Tesis Doctoral
Ingeniería Energética

Synthetic generation of high-temporal resolution direct normal irradiation time series

Autor: Miguel Larrañeta Gómez-Caminero
Director: Dr. Manuel A. Silva Pérez

Dep. Ingeniería Energética
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

2018

Ingeniería Energética

Synthetic generation of high-temporal resolution direct normal irradiation time series

Autor:

Miguel Larrañeta Gómez-Caminero

Director:

Dr. Manuel A. Silva Pérez

Dep. de Ingeniería Energética

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2018

Agradecimientos

A mi mujer, Belencita, por no leerse una línea de las aquí escritas y a mis niñas Ana y Lola que ahora pensaran que tienen al médico en casa.

A mis hermanos, Carlos, Barbara y Juancho, me encantan nuestros bucles infinitos. A mis compañeros del GTER, los que están y las cabezas que no están, juntos hemos aprendido a aprender disfrutando, son como mi segunda familia.

A los integrantes del departamento, especialmente a los directores del grupo de investigación Isidoro Lillo y Manuel Silva, la combinación perfecta de determinación y rigor.

Dedicar un par de líneas a mi director, mentor y padre académico, Manolo, que suerte la mía haberme cruzado con él. Bajo su dirección he aprendido a alcanzar mis metas sin necesidad de aspavientos, un ejemplo de vida en lo personal y lo profesional.

Y por último, a mi amada madre. Madre no hay más que una y la mía es la mejor del mundo.

Miguel Larrañeta Gómez-Caminero

Sevilla, 2018

A mi padre

Lista original de publicaciones

Esta Tesis doctoral se presenta por compendio de publicaciones de acuerdo con el artículo 9 de la normativa reguladora del régimen de tesis doctoral (Acuerdo 9.1/CG 19-4-12) en el marco del programa regulado por el RD 1393/2007. Los artículos publicados relacionados con el hilo argumental de la tesis son los siguientes (anexos A, B, C, D, E):

ANEXO A:

Larrañeta, M., Moreno-Tejera, S., Silva-Pérez, M.A., Lillo-Bravo, I., 2015. **An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series.** Sol. Energy 122, 517–528. Publicado antes de la fecha de la primera matrícula en el programa de doctorado.

ANEXO B:

Larrañeta, M., M.J. Reno., Silva-Pérez, M.A., Lillo-Bravo, I. 2017. **Identifying periods of clear sky direct normal irradiance**. Renewable Energy. 113:756–763.

ANEXO C:

Larrañeta, M., Moreno-Tejera, S., Lillo-Bravo, I., Silva-Pérez, M.A. 2018. **A methodology for the stochastic generation of hourly synthetic direct normal irradiation time series**. Theoretical and Applied Climatology 131, 3-4 1517-1527.

ANEXO D:

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

ANEXO E:

Larrañeta, M., Moreno-Tejera, S., Lillo-Bravo, I., Silva-Pérez, M.A., 2017. **Cloud transient characterization in different time steps**. AIP Conf. Proc. 1850. doi:10.1063/1.4984524. SolarPACES 2016 conference proceedings.

Administrativamente la tesis la componen 3 de estos 5 artículos, anexos B, C y D, publicados en revistas indexadas en plazo y forma.

Otros trabajos publicados relacionados con la temática de la tesis:

- **Publicaciones en revistas indexadas:**

Vera-Medina, J.., Hernández, J., Lillo-Bravo I. Larrañeta M., 2018. Experimental and numerical study on a freeze protection system for flat-plate solar collectors with silicone peroxide tubes. Applied Thermal Engineering. In Press: https://doi.org/10.1016/j.applthermaleng.2018.02.085.

Peruchena C.F., Larrañeta M., Blanco M., Bernardos A. 2018. High frequency generation of coupled GHI and DNI based on clustered Dynamic Paths. Solar Energy 159, 453-457.

Isidoro Lillo-Bravo, Pablo González-Martínez, Miguel Larrañeta and José Guasumba-Codena. 2018. Impact of Energy Losses Due to Failures on Photovoltaic Plant Energy Balance. Energies 11(2):363.

Silva-Perez, M.A., Barea-Garcia, J.M., Larrañeta, M., Moreno-Tejera, S., Lillo-Bravo, 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. doi:10.1016/j.egypro.2014.03.266

- **Publicaciones en actas de congresos:**

Lillo-Bravo, Isidoro; Silva-Pérez, Manuel Antonio; Larrañeta-Gómez-Caminero, Miguel. Experimental Comparison of Different Global Irradiation Models with and without Sun Tracking for the South of Spain. 28th EU PVSEC, 2013, Wip ISSN 2196-0992.

Silva-Pérez, Manuel Antonio; Barea-García, Jose María; Sancho-Caparrini, Natividad; Briceño-Cano, Susana; Larrañeta-Gómez-Caminero, Miguel. Analysis of the Electricity Generation of a Parabolic Trough Power Plant from the Simulation of Eleven Years of Direct Normal Radiation Measurements. Cost Action WIRE ES1002 Meeting. 2012.Roskilide, Denmark

Moreno-Tejera, Sara; Pérez-Aparicio, Elena; Larrañeta-Gómez-Caminero, Miguel; Silva-Pérez, Manuel Antonio. Influence of data type in the development of TMY. SolarPACES 2012 Granada, Spain.

Lillo-Bravo, Isidoro; Silva-Pérez, Manuel Antonio; Larrañeta-Gómez-Caminero, Miguel. Use of Phase Change Materials in Photovoltaic Modules with Solar Concentration Up to 2X. 26h EU PVSEC, 2011. Germany

Moreno, S., Pérez, E., Larrañeta, M., Silva, M. a., 2011. Influence of data type in the development of TMY. SolarPaces Conf. Granada, Spain.

Abstract

In this doctoral thesis, we present two methods for the synthetic generation of high temporal resolution Direct Normal Irradiance (DNI) time series from hourly means that can be applied globally without any local adaptation.

- The SA (Stochastic Adaptation) method, based on the concept proposed by Polo et al, (2011) with several improvements. It divides the irradiance into a deterministic and a stochastic component, i.e., the contribution from the hourly mean and the stochastic fluctuation which depends on the sky condition, respectively. The improvements are focused on the characterization and reproduction of the cloud transients' effects depending on the sky condition, and the identification of periods where fluctuations of the solar radiation take place. We have incorporated these improvements in two approaches, the initial approach (Larrañeta et al., 2015) and the second approach (Larrañeta et al., 2018-b).
- The ND (Non-Dimensional) method relies on the normalization of the daily profiles of an extensive database of solar radiation data by a clear-sky envelope that can be used to reproduce the high-frequency dynamics of DNI in any location. The method is based on the concept proposed by Fernández-Peruchena et al, (2015) but with several improvements, the most relevant being the identification of the most similar day to be reproduced not only in terms of energy, but also in terms of distribution and variability of the daily DNI curves.

It is worth highlighting that the methods presented in this work use measurements from one location to characterize the cloud transients and generate synthetic high resolution DNI data in any location where hourly DNI data is available, without any local adaptation. Each method requires different degrees of accuracy in the knowledge of local hourly DNI data. The SA method requires high-quality site hourly DNI series while the ND 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.

The models have been applied in three locations with different climatic conditions. In the analysis of the performance of each model, the measured and synthetic time series have been evaluated in terms of energy, frequency distribution and autocorrelation. In terms of energy, we obtain annual differences lower than 0.2%. The similitude between measured and generated DNI distributions has been evaluated through the Kolmogorov-Smirnov test Integral (*KSI*), and the performance of the synthetic series when used for the evaluation of the thermal power produced by a parabolic trough (PT) plant has been assessed using the daily normalized root mean square deviations (NRMSD). The autocorrelation have been addressed by evaluating the ramp rates of both measured and synthetic data sets, The generation methods provide, 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. For ramp rates greater than 500 W/m^2·min, the differences in the measured and synthetic sets are almost negligible. Sites selected for validating these methods are located at different climates and latitudes, suggesting their global applicability.

Resumen

En esta tesis doctoral, se presentan dos métodos para la generación sintética de datos de Irradiancia Directa Normal (DNI) de alta resolución temporal a partir de promedios horarios que se pueden aplicar en cualquier emplazamiento sin necesidad de ser adaptados localmente.

