI) while there is a growing need of extensive databases of direct normal radiation (DNI) mainly for the development of concentrated solar power and concentrated photovoltaic technologies. In the present work, we propose a methodology for the generation of synthetic DNI hourly data from the hourly average GHI values by dividing the irradiance into a deterministic and stochastic component intending to emulate the dynamics of the solar radiation. The deterministic component is modeled through a simple classical model. The stochastic component is fitted to measured data in order to maintain the consistency of the synthetic data with the state of the sky, generating statistically significant DNI data with a cumulative frequency distribution very similar to the measured data. The adaptation and application of the model to the location of Seville shows significant improvements in terms of frequency distribution over the classical models. The proposed methodology applied to other locations with different climatological characteristics better results than the classical models in terms of frequency distribution reaching a reduction of the 50% in the Finkelstein-Schafer (FS) and Kolmogorov-Smirnov test integral (KSI) statistics.

1 Introduction

The direct component of the solar radiation is the relevant component for Concentrated Solar Power (CSP) and Concentrated Photovoltaic (CPV) technologies. Long–term series of Direct Normal Irradiance (DNI) measurements are only available for a limited number of places around the world. In general, the available DNI series cover relatively short time periods and show more gaps than Global Horizontal Irradiance (GHI) series (Roesch et al. 2011) since the measurement equipment of DNI (tracker and pyrheliometer for direct measurement) is more complex and expensive and requires more attention. DNI, as well as Diffuse Horizontal Irradiance (DHI), is directly related with GHI. Many models to estimate DNI and DHI series from GHI series have been developed for decades. Boland et al. (2013) verified that estimating DHI from GHI, and then calculating the DNI from the estimated DHI, accomplishes similar results as modeling the DNI directly from GHI.

These models, commonly called separation or decomposition models in the literature, relate GHI with its components by mean of dimensionless indexes: the clear sky index (k_t), the diffuse fraction (k_d) and the direct fraction (k_b or k_n) are some examples. These indexes were introduced by Liu and Jordan (1960) in daily scales. Later, other authors have extended their use to other time scales (Tovar-Pescador 2008), mostly in hourly time resolution. In their simplest form, these models are presented as first-order empirical fittings

between k_t and k_d or k_b indexes (Orgill and Hollands 1977). Other authors (Erbs et al. 1982; Louche et al. 1991; Oliveira et al. 2002; Jacovides et al. 2006) have developed higher-order models. Both types of models usually divide the k_t range in several segments according to the shape of the scatter plot. Boland et al. (2001) propose a logistic curve model avoiding segmentation. In general, these models with only one predictor provide good outcomes when monthly and annual estimated data are compared with measured values but not so good in shorter time scales.

The value of DNI for a given value of GHI varies considerably depending on the atmospheric conditions (clouds, aerosols, etc.). Aerosols, that derive mainly from natural but also from anthropogenic (Das et al. 2016) emission sources, have great impact in local climate (Singh et al. 2017) and therefore in solar radiation reaching the surface. Absorbing aerosols (BC, organic carbon, dust etc.) primarily influence the solar radiation budget by absorbing radiation at all wavelengths while scattering aerosols (seas salts, sulphates, nitrates, etc.) scatter the radiation.

Empirical models as the ones referred above are location-dependent (Paulescu and Blaga 2016; Gueymard and Ruiz-arias 2014) and need to be verified, and sometimes adjusted, when used in places with different climate conditions. The scatter plot shape of the hourly k_d or k_b indexes against the hourly clearness index shows another drawback of these fittings. The wide spread of points in the central region suggests that other type of model is necessary for an accurate estimation of the solar radiation components from GHI measurements. Some authors propose to add more predictors to the model. In this regard, many options are found in the literature: astronomical parameters, as solar altitude and air mass (Perez et al. 1990; Maxwell 1987; Reindl et al. 1990); meteorological parameters, as air temperature and relative humidity (Reindl et al. 1990); and others parameters related with the variability of the solar radiation as variability, persistence or stability index (Ridley et al. 2010; Perez and Ineichen 1992; Skartveit et al. 1998). Some interesting studies compare models with different number of predictors (Torres et al. 2010; Jacovides et al. 2006; Gueymard and Ruiz-arias 2014; Paulescu and Blaga 2016; Dugaria et al. 2015). Paulescu and Blaga (2016) analyze the improvements in the performance of diffuse fraction models testing different predictors and comparing the results with similar models fitted in other climate areas. The study points out that increasing the numbers of predictors might improve the local performance of the model but degrade it when applied in other locations. Gueymard and Ruiz-arias (2014) find similar conclusions when validating 36 diffuse and direct fraction models with data from arid or desert areas According to Furlan et al. (2012) predictors which inform about the state of clouds are more relevant than the traditional meteorological variables and pollution indicators in separation models. Models that include indexes related with the variability/dynamics of the solar radiation as predictors exhibit the best results in other studies (Gueymard and Ruiz-arias 2014; Torres et al. 2010). Torres et al. (2010) show a good concordance between the probability distribution function of the diffuse irradiance values measured and calculated with two popular models which also include this type of predictors (Perez and Ineichen 1992; Ridley et al. 2010). This comparison, not considered in other works, helps to test how the separation model reproduces the dynamic performance of the radiation components.

The common intention when improving a separation model is to reduce statistical indicators such as RMSE, MAE, BIAS or R^2 and no attention is generally placed in the frequency distribution of the modeled data. Using data with unrealistic frequency distributions as input for CSP simulation software leads to unrealistic energy yields (Silva-Perez et al. 2014) reaching differences of up to 9% for sites with a similar annual DNI (Chhatbar and Meyer 2011).

In this paper, we present a novel methodology for the synthetic generation of hourly DNI values from hourly GHI data that keeps the same frequency distribution as the measured DNI data while the results in terms of daily, monthly and annual deviations are similar to those of the most common models. The methodology, which is based on the separation of the solar radiation into a deterministic and a stochastic component, has been validated in four locations with different climatic conditions, showing a satisfactory performance regardless of the location where applied.

2 Meteorological databases

The data set used consists of hourly average values of global and direct solar radiation recorded during 15 years (2000–2014) for the location of Seville. The measurements were taken with a sampling and storing frequency of 0.2 Hz. The GHI has been measured with a secondary standard pyranometer Kipp&Zonen CMP21. A first class Eppley NIP pyrheliometer coupled to 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 follow a maintenance and calibration procedure according to the recommendations of the instrument manufacturers. Data have been subject to quality-control procedures following the Baseline Surface Radiation Network (BSRN) recommendations (McArthur 2004) applying the tool proposed by Moreno-Tejera et al. (2015). Data recorded at sun altitudes lower than 10º have not been used in this study (Paulescu and Blaga 2016).

The study has been conducted by splitting the data into a training set and a validation set. The training set used to build and adjust a satisfactory model corresponds to the years 2000-2013, while the validation set used to assess the performance of the model corresponds to the year 2014.

In addition, the model has been validated in three other locations with different climates. The selected sites are presented in Table 1.

Table 1. Locations selected for the model validation.

	Latitude (°N)	Longitude (°W)	Altitude (m)	Climate	Training	Test
Pamplona	42.8	−1.6	450	Atlantic	2010-2011	2012
Pretoria	-25.75	28.22	1410	Sub-Tropical	2013-2014	2015
Payerne	46.81	6.94	491	Continental	2010-2011	2012
Seville	37.4	−6.0	10	Mediterranean	2000-2013	2014

The databases of Pamplona and Payerne (Vuilleumier et al. 2014) belong to the BSRN. The database from Pretoria has been accessed from the Southern African Universities Radiometric Network (SAURAN) (Brooks et al. 2015). The methodology presented in this paper requires a dataset in the corresponding location for the training of the model. The length of training set is recommended in at least two years to avoid an adjustment to a year particularly different from the average, therefore, the databases have also been divided into training (two years for each location) and test sets for the validation of the methodology.

3 Methodology for the synthetic generation

The methodology follows the idea of dividing the solar radiation into a deterministic and a stochastic component. For a given value of global radiation, there is a range of possible direct radiation values conditioned to the state of the atmosphere. The proposed methodology intends to emulate, in a simplified form, the variability of the atmospheric components that affect the attenuation of the direct component. The deterministic component is generated through a classical separation model (Erbs et al. 1982). The stochastic component, which is added to the deterministic component, is calculated from the cumulative frequency distribution (CFD) curve of a sufficiently large database.

3.1 Deterministic component

An extensive database of hourly DNI and GHI data (thirteen years in the case of Seville) is used to generate an empirical separation k_t^h- k_b^h model for solar elevations higher than 10°. k_t^h is the hourly clearness index defined as:

$$k_t{}^h = I_{g0}{}^h / I_0{}^h \quad (1)$$

Where $I_{g0}{}^h$ is the observed hourly average global horizontal irradiance and $I_0{}^h$ is the hourly average extraterrestrial irradiance.

$k_b{}^h$ (Skartveit and Olseth 1992) is the direct fraction index defined as:

$$k_b{}^h = I_{bn}{}^h / I_{bn_{cs}}{}^h \quad (2)$$

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

The purpose of using a clear sky model in Eq. (2) is to provide an "envelope" for the DNI, corresponding to the cleanest clear sky conditions found on our site. To obtain accurate results, an empirical fit of the clear sky DNI with local data is strongly recommended. The clear sky model is adjusted by selecting the maximum DNI values for different solar altitudes and fitting them to an empirical clear sky model. Any of the well-known clear sky DNI models (Behar et al. 2015) could be used in this step. We have implemented the model AB proposed by Silva-Pérez in (Silva-Pérez 2002):

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

Where m_R is the relative air mass determined according to the expression of (Kasten and Young 1989), I_{cs} is the solar constant, E_0 the Earth-Sun distance correction and A and B are empirical parameters intended to model the state of transparency or turbidity of the atmosphere. For the generation of an envelope curve independent of the time of the year, we propose the empirical fit to the maximum DNI values divided by the Earth-Sun distance correction obtained for each solar elevation higher than 5°.

Finally, we calculate the fourth order polynomial fit to the point cloud of $k_t{}^h$ vs $k_b{}^h$ values for solar elevations higher than 10° and $k_t{}^h$ lower than 0.85 to avoid incorrect measurements caused by horizon obstacles and the 'enhancement effect' due to the reflection from the base of the clouds (Tapakis and Charalambides 2014). Figure 1 shows the fourth order polynomial fit to the empirical separation model calculated with thirteen years of hourly data available for Seville.

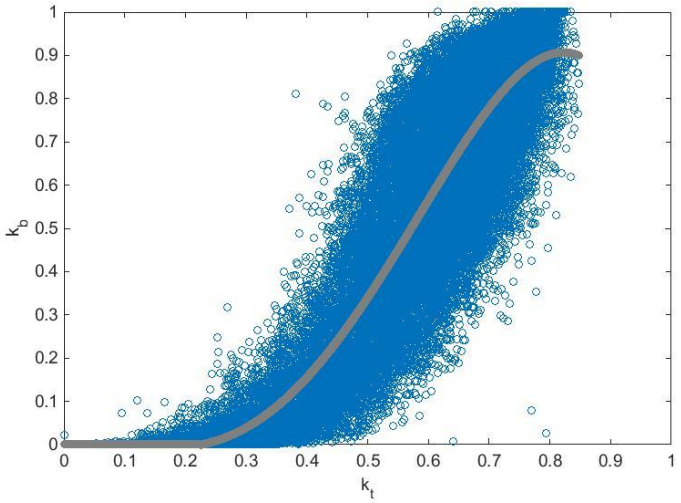

Fig 1. Fourth order polynomial fit to the to the point cloud of $k_t{}^h$ vs $k_b{}^h$ hourly values.