- El método SA (Adaptación estocástica). Se basa en el concepto planteado por Polo et al., (2011) pero con varias mejoras. Conceptualmente, la irradiancia se divide en una componente determinista y otra estocástica, esto es, la aportación del promedio horario y la fluctuación de la irradiancia en torno a dicho promedio, condicionada por el estado del cielo. Las mejoras se centran en la caracterización y reproducción de los transitorios de nubosidad en función de las condiciones del cielo y la identificación de los períodos en los que se producen las fluctuaciones de la radiación solar. Dichas mejoras se han abordado en dos enfoques, uno inicial (Larrañeta et al., 2015) y otro más reciente (Larrañeta et al., 2018-b).
- El método ND (no-dimensional) se basa en la normalización de los perfiles diarios de radiación de una extensa base de datos de radiación solar en alta resolución temporal con respecto a una envolvente de cielo despejado que posteriormente se puede utilizar para reproducir las dinámicas de la DNI en cualquier emplazamiento. El método se basa en el concepto planteado por Fernández-Peruchena et al, (2015) pero introduce varias mejoras, la más relevante de las cuales es la introducción de un nuevo método para la identificación del día más similar a ser reproducido no solo en términos de energía, sino también en términos de distribución y variabilidad diaria de la irradiancia directa normal.

Los modelos presentados en esta tesis doctoral solo requieren medidas de DNI de alta resolución temporal en un único emplazamiento y a partir de ellas se pueden simular los transitorios de nubosidad y generar datos sintéticos de DNI de baja frecuencia en cualquier emplazamiento en el que se dispongan de datos horarios de

irradiancia directa normal. La información disponible respecto a la precisión de los datos horarios de DNI conducirá al uso de uno u otro modelo. El modelo SA requiere una serie de DNI horaria en el emplazamiento de alta calidad, mientras que el método ND solo requiere información de los perfiles diarios de DNI en términos de energía, variabilidad y distribución.

Los modelos se han aplicado en tres ubicaciones con diferentes condiciones climáticas. En el análisis del comportamiento de cada modelo se han evaluado las similitudes en términos de energía, distribución de frecuencia y autocorrelación entre las series medidas y sintéticas. En términos energéticos, se obtienen diferencias anuales inferiores al 0.2%. La similitud entre las distribuciones DNI medidas y generadas se ha evaluado a través del test integral Kolmogorov-Smirnov (KSI), y para la evaluación de su impacto en la estimación de la potencia térmica producida por una central termosolar de canal parabólico se ha utilizado la desviación cuadrática media (NRMSD). La autocorrelación se ha evaluado a partir del análisis de las rampas de las series de irradiancia. Los métodos de generación proporcionan, para un conjunto anual de datos sintéticos en paso minutal, valores de KSI entre ~ 3.3 W/m² y ~ 12.9 W/m² (dependiendo del método de generación utilizado), y NRMSD diario de ~ 0.9 % y ~ 3.4 %, respectivamente. Para rampas superiores a 500 W/(m²·min) las diferencias observadas son prácticamente despreciables. Los emplazamientos seleccionados para validar estos métodos se encuentran en climas y latitudes diferentes, lo que sugiere su aplicabilidad global.

Document structure

This document is structured according to the recommendations established in article 9 of the regulations governing the doctoral thesis regime (Agreement 9.1 / CG 19-4-12) for the modality of Thesis by compendium of publications. The purpose of this document is to summarize and integrate the work done in the annexed publications that make up the thesis. Given the international nature of the publications, it has been decided to use the English language in the document. The sections referred are: Introduction, Motivation and Objective, Methodology, Results and Discussion and Conclusions.

According to this structure, the section Introduction, Motivation and Objective, presents the needs of the industry of the solar thermal sector that motivate the development of the present thesis, as well as the approaches to the synthetic generation high temporal resolution solar radiation time series found in the literature.

The next section, Methodology, is the most extensive of the document. It includes 3 subsections. We initially describe the databases used in the training and application of the models. In the second sub-section we describe the stochastic adaptation method (Polo et al., 2011) and the two approaches that have been applied to improve this algorithm, the initial one (Larrañeta et al., 2015, Annex A) and the latest and improved approach (Larrañeta et al., 2018-b Annex D). For the development of the improvements related to the last approach of the stochastic adaptation method, we have carried out several investigations that have derived in additional publications. These are (Larrañeta et al., 2017-a) and (Larrañeta et al., 2017-b), Annexes B and E respectively. In addition, the concept on which the stochastic adaptation model is based, can be applied in other areas of the synthetic generation of solar radiation time series (Larrañeta et al., 2018-a), Annex C, as exposed at the end of the sub-section.

The third sub-section of the methodology describes the method and the improvements implemented in the non-dimensional model in its only approach to date (Larrañeta et al., 2018-b), Annex D.

The Results and Discussion section presents and discusses the results of applying both models in their latest approaches to several sites with different climatic conditions, comparing the performance of each model in terms of energy, frequency distribution and autocorrelation.

The Conclusions section, where we summarize the work done and the main results obtained and propose future lines of work related to the improvement of the models closes the document.

Estructura del documento

El presente documento se estructura de acuerdo a las recomendaciones establecidas en el artículo 9 de la normativa reguladora del régimen de tesis doctoral (Acuerdo 9.1/CG 19-4-12) dentro de la modalidad Tesis por compendio de publicaciones. El objeto de este documento es resumir e integrar el trabajo realizado en las publicaciones anexas que conforman la tesis. Dada la naturaleza internacional de las publicaciones, se ha decidido utilizar el idioma inglés en el documento resumen. Los apartados contemplados son: Introduction, Motivation and Objective, Methodology, Results and Discussion and Conclusions.

Conforme a dicha estructura, en el apartado Introduction, Motivation and Objective, se presentan las necesidades de la industria del sector de la energía termosolar que motivan el desarrollo de la presente tesis, así como las aproximaciones existentes en la literatura a la generación sintética de series de irradiancia de alta resolución temporal.

El siguiente apartado, Methodology, es el más extenso del documento. En él se contemplan 3 subapartados. Inicialmente se detallan en las bases de datos utilizadas tanto en el entrenamiento como en la aplicación de los modelos. Posteriormente se profundiza en las diferentes aproximaciones a las mejoras de cada uno de los dos modelos implementados. En el segundo subapartado se describe el método de adaptación estocástica (Polo et al., 2011) y los dos enfoques que se han aplicado para la mejora de dicho algoritmo, uno inicial (Larrañeta et al., 2015 Anexo A), y otro más reciente y que mejora la aproximación inicial (Larrañeta et al., 2018-b, Anexo D). Para el desarrollo de las mejoras relacionadas con el último enfoque del método de adaptación estocástica, se han realizado investigaciones que han derivado en otras publicaciones contempladas en esta tesis. Estas son (Larrañeta et al., 2017-a) y (Larrañeta et al., 2017-b), Anexos B y E respectivamente. Además, el concepto en el que se basa dicho modelo se puede aplicar en otros ámbitos de la generación sintética de series de irradiancia (Larrañeta et al., 2018-a), Anexo C, como se expone al final del sub-apartado.

El tercer sub-apartado de la metodología describe el método y las mejoras implementadas en el modelo no-dimensional en su única aproximación hasta la fecha (Larrañeta et al., 2018-b), Anexo D.

El apartado Results and Discussion presenta y discute los resultados de aplicar ambos modelos en su última aproximación a varios emplazamientos con distintas condiciones climáticas comparando el funcionamiento de cada modelo en la generación sintética de series de irradiancia directa normal en términos de promedio energético, distribución de frecuencia y autorregresión.

El documento se cierra en el apartado Conclusions, en el que además de resumir el trabajo realizado y los principales resultados obtenidos, se expresan las futuras líneas de trabajo relacionadas con la mejora de los modelos.

Table of Contents

List of figures

List of Tables

Glossary

$I_{bn}{}^{i}$	Direct normal irradiance in the instant *i*
$I_{bn}{}^{h}$	Hourly direct normal irradiance
$I_{bn}{}^{d}$	Daily direct normal irradiance
$I_{bn}{}^{d}_{md}$	Direct normal irradiance recorded from the sunshine to the solar noon
$I_{bn}{}^{i}_{synth}$	Synthetic direct normal irradiance in the instant *i*
$I_{bn}{}^{i}_{i3}$	Cubic interpolation of the hourly values in the high resolution time scale
$I_{bn}{}^{h}_{cs}$	Hourly clear-sky direct normal irradiance
$I_{bn}{}^{d}_{cs}$	Daily clear-sky direct normal irradiance
$k_b{}^{h}$	Hourly direct fraction index
$k_b{}^{d}$	Daily direct fraction index
I_{cs}	Solar constant
E_0	Correction due to Earth-Sun distance
m_R	Relative air mass
s	Silhouette index
dif_{1-min}	Differences between the 1-min DNI data to the cubic interpolation of the hourly values in the 1-min scale

σ_{1-min}^{h}	Standard deviation of the differences between the 1-min DNI data to the cubic interpolation of the hourly values in the 1-min scale
$\max_{1-min}^{h}$	Maximum value of the differences between the 1-min DNI data to the cubic interpolation of the hourly values in the 1-min scale
τ	Tortuosity
${k_t}^i$	Clearness index in the instant *i*
${I_{g0}}^i$	Global horizontal irradiance in the instant *i*
${I_0}^i$	Extra-terrestrial irradiance in the instant *i*
${k_t}^h$	Hourly clearness index
P2	Perturbation coefficient
${D_{cs-m}}^i$	Absolute percentage differences of the measured and clear sky radiation
${S_{cs}}^i$	Slope of the straight lines that join two hourly direct normal clear sky irradiance values
${S_m}^i$	Slope of the straight lines that join two hourly direct normal measured irradiance values
${L_{cs}}^i$	Length of the straight lines that join two hourly direct normal clear sky irradiance values
${L_m}^i$	Length of the straight lines that join two hourly direct normal measured irradiance values
R	Random number calculated from an uniform distribution between 0 and 1
VI	Variability index
F_m	Morning fraction index
KSI	Kolmogorov-Smirnov test integral index

$a_{critical}$	Critical area
V_c	Critical value
$NRMSD$	Normalized root mean squared deviation
$RMSD$	Root mean squared deviation
P_{max}	Maximum thermal power production
P_{min}	Minimum thermal power production
P_m^i	Thermal power produced when using the measured DNI as input for the simulation in SAM
P_s^i	Thermal power produced when using the synthetic DNI as input for the simulation in SAM
RR	Ramp rate

1. INTRODUCTION, MOTIVATION AND OBJECTIVES

The direct normal irradiance time series are one of the basic inputs for the simulation of Concentrating Solar Thermal (CST) systems and 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. However, the availability of high resolution DNI data is often limited in terms of extension for most of locations, and typically historical solar resource data are available at hourly scale (Fernández-Peruchena et al., 2010).

On the other hand, STE plant operators frequently use irradiance predictions based on satellite estimates or meteorological, solar radiation models or a combination of both 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). Nowadays, 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 do not maintain the frequency distribution of the ground measured data (Hammer et al., 2009).

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 earlier attempt conducted by Skartveit and Olseth (1992) on the synthetic generation of irradiance values at different time intervals 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 form the bases for a scheme of data synthetization.

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.

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 used as a driver for the generation of an on/off sequence of beam irradiance (Morf, 2011). 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. This model was improved in two attempts by Larrañeta et al. (2015) and Larrañeta et al. (2018-b) for a more accurate DNI generation. Grantham et al. (2017) modified the model 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. It improved the first-order MTM model employed by Richardson and Thomas (2011) by including statistical characteristics associated with the atmospheric condition (clear, cloudy, and overcast). 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., 2013), 3-h (Fernández-Peruchena et al., 2017) and hourly (Fernández-Peruchena et al., 2015) DNI means. 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 at the location under study. 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, it required, besides a high frequency database on the site, a great computational cost. For the generation of high frequency DNI time series on a given day it sought for the most appropriate combination of clear sky envelope and dimensionless daily curves within all the possibilities in terms of the Euclidean distance between measured and synthetic values. The same concept was used by Fernández-Peruchena et al. (2016) to generate synthetic 1-min GHI from hourly means and coupled DNI and GHI 1-min time series (Fernández-Peruchena et al., 2018). The model was also improved by Larrañeta et al, (2018-b) by modifying the clear sky envelope approximation and by selecting the most similar day to be downscaled from the normalized dataset in terms of energy, variability and distribution.

The use of machine learning techniques such as artificial neural networks (ANN) for the generation of solar radiation values is mainly focused on the forecast applications of GHI (Mellit et al., 2005; Linares-Rodriguez et al., 2011). However, they open promising possibilities in the field of the synthetic generation of high temporal resolution solar radiation time series.

In this doctoral thesis, 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

2. DATA & METHODOLOGY

In the next subsections, we describe the two methodologies implemented for the synthetic generation of high frequency synthetic DNI data from hourly means and the database used for the training and testing of the methods.

2.1. Meteorological database

In this work, an extensive database is used for training the proposed methods (Table 2-1). This database is composed of 1-min average values of DNI recorded during 14 consecutive years (2002–2015) and two years of 10-min DNI averages (2000-2001) 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 Kipp & Zonen 2AP sun tracker 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 (Moreno et al., 2016).

Table 2–1. Location selected for training the methods.

	Latitude (°N)	Longitude (°W)	Altitude (m)	Climate	Period
Seville	37.4	6.0	12	Mediterranean	2000-2015

In addition, the models have been validated in three other locations belonging to different climates and latitudes (Table 2-2). 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, have been provided by the Baseline Surface Radiation Network (BSRN) (Ohmura et al., 1998); DNI data from Pretoria have 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

belong 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 2–2. Locations selected for the application of 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

2.2. Stochastic adaptation (SA) model

The SA model follows the methodology proposed by Polo et al. (2011) for the generation of 10-min synthetic irradiance values from a given hourly time set. The original method is briefly described in the following lines.

In the first step, the 10-min data available were clustered as a function of the atmospheric conditions since the dynamics of the DNI vary considerably depending on these atmospheric conditions (clouds, aerosols, etc.). The clustering was performed by calculating the normalized clearness index, kt' (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 was generated by the cubic interpolation of the hourly means calculated every 4 hours in the 10-min time scale. The stochastic component was 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 procedure is roughly described below:

i. Calculation of the cubic interpolation of the hourly values in the high resolution time scale to generate the shape where the fluctuations will be added (Ibn_{i3}^{i}).
ii. Generation of random numbers from a 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, the estimation of the synthetic data for the instant i (Ibn_{synth}^{i}) is performed by means of the following equation:

$$Ibn_{synth}^{i} = Ibn_{i3}^{i} + sign(r) \cdot A \qquad \text{Ec. (2–1)}$$

Where, the subscript $i3$ represents the cubic interpolated value, r is the random number from the beta distribution, and A is the amplitude of the fluctuation.

We have improved the original method in two approaches, the initial approach (Larrañeta et al., 2015) and the last approach (Larrañeta et al., 2018-b). The improvements are focused in the stochastic component; no modifications have been included for the deterministic component. They can be divided in three main blocks.

- The sky condition classification. This improvement is focused on the clustering of the available high resolution database into groups that represent similar sky condition leading to similar cloud transients to be reproduced in a given moment. This way, the stochastic component, related to the fluctuations of the solar radiation would reproduce more accurately real situations.

- Clear sky equivalent DNI condition identification. In many periods when there are not visible clouds between the observer and the sun (although there may be clouds in the rest of the sky dome), there are no cloud transients and the synthetic data will be equivalent to the deterministic component.

$$Ibn^{i}_{synth} = Ibn^{i}_{i3}$$ Ec. (2–2)

This improvement is focused on the identification of these clear sky equivalent DNI periods.

- Stochastic component reproduction. This improvement is focused on the way the synthetic time series are dynamically reproduced since it has a significant impact in the accuracy of the algorithms.

- Similarity of the daily sums. This improvement is focused on generating synthetic data sets with similar daily energy as the input data to be downscaled.

In the initial approach, we aimed to generate synthetic data at 10-min resolution, following Polo et al, (2011) practice, but recent publications suggested a greater resolution up to 1-min to 5-min improves the detailed performance modelling of CST systems (Ramirez et al., 2017). Hence in the second approach we aim to generate synthetic DNI time series in the 1-min resolution.

2.2.1. The sky condition classification

i Initial approach

Polo et al, (2011) used *kt'* for the classification of the sky condition. The cloud transients 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. *kt'* is based on GHI and does not properly define the sky conditions for this application. We propose the use of the direct fraction index, k_b, (Skartveit and Olseth, 1992) divided in intervals of 0.1 as described in Table 2-3 instead of the original division into four intervals of the normalized clearness index.

$$k_b{}^{h} = I_{bn}{}^{h} / I_{bn_{cs}}{}^{h}$$ Ec. (2–3)

Where, Ibn is the observed direct normal irradiance and $I_{bn_{cs}}$is the clear-sky DNI. We use the clear sky model A-B Proposed by Silva-Perez (2002).

$$I_{bn_{cs}}{}^{h} = I_{cs} \cdot E_0 \cdot \frac{A}{1+B \cdot m_R}$$ Ec. (2–4)

Where m_R is the relative air mass determined according to the expression of Kasten

and Young, (1989) Ics is the solar constant, E_0 the correction due to Earth-Sun distance and *A* and *B* are empirical parameters intended to model the state of transparency or turbidity of the atmosphere. This model is a modification of the Kastov´s formula quoted, among others, by Kondratyev (1969):

$$I_{bn_{cs}}{}^{h} = I_{cs} \cdot E_0 \cdot \frac{1}{1+c \cdot m_R} \qquad \text{Ec. (2–5)}$$

The introduction of the second parameter is justified by the fact that, as noted by Murk (1959), at least two parameters are required to model the time evolution of the solar irradiance. The parameter *A* mainly accounts for the absorption in certain spectral bands, particularly those in which the absorption is stronger, while *B* realizes primarily scattering phenomena, but also of weak absorption phenomena. In any case, A and B are not independent considering that certain elements (including aerosols) play an important role in both dispersion and absorption processes; therefore, their presence in the atmosphere will be reflected in both parameters.