3.2 Stochastic component

The second step for the calculation of the synthetic DNI relies on the frequency distribution of the measured database. For a given $k_t{}^h$ value there is a range of possible $k_b{}^h$ values as it can be observed in Figure 1. These

possibilities are modeled through a fit to the Empirical Cumulative Frequency Distribution Function (ECDF) of the training dataset. For each k_t^h, the stochastic component is fitted to the ECDF of the difference between the k_b^h measured and the k_b^h deterministic ($k_b^h{}_{\text{difference}}$) with the aim of generating only synthetic data that have already been measured. In order to have a statistically significant ECDF, we divide the k_t^h range into intervals of ±0.02 points and we cluster the data into three sun elevations intervals; 10-30°, 30-60°, and 60-90°, understanding that the behavior of the k_b varies significantly depending on this parameter (Tovar-Pescador 2008). To illustrate the procedure, Figure 2 shows an example of the stochastic component ECDF curves for an hourly k_t^h= 0.55 depending on the sun position. For a value of k_t^h= 0.55 we take the $k_b^h{}_{\text{difference}}$ data corresponding to an interval of k_t^h= 0.55 ± 0.02 and the ECDF to which the stochastic component is fitted will also depend on the sun position.

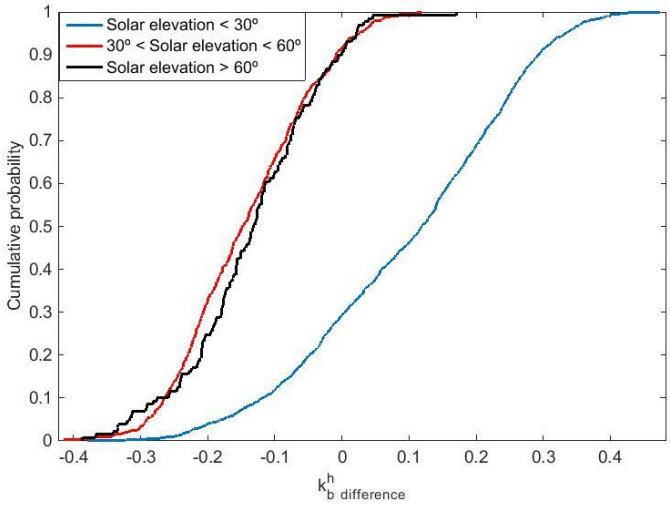

Fig 2. Example of the ECDF curves of the $k_b^h{}_{\text{difference}}$ for k_t^h= 0.55 and for each sun elevation interval.

It can be observed that for sun elevations higher than 30°, most of the differences between the measured k_b^h and the deterministic k_b^h are lower than zero while for the lowest elevations the majority of the differences are higher than zero.

3.3 Synthetic generation

The procedure for the generation of synthetic DNI values given an hourly value of GHI divides the solar radiation into a deterministic and stochastic component. The first is generated by the fourth order polynomial fit of the hourly k_t^h and k_b^h values of an extensive database of at least one year. The stochastic component is dynamically reproduced by using random numbers from the ECDF curve fitted to each $k_b^h{}_{\text{difference}}$ and solar elevation intervals corresponding to each k_t^h. The procedure is described below:

i. Calculate the deterministic component ($k_b^h{}_{\text{det}}$) by the determination of the k_t^h corresponding to the hourly GHI mean and obtain its corresponding k_b^h using the fourth order polynomial fit adjusted to the empirical database (3.1).
ii. Calculate the ECDF of the difference between the k_b^h measured and the $k_{\text{b\_det}}^h$ concerning the k_t^h± 0.02 and its corresponding solar elevation (3.2).
iii. Generate random numbers from a uniform distribution curve [0,1].
iv. Locate the value whose cumulative probability is the same as the generated with the random number thus obtaining the stochastic component ($k_b^h{}_{\text{stoc}}$). Figure 3 presents a graphical explanation of the step iv.

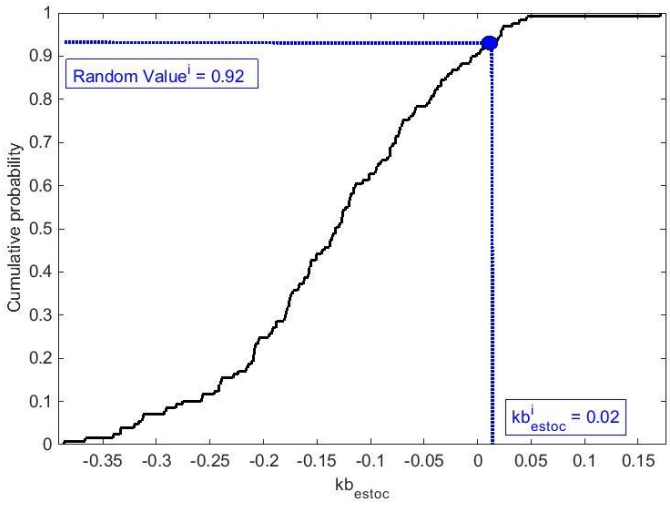

Fig 3. Graphical representation of step iv of the synthetic generation.

To estimate the synthetic value for each hour, the operation is performed by the following equation:

$$k^{h}_{\mathrm{b\,synth}} = k^{h}_{\mathrm{b\_det}} + k^{h}_{\mathrm{b\,stoc}} \tag{4}$$

To avoid undesired fluctuations of the DNI in clear-sky conditions, we have imposed the condition that the $k_t{}^h$ must vary by more than 0.05 points from the previous value; otherwise, the stochastic component $k^{h}_{\mathrm{b\_stoc}}$ remains constant. The $k^{h}_{\mathrm{b\_synth}}$ is limited to a maximum value of one.

Finally, the hourly synthetic data I^{h}_{synth} is calculated by multiplying the $k^{h}_{\mathrm{b\_synth}}$ by the clear sky irradiance value for that hour.

$$I^{h}_{\mathrm{synth}} = I^{h}_{bn_{cs}} \cdot k^{h}_{\mathrm{b\_synth}} \tag{5}$$

To avoid negative results, there is a minimum value of zero during daytime and an imposed value of zero during nighttime.

4 Results and discussion

The analysis of the results has been made in four time scales, hourly, daily, monthly and yearly for the year 2014 in the location of Seville. Due to the stochastic nature of the methodology, simulating only one year may lead to unrepresentative results. Therefore, we have generated synthetically the DNI of the year 2014 fifty times for the yearly and monthly analysis. The results are presented in a box plot together with the measured monthly values in Figure 4.

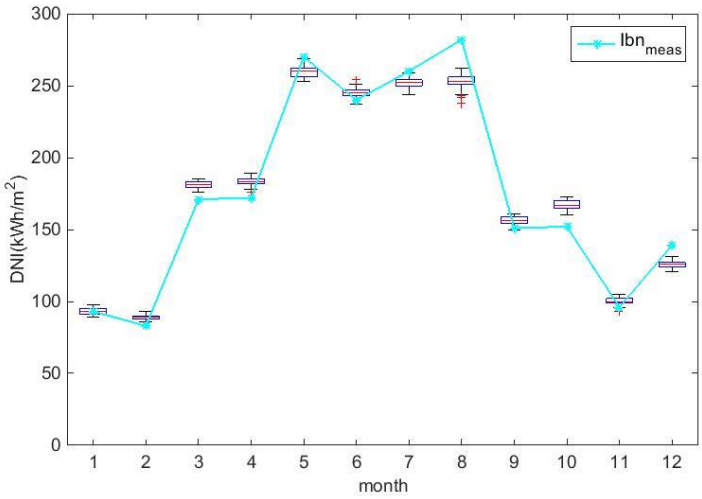

Fig 4. Monthly Box plot of fifty simulations of the synthetic generation of the DNI for the year 2014 in comparison with the measured DNI.

The annual mean of the synthetic data differs in less than 0.5% from the measured value. The monthly values show a stable performance with small variations from the mean and the median that is also very close to the measured monthly mean demonstrating the robustness of the methodology. The results are summarized in Table 2.

Table 2. Main results of the simulation of the year 2014 fifty times in comparison with the measured DNI.

Month	**DNI measured (kWh/m^2)**	**DNI modeled (50 times) (kWh/m^2)**					
		Mean	**Median**	**P25**	**P75**	**Max**	**Min**
1	93	93	93	91	95	98	89
2	83	89	89	88	90	93	86
3	171	181	181	179	183	185	176
4	172	183	183	182	185	189	176
5	270	260	260	256	262	269	253
6	240	245	245	243	247	254	237
7	260	252	252	250	254	259	244
8	282	253	253	251	256	262	238
9	151	156	156	154	159	161	150
10	152	167	167	165	170	173	160
11	96	100	100	99	102	105	93
12	139	126	126	124	127	131	121
Annual	2110	2105	2105	2099	2112	2126	2086

In Figure 5, we represent the hourly ECDFs of the fifty synthetically generated annual data sets together with the ECDF of the measured data set. We find little differences in terms of frequency distribution between all the generated years.

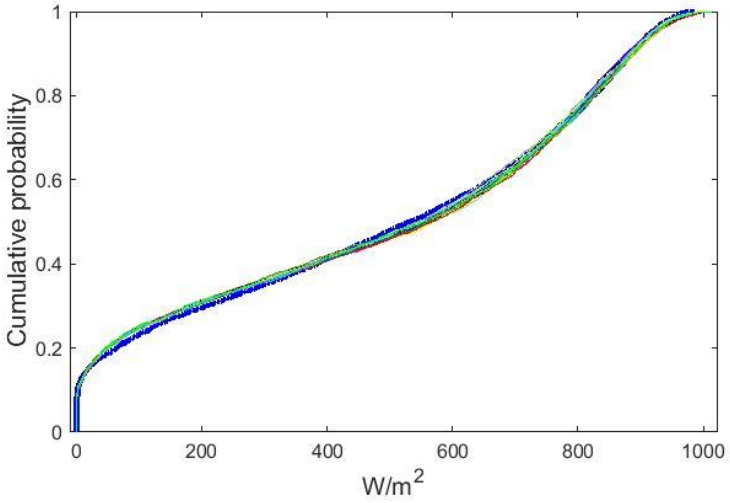

Fig 5. ECDF of the fifty synthetically generated annual data sets compared to the measured ECDF for the year 2014 in Seville (blue thicker line).

For the rest of the results analysis, we have selected one of these fifty years since results barely change in terms of frequency distribution. He have selected one year whose annual cumulative DNI value was the same as the mean of the 50 generated years (2105 kWh/m^2).

Figure 6 presents the modeled and measured k_b^h versus the k_t^h. A similar performance in both figures can be observed.