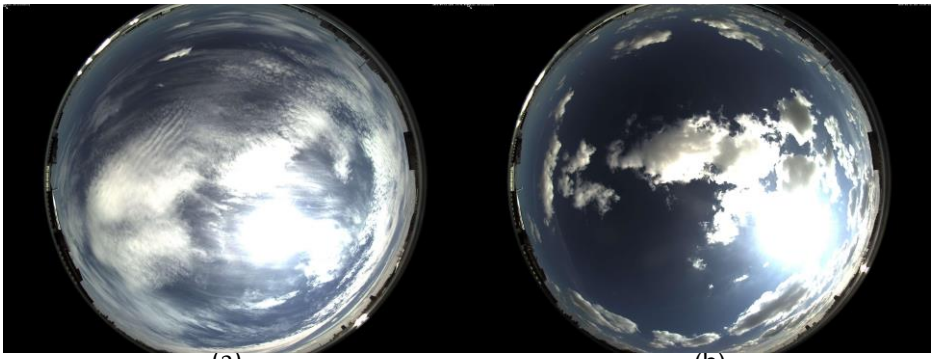
Water vapour and atmospheric aerosols are predominant components in the absorption processes. Regarding dispersion processes, aerosols play a major role. Furthermore, the presence of these two components in the atmosphere is quite variable. According to the above reasoning, the presence of water vapour will be reflected mainly in the parameter *A*, which will be lower the greater the presence of water vapour in the atmosphere, and the presence of aerosols is mainly reflected in the *B* parameter, whose value will be greater the higher content of aerosols.

Table 2–3. Sky condition for each kb interval

k_b interval	Sky condition
$k_b \leq 0.1$	Totally covered
$0.1 < k_b \leq 0.2$	Totally covered
$0.2 < k_b \leq 0.3$	Mostly covered
$0.3 < k_b \leq 0.4$	Mostly covered
$0.4 < k_b \leq 0.5$	Partly covered
$0.5 < k_b \leq 0.6$	Partly covered
$0.6 < k_b \leq 0.67$	Mostly clear
$k_b > 0.67$	Totally clear

ii Last approach

In the initial approach, 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 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 2-1, we show two images of the sky with similar direct fraction index while the passing clouds are very different. The left picture shows cirrostratus and the right picture shows broken clouds.



(a) (b)

Figure 2-1. 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 has the smallest average dissimilarity to all other objects in the cluster. 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.

- 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_{1-min}^{h}).
- 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 ($\max_{1-min}^{h}$).

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.

$$s(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}} \qquad \text{Ec. (2–6)}$$

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 high *s (i)* values are well clustered. The average silhouette value can be used to evaluate the quality of the classification.

In Figure 2-2 we present the silhouette plot and the cluster groups together with their centroids of the interval $0.4 < k_b \leq 0.5$, 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.

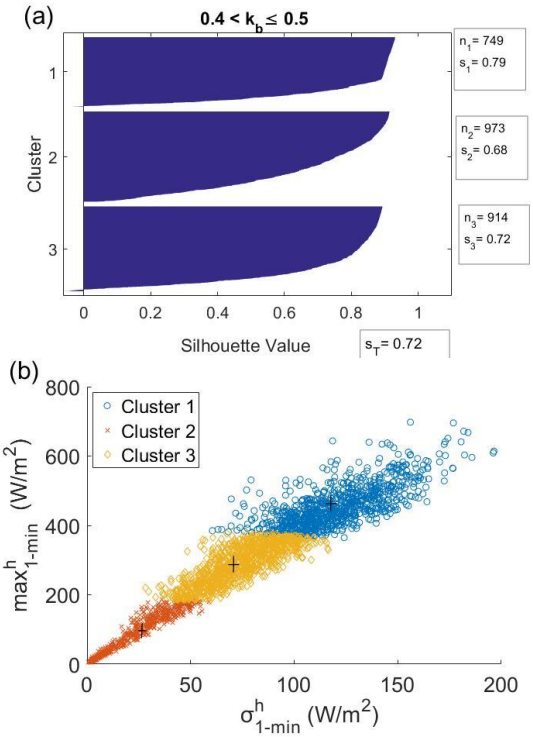


Figure 2-2. Silhouette plot of the classification obtained for the interval $0.4 < k_b \leq 0.5$. (a) Clusters and centroids for the interval $0.4 < k_b \leq 0.5$. (b)

Twenty-four groups are constructed using this cluster methodology: three for each k_b bin of width 0.1. Table 2-4 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 2–4. 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/m²)	max_{1-min}^{h} (W/m²)	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

2.2.2. Clear sky equivalent DNI condition identification

We use the term "*clear sky equivalent DNI*" to refer to either DNI during clear sky conditions or DNI during all-sky conditions that is similar to clear sky irradiance. In these situations the stochastic component is null and the synthetic data will be equal to the deterministic component. In order to generate synthetic data according to that sky condition, we aim to identify these clear sky equivalent DNI periods.

i Initial approach

In the initial approach, the decision of whether or not to include fluctuations in the synthetic data was led by a daily index. We assumed a daily direct fraction index greater than 0.65 represents a clear sky profile along all the day, but we found the model did not properly reproduce the hazy days (Gueymard, 2005) when the daily k_b values suggest partly cloudy days ($k_b \approx 0.55$) but the DNI may show negligible fluctuations. At this point, we defined a new coefficient P2 in order to identify hazy days when the stochastic component is equal to zero.

The coefficient P2 is based on the concept of tortuosity, τ, 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.

$$\tau = {}^{L}/_{X} \qquad \text{Ec. (2–7)}$$

Figure 2-3 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.

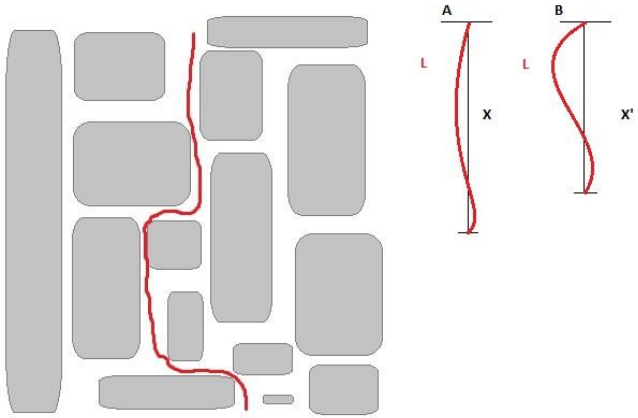


Figure 2-3. Tortuosity in porous media.

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 the estimation as the integral of the squared derivative of the curve divided by the curve length, *L*.

$$\tau = \frac{\int_{t1}^{t2} (f'(t))^2 dt}{L} \qquad \text{Ec. (2–8)}$$

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

$$S2 = \sum_h \left\{ k_t{}^{h+2} - \left(2 \cdot k_t{}^{h+1}\right) + k_t{}^h \right\}^2 \qquad \text{Ec. (2–9)}$$

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

$$P2 = \sum_h \{ Ibn^{h+2} - (2 \cdot Ibn^{h+1}) + Ibn^h \}^2 \qquad \text{Ec. (2–10)}$$

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 only for daily values of k_b greater than 0.3. In an empirical approach observing the daily DNI profiles of an entire year for the location of Seville, we identified a threshold value of P2 of $215 \cdot 10^3$ W^2/m^4 to distinguish days with and without fluctuations with an 82% of success.

i Last approach

Analysing the initial approach, we found that in many cases, we added fluctuations in clear sky equivalent DNI hourly 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 in an hourly basis (Larrañeta et al., 2017-a). The approach is performed following an iterative process to fit a clear sky envelope day by day and then, by the comparison of the differences between the means, the slopes and the lengths of the measured and theoretical clear sky curves, flag hourly periods as clear or with fluctuations.

In the first step of the iterative process, intending to have an initial envelope clear sky curve, we use the parameters A and B of the clear sky model A-B (paragraph 2.2.1-i) fitted to the maximum irradiance values divided by the correction due to Earth-Sun distance obtained for each solar angle higher than 5°. The hourly direct fraction index ${k_b}^h$ is afterwards calculated as in Ec. 2-3. The initial hourly values flagged as clear are defined by means of a direct fraction index greater than 0.65.