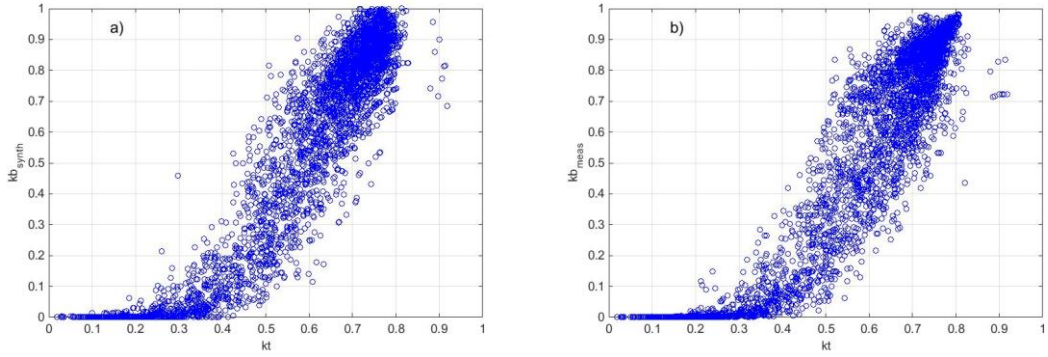

Fig 6. Scatter plots of the hourly direct fraction index versus the hourly clearness index of the modeled data (a) and the hourly direct fraction index versus the hourly clearness index of the measured data (b) for the year 2014 in Seville.

Following the common practices for benchmarking modeled irradiance datasets (Beyer et al. 2008), we use the root mean squared difference (RMSD) as the main statistic for the comparison of the observations and the data generated synthetically. In the analysis, only daylight hours are considered.

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I_{meas}^{i} - I_{synth}^{i}\right)^2} \quad (6)$$

Where, N is the number of data pairs, I_{synth} is the synthetic DNI and I_{meas} is the measured DNI.

The corresponding relative differences are calculated as follows:

$$rRMSD = {RMSD}\Big/{\overline{I_{meas}}} \quad (7)$$

Table 3 presents the RMSD and rRMSD of the, hourly, daily and monthly means of the presented model and the model only with the deterministic component.

Table 3. RMSD and rRMSD of the, hourly, daily and monthly means of the presented model and the model only with the deterministic component for the year 2014 in Seville.

	Deterministic + Stochastic			**Deterministic**		
	Hourly (W/m^2)	Daily (kWh/m^2)	Monthly (kWh/m^2)	Hourly (W/m^2)	Daily (kWh/m^2)	Monthly (kWh/m^2)
RMSD	128	1	13	111	0.8	12
rRMSD (%)	26	18	7	23	13	7

Some examples of the daily profiles are illustrated in Figure 7 where the goodness of the method in reproducing the stochastic variation of the atmosphere composition is qualitatively illustrated. The figures should be observed in pairs. In each pair, we represent days with similar sky condition but different atmospheric composition. The figures on the left represent a clearer atmosphere than figures on the right. The first pair represent partially clear days (a-b), the second pair represent partially cover days (c-d), and the third pair represent clear days (e-f).

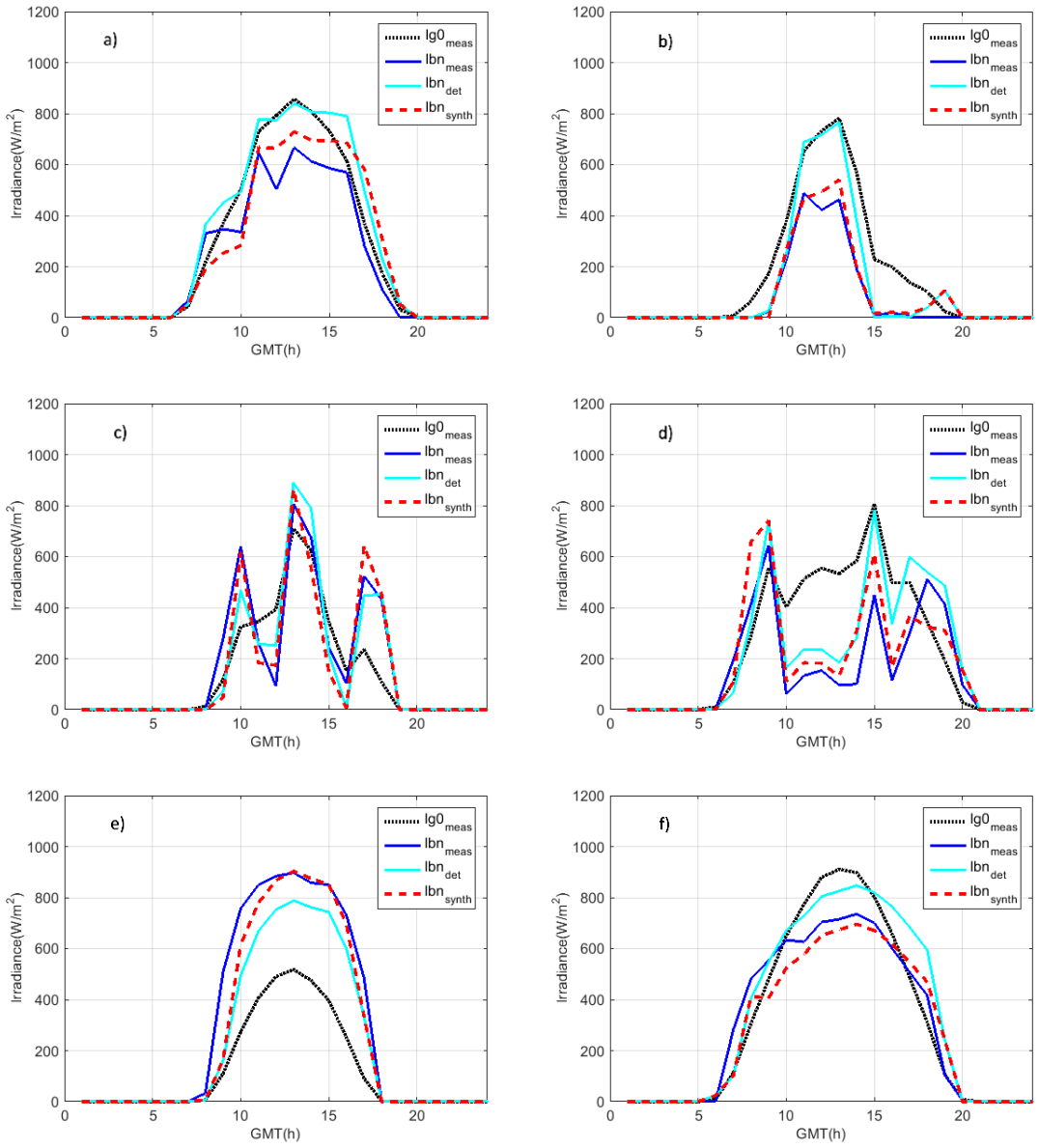

Fig 7. Illustrative examples of the results of the synthetic generation compared to the original dataset.