In a second step, we identify, for each day, those hours with a direct fraction index greater than 0.65. If we found more than two hours matching this condition, we run an empirical fit of the clear sky envelope to those hourly values using a least square procedure. Under the assumption that the state of the atmosphere does not change substantially from one day to the next, the fitted A and B parameters will remain constant until the exposed conditions (more than two hours with ${k_b}^h$ greater than 0.65) appear again. In days with only few hours flagged as initially clear separated in time, the method will generate unreal A and B parameters. To overcome those situations, if the fitted A and B parameters exceed the threshold maximum and minimum values, they values from the previous day will be used.

Those threshold maximum and minimum values are calculated using the measurements of direct solar radiation during 14 years (2000–2013) from the training database. We calculate the A and B parameters assuming the same conditions as

exposed in the previous paragraph, but in this case, due to the high resolution of the available data, a few hours flagged as initially clear correspond to many more points and therefore, the fit is generally better. The threshold maximum and minimum values are the Percentile 99 and Percentile 01 of the obtained *A* and *B* values for the entire dataset. Figure 2-4 represents the boxplot of the *A* and *B* values calculated for 14 years at the location of Seville.

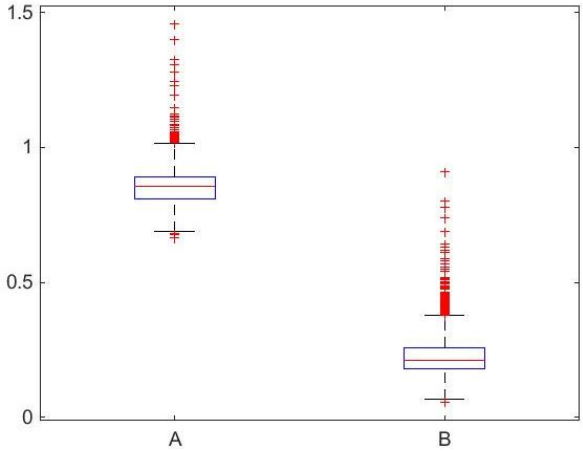


Figure 2-4. A-B clear sky model parameters boxplot for 14 years at the location of Seville.

In order to appreciate the strength of correlation between both parameters we present a scatter plot of the daily fitted values in Figure 2-5. Outliers have not been included in the plot.

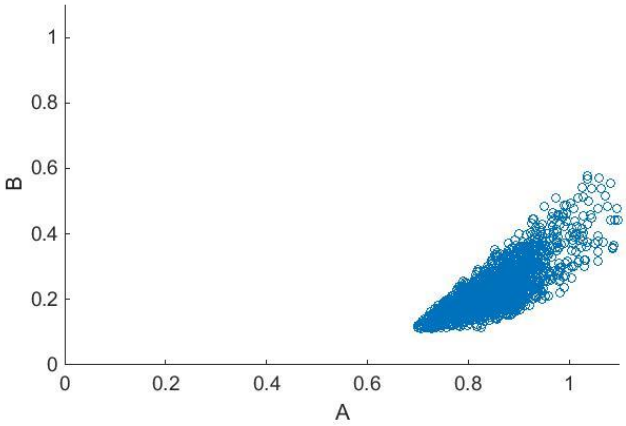


Figure 2-5. A-B clear sky model parameters scatter plot for 14 years at the location of Seville.

From Figures 2-4 and 2-5, we can observe that there is a strong correlation between both parameters and that most of the outliers appear for *A* values greater than 1.03 and *B* values greater than 0.4.

Once obtained a clear sky curve fitted to each day, the next step consists of identifying periods with irradiance similar to the clear sky irradiance from the comparison of both curves hour by hour. The criteria used are based on the method proposed by Reno et al, (2016) to identify clear sky equivalent periods of GHI from 1-min observations but in our case, because of the lower resolution data (hourly time step), we use only 3 of the 5 metrics originally proposed. Each hour is defined by two points, which leads to the comparison of linear segments. In clear conditions, at lower solar angles the variation of the hourly mean from one hour to the next is greater than for solar angles close the solar noon because the daily clear sky solar radiation curve is similar to a Gaussian curve. Therefore, we divide the daily curves into three intervals each covering 1/3 of the maximum solar angle. Figure 2-6 presents an example of the threshold values for each interval on the clear sky curve of a summer day.

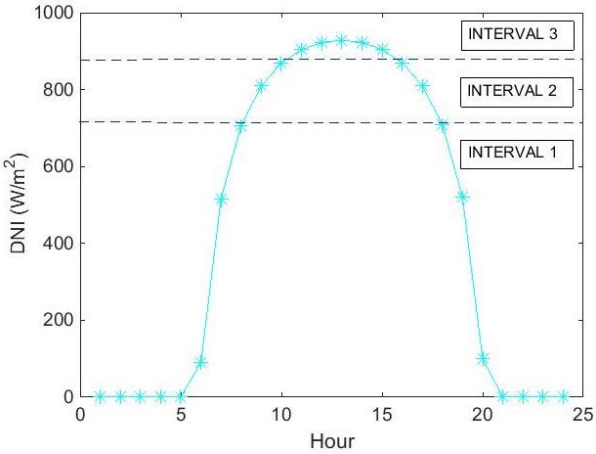


Figure 2-6. Example of the three independently analysed intervals depending on the maximum solar elevation for a summer day.

The identification of the clear sky equivalent DNI consists on the concurrence of three criteria depending on the hourly mean and the slope and length of the straight line. Note that each criterion is not overly restrictive on its own, but the combination of the three of them would determine whether the analysed hour is flagged as clear or not. We quantify the threshold values by empirically analysing the daily profiles of fourteen years of aggregated hourly DNI values (2000-2013) of the training set.

Hourly mean criterion

The first criterion involves the comparison of the hourly clear sky values and the measured means. The analysis consist in the calculation of the absolute percentage differences of the measured and clear sky radiation $D_{cs\text{-}m}{}^{i}$. A large difference in the means represents an alteration condition due to passing clouds. Table 2-5 presents the limit values for each solar interval.

$$D_{cs-m}{}^{i} = \text{abs}\left(100 \cdot \frac{I_{bn_{cs}}{}^{i} - I_{bn}{}^{i}}{\mathrm{I}_{bn}{}^{i}}\right) \qquad \text{Ec. (2–11)}$$

Where the subscript *i* represents the time instant, *cs* the clear sky and *m* the measured data.

Table 2–5. Threshold clear sky identification values for the hourly mean criterion

Solar elevation	*Condition*	*Classification*
Interval 1	$D_{cs\text{-}m}{}^{i} < 35\%$	Clear
Interval 2	$D_{cs\text{-}m}{}^{i} < 15\%$	Clear
Interval 3	$D_{cs\text{-}m}{}^{i} < 10\%$	Clear

The shadows that come from the horizon obstacles and the fact of working with hourly means involves a source of error affecting mainly low solar elevations (interval1). For these reason, the identification of threshold clear sky values in this criterion are less restrictive the lower the solar angle. The threshold $D_{cs\text{-}m}{}^{i}$ values correspond to quantiles of about 0.5 (0.46, 0.54 and 0.52 for the intervals 1, 2 and 3 respectively). Values adjusted to the location under study.

Whenever $D_{cs\text{-}m}{}^{i}$ is lower than 2.5%, the hour is classified as clear regardless the rest criteria output. This statement includes points where the measured DNI is higher than the clear sky DNI.

Slope criterion

The second criterion consists on the comparison of the slopes of the straight lines that join two hourly mean values. For each hour, the slope would be the variation of the hourly DNI divided by the time interval.

$$S_{cs}{}^{i} = {I_{bn_{cs}}{}^{i} - I_{bn_{cs}}{}^{i-1}}\Big/{T^{i} - T^{i-1}} \qquad \text{Ec. (2–12)}$$

$$S_{m}{}^{i} = {I_{bn}{}^{i} - I_{bn}{}^{i-1}}\Big/{T^{i} - T^{i-1}} \qquad \text{Ec. (2–13)}$$

We assume that a difference between the measured and the clear sky slope signs implies an alteration on the sky condition caused by passing clouds. The condition for the clear sky identification in this criterion is expressed in Table 2-6.

Table 2–6. Condition for the clear sky identification for the slope criterion.

Solar elevation	*Condition*	*Classification*
Interval 1,2,3	Sign ($S_{cs}{}^{i}$) = Sign ($S_{m}{}^{i}$)	Clear

Line length criterion

The third criterion consist on the analysis of the length of the straight lines that join two hourly mean values and their corresponding absolute percentage differences calculated as follows:

$$L_{cs}{}^{i} = \sqrt{(I_{bn_{cs}}{}^{i} - I_{bn_{cs}}{}^{i-1})^2 + (T^{i} - T^{i-1})^2} \quad \text{Ec. (2–14)}$$

$$L_{m}{}^{i} = \sqrt{(I_{bn}{}^{i} - I_{bn}{}^{i-1})^2 + (T^{i} - T^{i-1})^2} \quad \text{Ec. (2–15)}$$

$$LD_{cs-m}{}^{i} = \text{abs}\,(100 \cdot \frac{L_{cs}{}^{i} - L_{bn}{}^{i}}{\mathrm{L}_{bn}{}^{i}}) \quad \text{Ec. (2–16)}$$

Following a similar assumption as in the hourly mean criterion, a large difference in the line length represents an alteration condition due to passing clouds. The length of the line that joins two hourly mean clear sky values varies substantively depending on the solar elevation therefore we combine the absolute differences on the line lengths and the clear sky line length. The combination of these calculations leads to the line length clear sky classification as presented in Table 2-7.

Table 2–7. Line length classification conditions for the clear sky identification.

Solar elevation	Condition	Classification
Interval 1	$\mathrm{LD}_{\mathrm{cs-m}}{}^{\mathrm{i}} < 25\ \%$ & $L_{cs}{}^{i} > 220$	Clear
Interval 2	$\mathrm{LD}_{\mathrm{cs-m}}{}^{\mathrm{i}} < 35\ \%$ & $L_{cs}{}^{i} > 110$	Clear
Interval 3	$\mathrm{LD}_{\mathrm{cs-m}}{}^{\mathrm{i}} < 120\ \%$ & $L_{cs}{}^{i} < 30$	Clear

For low solar angles, the hourly mean criterion is lower than for high solar angles

however, the line lengths follow the inverse pattern. The threshold $LD_{cs-m}{}^{i}$ values correspond to quantiles of 0.46, 0.54 and 0.52 for the intervals 1, 2 and 3 respectively.

Once identified the periods with irradiance similar to the clear sky irradiance, we execute a daily iterative process that consists on the readjustment of the clear sky model to those flagged clear sky periods through the process previously exposed until the *A* and *B* parameters converge. Figure 2-7 shows the block diagram of the process. The subscript j represents the modelled day

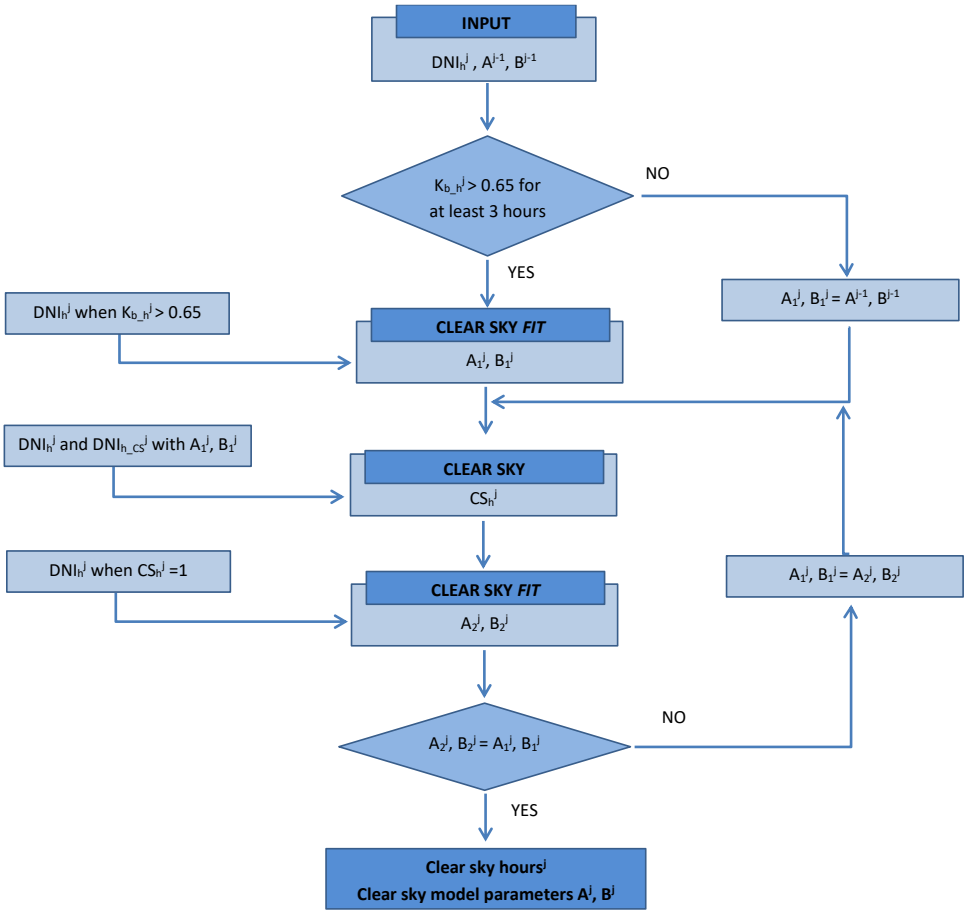
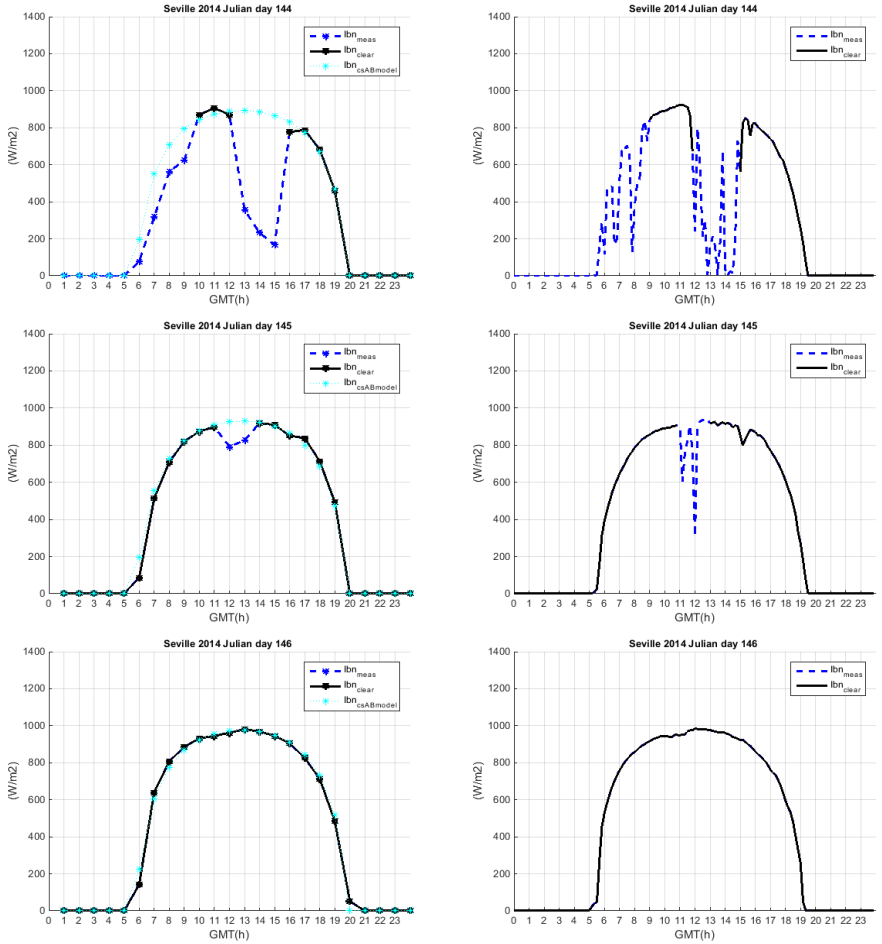
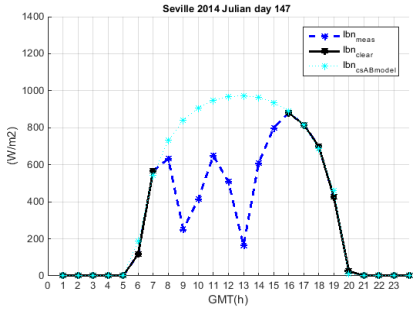
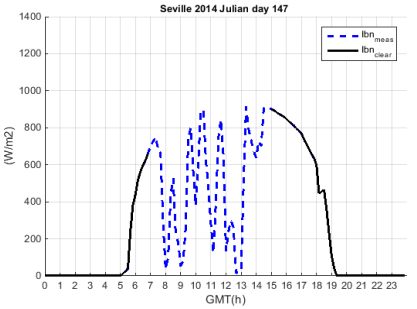


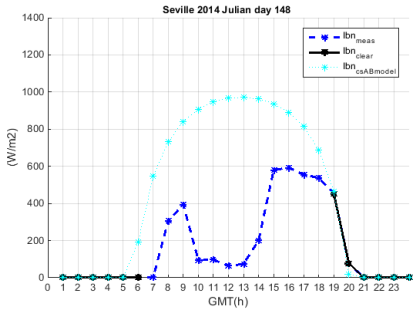
Figure 2-7. Block diagram of the process

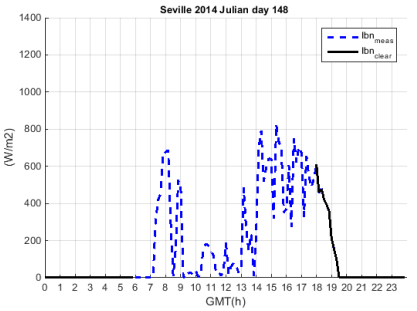
To assess the performance of the model, we have utilized one year (2014) of the training dataset. The instantaneous DNI data have been integrated into hourly means for the execution of the model and to 10-min resolution to observe the performance of the model. Ten consecutive daily profiles are illustrated in Figure 2-8. The charts on the left side show the DNI hourly means of the daily profiles and the charts on the right show the daily profiles in 10-min means. The measured data is printed in discontinuous blue, the identified clear sky equivalent DNI periods are presented in continuous black and for the case of hourly means, and the clear sky fitted shape is presented in a dotted cyan line.