The assessment of a method for the synthetic generation should aim to find similarities with the frequency distribution of measured data. With this purpose, we have calculated the Finkelstein-Schafer (FS) statistic (Finkelstein and Schafer 1971) for each dataset. This statistic takes into account the differences between the ECDF of the measured and synthetic datasets, and permits the comparison with the results of other models regardless of the time resolution and analysis period.

$$FS = {}^{1}/_{n} \sum_{i=1}^{n} \delta_i \qquad (8)$$

Where, δ is the absolute difference between the measured and synthetic ECDF at each point, I and n represents the number of readings. Only daylight hours are considered.

To assess the similarity of the ECDFs over the whole range of observed values, we also calculate the KSI (Kolmogorov-Smirnov test integral) index which is used to check the similitude between generated and measured DNI values (Espinar et al. 2009) .

$$KSI(\%) = 100 \cdot \frac{\int_{x_{min}}^{x_{max}} D_n dx}{a_{critical}} \quad (9)$$

Where x_{max} and x_{min} are the extreme values of the independent variable, and $a_{critical}$ is calculated as

$$a_{critical} = V_c \cdot (x_{max} - x_{min}) \quad (10)$$

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

$$V_c = {}^{1.63}/_{\sqrt{N}} \quad N \geq 35 \quad (11)$$

In this case $V_c = 0.026$ and D_n are the absolute differences between the ECDFs for each interval. The higher the KSI value, the worse the model fit.

Figure 8 shows the ECDF of the measured and synthetic datasets in addition with the synthetic dataset generated only with the deterministic component (left) and their differences (right), for the different solar elevation intervals considered in the methodology. The blue dotted line represents the critical value.

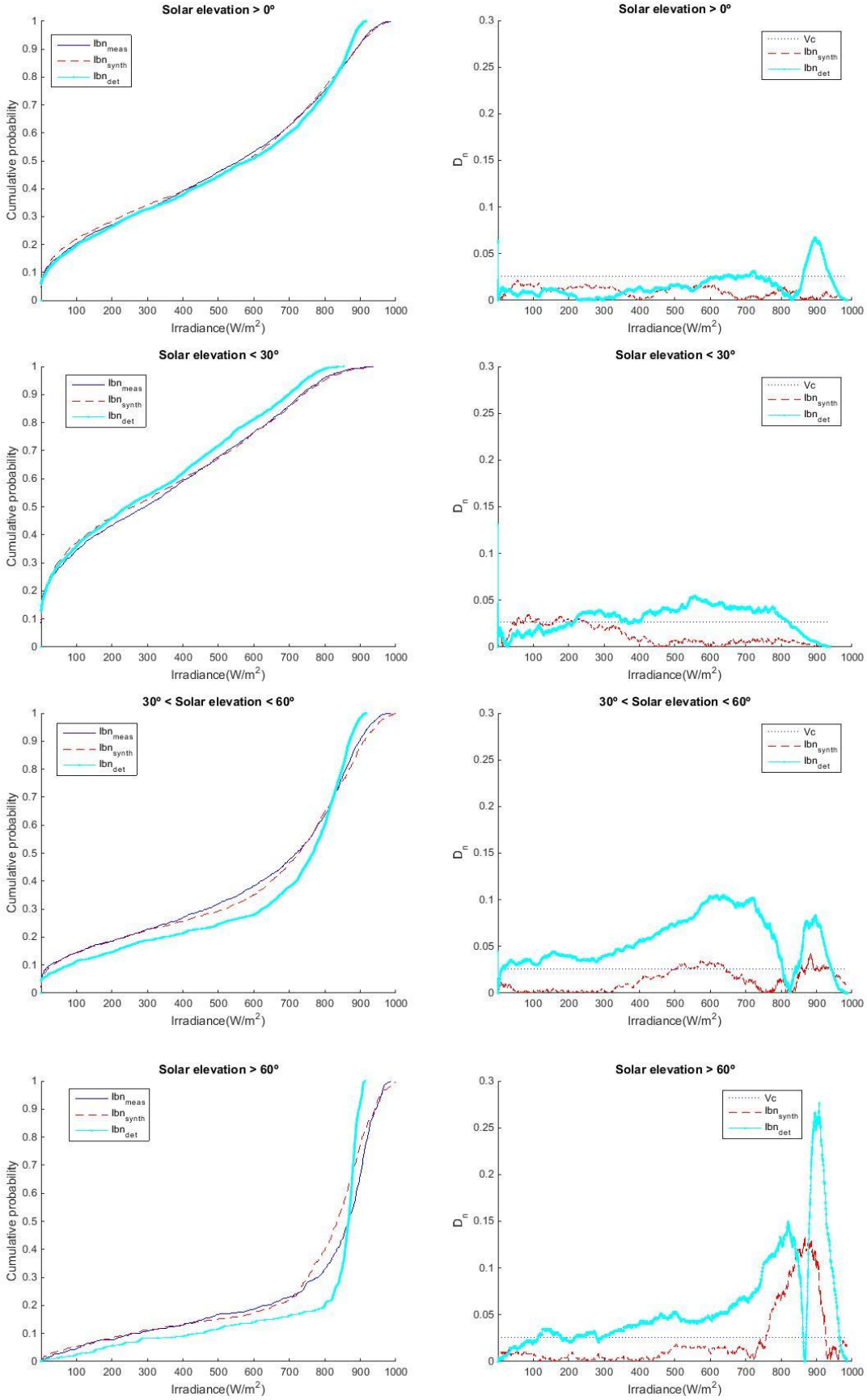

Fig 8. ECDF of the measured and synthetic datasets for each solar elevation selected interval (left) and their differences (right) for the year 2014 at Seville.