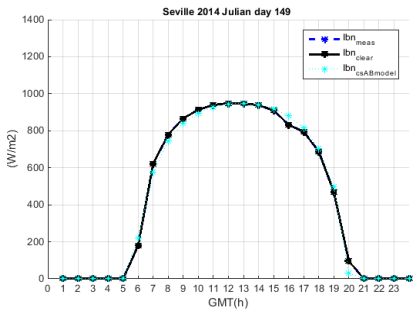


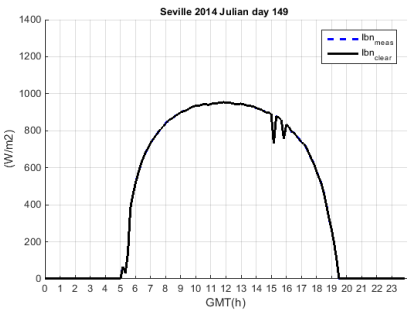

Seville 2014 Julian day 147
Ibn meas
Ibn clear
Ibn csABmodel
(W/m2)
GMT(h)

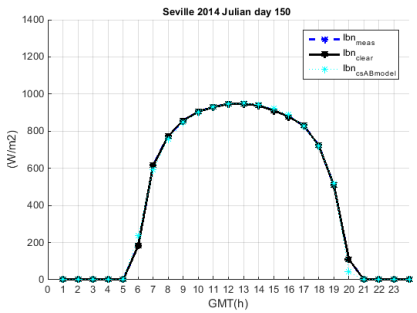

Seville 2014 Julian day 147
Ibn meas
Ibn clear
(W/m2)
GMT(h)

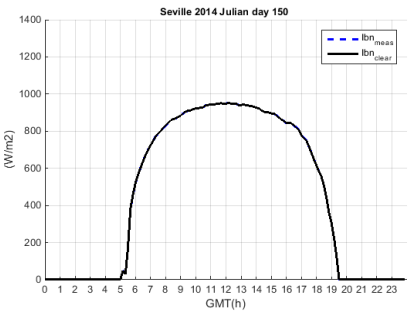

Seville 2014 Julian day 148
Ibn meas
Ibn clear
Ibn csABmodel
(W/m2)
GMT(h)


Seville 2014 Julian day 148
Ibn meas
Ibn clear
(W/m2)
GMT(h)


Seville 2014 Julian day 149
Ibn meas
Ibn clear
Ibn csABmodel
(W/m2)
GMT(h)


Seville 2014 Julian day 149
Ibn meas
Ibn clear
(W/m2)
GMT(h)


Seville 2014 Julian day 150
Ibn meas
Ibn clear
Ibn csABmodel
(W/m2)
GMT(h)


Seville 2014 Julian day 150
Ibn meas
Ibn clear
(W/m2)
GMT(h)

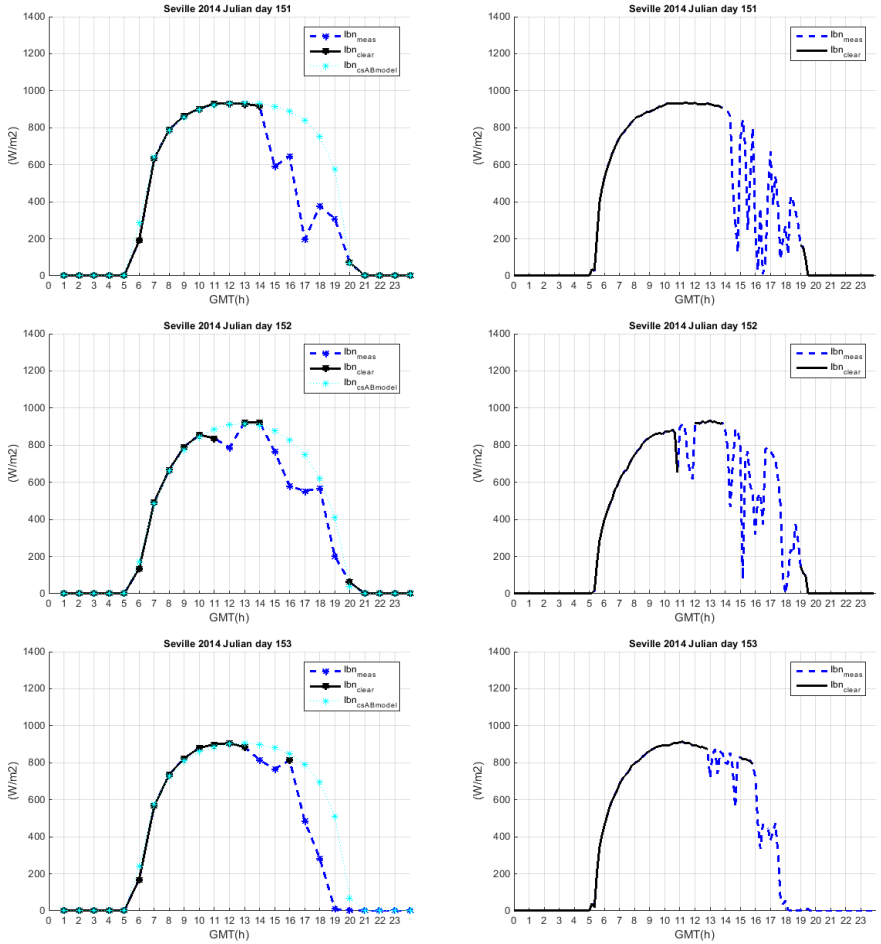


Figure 2-8. Illustrative daily examples of the last approach clear sky equivalent DNI model performance.

2.2.3. Stochastic component reproduction

i Initial approach

In the original SA model (Polo et al., 2011), the stochastic component was reproduced by means of beta distributions fitted to the normalized standard deviations of the instant values to the hourly means from the training dataset, being the sign of the fluctuation randomly added from a normal distribution.

The standard deviations of the high resolution values of the training set were normalized 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.

In the original model the normalization was performed using a unique maximum value of the deviation of the 10-min direct normal irradiance data with respect to the hourly. In the improved model, there is a maximum deviation value for each defined interval. The largest fluctuations occur during the passage of clusters of low clouds with high density like cumulus or stratocumulus, therefore the maximum deviations are found in the central intervals of the clearness index as shown in Table 2-8.

Table 2–8. Maximum deviation of the 10-min direct normal irradiance with respect to the hourly mean for each k_b interval.

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/m^2)	215	345	401	431	442	435	430
σ_{max}^{orig} (W/m^2)	442	442	442	442	442	442	442
Diff (W/m^2)	227	97	41	11	0	7	12

The highest differences between the improved and original models are found for low values of the clearness index because of the smaller DNI fluctuations that occur due to the presence of dense and compact clouds during the mostly-covered sky conditions.

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, when the DNI hourly values are lower than 90 W/m², the stochastic component is equalled to zero.

ii Last approach

Following the initial approach, 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 high resolution time scale (dif_{1-min}) of each cluster. The procedure is described below:

1. Calculate the ECDF of the dif_{1-min} for each cluster (Section 2.2.1-ii).
2. Select the group from which the 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.
3. 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 2), whose cumulative probability is the same as the value of the random number R, obtaining thus the stochastic component (Ibn_{stoc}^{i}). Figure 2-9 shows a graphical explanation of the step 3.

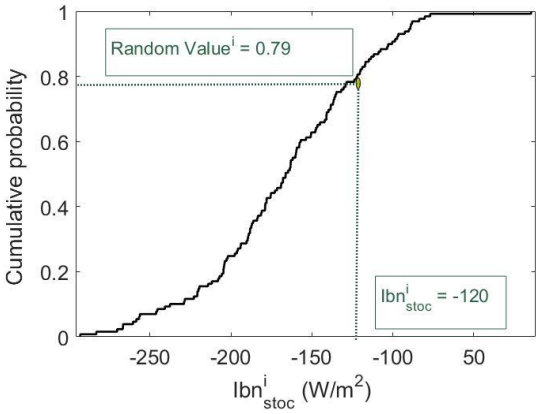


Figure 2-9. Graphical reproduction of step 3 of the last approach for the stochastic component reproduction. For a randomly generated value of probability of probability of 0.79 we obtain a stochastic component of -120 W/m².