The synthetic dataset generated using only the deterministic component exhibits a completely different frequency distribution compared to the frequency distribution of the measured dataset, especially for solar elevations higher than 30° (when the optical efficiency of the solar thermal concentration systems is higher).. The FS and KSI statistics indicate a better performance of the synthetic generation with both, the stochastic and deterministic components for every solar elevation interval.

Table 4 presents the KSI and FS of the synthetic dataset generated with both the deterministic and stochastic component and the synthetic set only generated with the deterministic component for the different solar elevation intervals taken into account. Note that the improvements in the reproduction of the frequency distribution for the three intervals considered are greater than the overall improvement. This is because for solar elevations higher than 30° the ECDF of synthetic data generated only with the deterministic component stands below the ECDF of the measured data (see Fig. 8), while for solar elevations lower than 30° occurs just the opposite. As a result, the distances are balanced and the statistics KSI and FS become smaller in the aggregated values ($\alpha > 0$, Table 4).

Table 4. Performance of the methodology in comparison to the deterministic synthetic generation.

	Deterministic + Stochastic		**Deterministic**	
solar elevation	**KSI (%)**	**FS**	**KSI (%)**	**FS**
$\alpha > 0$	29.2	0.007	56.6	0.015
$0 < \alpha \leq 30$	39.0	0.010	112.8	0.030
$30 < \alpha \leq 60$	46.9	0.012	212.4	0.056
$\alpha > 60$	74.2	0.018	222.7	0.058

The robustness of the methodology can be observed when applying it to other sites with different climatic conditions. We tested it on three locations with Atlantic, Sub-tropical and Continental climates analyzing the hourly FS and RMSD. The results, summarized in Table 5, show a better performance of the proposed methodology compared to the deterministic model performance in terms of frequency distribution in every location and for every solar elevation interval.

Table 5. FS analysis of the synthetic generation in different locations.

Location	*Model*	*Statistic*	*α > 0*	*0 < α ≤ 30*	*30 < α ≤ 60*	*α > 60*
Pamplona	Det. + Stoc.	RMSD (W/m^2)	83	67	87	119
		KSI(%)	26.8	34.4	41.2	29.6
		FS	0.007	0.009	0.011	0.008
	Det.	RMSD (W/m^2)	69	51	65	118
		KSI(%)	60.3	61.7	104.5	69.3
		FS	0.016	0.019	0.028	0.018
Pretoria	Det. + Stoc.	RMSD (W/m^2)	124	129	122	116
		KSI(%)	29.2	42.0	28.2	33.9
		FS	0.007	0.010	0.007	0.008
	Det.	RMSD (W/m^2)	96	110	83	92
		KSI(%)	36.9	75.7	88.2	188
		FS	0.009	0.021	0.022	0.043
Payerne	Det. + Stoc.	RMSD (W/m^2)	136	130	142	147
		KSI(%)	45.2	78.7	85.2	85.9
		FS	0.016	0.024	0.022	0.023
	Det.	RMSD (W/m^2)	113	108	118	126
		KSI(%)	51.7	157.1	229	262.1
		FS	0.018	0.049	0.059	0.070

5 Conclusions

In this paper, we present a novel and simple methodology for the generation of synthetic hourly DNI data taking hourly mean GHI values as input. The goal of this methodology is to improve the results in terms of frequency distribution of the synthetic values, which is relevant for the performance of concentrating solar energy technologies, especially CSP. The proposed method upgrades the traditional models by dividing the solar radiation into a deterministic and a stochastic component. While there is no improvement in terms of RMSE, the results in reproducing the solar radiation dynamics quantified in terms of frequency distribution are outstanding. The FS and KSI statistics exhibits a reduction of 50% for the year 2014 at the location of Seville and reaches a reduction of 70% for solar elevations higher than 30°. In addition, the proposed methodology has been applied to other locations with diverse climatic conditions, achieving again excellent results compared to the deterministic model. The reduction of the KSI and FS indicators are greater than 50% for every solar elevation interval analyzed at all the locations, demonstrating the robustness of the proposed methodology.

Acknowledgments

The authors are grateful to C. Fernandez-Peruchena and L. Vuilleumier for providing the Pamplona and Payerne databases, respectively and F. Dinter for his advice in the use of SAURAN database used to access the Pretoria data.

This article has been developed under the frame of the project Codisol (cod. DPI2013-44135-R) of the RETOS Programme 2013 funded by the Ministry of Economy and Competitiveness of the Government of Spain

References

Behar, O., Khellaf, A., Mohammedi, K., 2015. Comparison of solar radiation models and their validation under Algerian climate - The case of direct irradiance. Energy Conversion and Management, 98, pp.236–251.

Beyer, H.G., Polo, J., Suri, M., Torres, J.L., Lorenz, E., Hoyer-Klick, C., Ineichen, P., 2008. D 1.1.3. Report on benchmarking of radiation products. Sixth framework programme contract no (038665).

Boland, J., Huang, J. and Ridley, B., 2013. Decomposing global solar radiation into its direct and diffuse components. Renewable and Sustainable Energy Reviews, 28, pp.749–756.

Boland, J., Scott, L. and Luther, M., 2001. Modelling the diffuse fraction of global solar radiation on a horizontal surface. Environmetrics, 12, pp.103–116.

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. Journal of Energy in Southern Africa, 26, pp.2–10.

Chhatbar, K., Meyer, R., 2011., The influence of meteorological parameters on the energy yield of solar thermal power plants. In: SolarPACES Proceedings, Granada, Spain.

Das, S., Dey, S., Dash, S.K. 2016., Theor Appl Climatol 124: 629. doi:10.1007/s00704-015-1444-8

Dugaria, S., Padovan, A., Sabatelli, V., Del Col, D. 2015., Assessment of estimation methods of DNI resource in solar concentrating systems. Solar Energy, 121, pp.103–115.

Erbs, D.G., Klein, S.A., Duffie, J.A., 1982. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Solar Energy, 28(4), pp.293–302.

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, pp.118–125.

Finkelstein, J.M., Schafer, R.E., 1971. Improved goodness-of-fit tests. Biometrika, 58, pp.641–645.

Furlan, C., de Oliveira, A.P., Soares, J., Codato, G., Escobedo, J.F., 2012. The role of clouds in improving the regression model for hourly values of diffuse solar radiation. Applied Energy, 92, pp.240–254.

Gueymard, C.A., Ruiz-arias, J.A., 2014. Performance of Separation Models to Predict Direct Irradiance at High Frequency : Validation over Arid Areas. International Solar Energy Conference, pp.16–19.

Jacovides, C.P., Tymvios, F. S., Assimakopoulos, V. D., Kaltsounides, N.A., 2006. Comparative study of various correlations in estimating hourly diffuse fraction of global solar radiation. Renewable Energy, 31(15), pp.2492–2504.

Kasten, F., Young, A.T., 1989. Revised optical air mass tables and approximation formula. Applied Optics, 18(22), pp.4735–4738.

Liu, B.Y.H., Jordan, R.C., 1960. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. Solar Energy, 4(3), pp.1–19.

Louche, A., Notton, G., Poggi, P., Simonnot, G., 1991. Correlations for direct normal and global horizontal irradiation on a French Mediterranean site. Solar Energy, 46(4), pp.261–266.

Maxwell, E., 1987. A quasi-physical model for converting hourly global horizontal to direct normal insolation, Golden, Colorado. Available at: http://adsabs.harvard.edu/abs/1987qpmc.rept.....M.

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

Moreno-Tejera, S., Ramírez-Santigosa, L. and Silva-Pérez, M.A., 2015. A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data. Renewable Energy, 78, pp.531–537.

Oliveira, A.P., Escobedo, J. F., Machado, A. J., Soares, J., 2002. Correlation models of diffuse solar-radiation applied to the city of São Paulo, Brazil. Applied Energy, 71(1), pp.59–73.

Orgill, J.F., Hollands, K.G.T., 1977. Correlation equation for hourly diffuse radiation on a horizontal surface. Solar Energy, 19(4), pp.357–359.

Paulescu, E., Blaga, R., 2016. Regression models for hourly diffuse solar radiation. Solar Energy, 125,

pp.111–124.

Perez, R., Seals, R., Zelenka, A., Ineichen, P., 1990. Climatic evaluation of models that predict hourly direct irradiance from hourly global irradiance: Prospects for performance improvements. Solar Energy, 44(2), pp.99–108.

Perez, R., Ineichen, P., 1992. Dynamic Global-To-Direct Irradiance Conversion Model (Ashrae). ASHRAE Trans Res, 98, pp.354–69.

Reindl, D.T., Beckman, W.A., Duffie, J.A., 1990. Diffuse fraction correlations. Solar Energy, 45(1), pp.1–7.

Ridley, B., Boland, J., Lauret, P., 2010. Modelling of diffuse solar fraction with multiple predictors. Renewable Energy, 35(2), pp.478–483.

Roesch, A., Wild, M., Ohmura, A., Dutton, E. G., Long, C. N., Zhang, T., 2011. Assessment of BSRN radiation records for the computation of monthly means. Atmospheric Measurement Techniques, 4(2), pp.339–354.

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

Silva-Perez, M.A. Barea-García, J.M., Larranñeta, 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, pp.2512–2520.

Singh, C., Thomas, L. & Kumar, K.K., 2017. mpact of aerosols and cloud parameters on Indian summer monsoon rain at intraseasonal scale: a diagnostic study. Theor Appl Climatol (2017) 127: 381. doi:10.1007/s00704-015-1640-6

Skartveit, A., Olseth, J.A., 1992. The probability density and autocorrelation of short-term global and beam irradiance. Solar Energy, 49(6), pp.477–487.

Skartveit, A., Olseth, J.A., Tuft, M.E., 1998. An hourly diffuse fraction model with correction for variability and surface albedo. Solar Energy, 63(3), pp.173–183.

Tapakis, R., Charalambides, a. G., 2014. Enhanced values of global irradiance due to the presence of clouds in Eastern Mediterranean. Renewable Energy, 62, pp.459–467.

Torres, J.L., De Blas, M., García, A., de Francisco, A., 2010. Comparative study of various models in estimating hourly diffuse solar irradiance. Renewable Energy, 35(6), pp.1325–1332.

Tovar-Pescador, J., 2008. Modelling the Statistical Properties of Solar Radiation and Proposal of Technique Based on Boltzmann Statistics. In Modeling Solar Radiation at the Earth Surface. Springer, pp. 55–91.

Vuilleumier, L. Hauser, M., Félix, C., Vignola, F., Blanc, P., Kazantzidis, A., Calpini, B., 2014. Accuracy of ground surface broadband shortwave radiation monitoring. Journal of Geophsical Research: Atmospheres, 119, pp.1365–1382.