Within the last approach, the ECDFs of the training set 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.

In the last approach, we intend to generate synthetic DNI datasets in the 1-min resolution but most of the cloud transients are gradual and last longer than 1-min (Larrañeta et al., 2017 -b). In that research work we intended to delve deeper into the knowledge of the cloud transients. We evaluated the cloud transients by analysing the dynamics of the direct fraction index k_b for one year (2014) of DNI measurements of the training dataset in different time steps. We use six temporal resolutions calculated as the integration of the 5-seg measurements. 1-min, 5-min, 10-min, 15-min, 30-min and 1-h besides the instant 5-seg values. We calculate the direct fraction index in each of the selected time steps dividing the observed DNI data by the clear-sky DNI using the A-B model (2.2.1-i). The sky condition (related to cloudiness or cloud type) is defined by means of the levels of attenuation of the direct solar radiation reaching the earth surface. Based on Martinez-Chico et al. (2011) we can classify the type of clouds into 5 groups through intervals of the direct fraction index of 0.2. The k_b thresholds selected for performing this classification are presented in Table 2-9.

Table 2–9. Classification of sky conditions.

Group	G1	G2	G3	G4	G5
k_b	$k_b \leq 0.2$	$0.2 < k_b \leq 0.4$	$0.4 < k_b \leq 0.6$	$0.6 < k_b \leq 0.8$	$k_b > 0.8$

To quantify the occurrence of different cloud transients, we count the number of times that the k_b varies from one group to another, this is, counting changes in the sky condition. We also quantify the duration of the transient by counting the time that the sky condition remains inside the threshold values.

For the assessment of the results we presented the transition matrices. Each value represents the transition from an initial sky condition (column) to a final sky condition (row). In this manner, values in the diagonal represent the persistence, values over the diagonal represent the entrance of clouds and values above the diagonal represent the exit of clouds. We divide the tables into number of occurrences, and percentage of occurrences where we omit the persistence in order to quantify only the cloud transients. Table 2-10 presents the transition matrixes for the 1-min resolution.

Table 2–10. Transition matrix for the integrated 1-min resolution.

		Number of occurrences Entrance of clouds					Percentage of occurrences Entrance of clouds				
		G1	G2	G3	G4	G5	G1	G2	G3	G4	G5
Exit of clouds	G1	57613	1630	76	22	9		11%	1%	0%	0%
	G2	1666	6416	1632	100	10	12%		11%	1%	0%
	G3	99	1650	7167	1700	98	1%	11%		12%	1%
	G4	17	129	1748	11932	1789	0%	1%	12%		12%
	G5	11	20	101	1861	142635	0%	0%	1%	13%	

Certain symmetry between values above and below the diagonal is observed. The values closer to the diagonal are the most repeated, i.e., transitions are progressive. As the time step increases, this trend is reduced. In Figure 2-10 we present the percentage of progressive transitions depending on the time step.

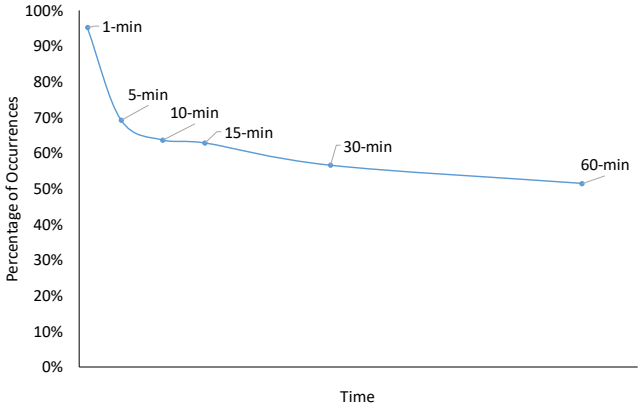


Figure 2-10. Percentage of occurrences of the progressive transients depending on the time step.

This same conclusion can be observed on Table 2-11. Aiming to compute the entrance of the clouds, we count the number of occurrences that the sky condition changes from clear ($k_b > 0.8$, G5) to any of the other defined state. Table 11 shows that as we increase the time step there is a shift towards a reduction in the percentage of progressive transitions and an increase in the steeper transitions. This can be explained because in a larger time step, more situations are integrated into a single value, averaging the result of several phenomena.

Table 2–11. Transition from G5 to the rest of the groups.

		Number of occurrences					Percentage of occurrences			
		Group				TOTAL	Group			
		G1	G2	G3	G4		G1	G2	G3	G4
Time Step	5-seg	246	696	1644	9167	11753	2%	6%	14%	78%
	1-min	9	10	98	1789	1906	0%	1%	5%	94%
	5-min	52	123	239	741	1155	5%	11%	21%	64%
	10-min	47	73	173	436	729	6%	10%	24%	60%
	15-min	38	56	123	306	523	7%	11%	24%	59%
	30-min	24	40	55	159	278	9%	14%	20%	57%
	1h	11	22	48	101	182	6%	12%	26%	55%

In view of the results of this research, in the last approach of the stochastic component reproduction, the random number R (step 3) of each instant should have a dependence on the previous instant, since in the 1-min resolution, most of the transients are progressive (Figure 2-10). In an empirical approach, we have taken the decision of limiting the difference between two consecutive values of R to ±0.3.

$$R^i = R^{i-1} \pm 0.3 \qquad \text{Ec. (2–17)}$$

2.2.4. Similarity on the daily sums

The application of the original model in the generation of synthetic series often resulted in significant differences between the cumulative daily values of the hourly time series to be downscaled and the synthetic 1-min 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 uncertainty of a 2% is accepted since that value represents the uncertainty of most of the first class pyrheliometers. This improvement is implemented in both, the initial and last approach.

2.2.5. Non-published improvements

We have already implemented the coupled GHI plus DNI synthetic generation to

the last approach of the SA model. The method follows the same methodology as the synthetic generation of the DNI but also for the GHI. The deterministic component of the high resolution GHI is calculated from the cubic interpolation of the hourly means every four hours. The stochastic component is calculated from the ECDFs of the clustered database (2.2.1-ii). To generate the GHI coupled to the transients of the DNI, we follow the stochastic component reproduction steps (2.2.3-ii). In the third step, we generate a single random value R for each instant that is used to locate the value whose cumulative probability is the same as the value of the random number but in this case within the ECDF of the selected cluster of the DNI and GHI in parallel. This way, transients will be either negative (entrance of clouds) or positive (exit of clouds) at the same time in the GHI and the DNI with a value corresponding to the same probability of occurrence.

2.2.6. Other applications for the SA model

The main concept of the SA model can be applied to many phenomena related to the solar radiation of which a relatively large database is available. In the research paper "A methodology for the stochastic generation of hourly synthetic direct normal irradiation time series" (Larrañeta et al., 2018-b) we use this concept for the synthetic generation of hourly DNI values from hourly GHI means intending to keep the same frequency distribution as the measured DNI data with similar results to those of the most common decomposition models (Gueymard and Ruiz-arias 2014) in terms of daily, monthly and annual deviations.

This work was motivated by the necessity of the CSP industry to have around fifteen years of high resolution DNI data for plant performance evaluation purposes when only two or three years of hourly DNI and GHI measured data are available on the site under study and fifteen years of GHI data can be obtained from a nearby location. In (Larrañeta et al., 2018-b) we intend to fulfil that necessity.

For a given value of global radiation, there is a range of possible direct radiation values conditioned to the state of the atmosphere. In Larrañeta et al., 2018-b, we intend to emulate, in a simplified form, the variability of the atmospheric components that affect the attenuation of the direct component. The methodology proposed in that research paper follows the SA model approach of dividing the solar radiation into a deterministic and a stochastic component. The deterministic

component is generated through a classical decomposition model (Erbs et al., 1982). The stochastic component, which is added to the deterministic component, is calculated from the cumulative frequency distribution of a sufficiently large database. In this research, we used the aggregated hourly DNI and GHI values of 14 years (2000–2013) for the location of Seville for the training of the stochastic component. Once trained, the stochastic component could be reproduced globally without local adaptation. For the deterministic component we require at least two years of hourly aggregated DNI and GHI measurements on the site under study. We used the measurements of three consecutive years of three locations with different climates for the training of the deterministic component (two years) and to test the model performance (1 year) (Table 2-12).

Table 2–12. Locations selected for the application of the model for the stochastic generation of hourly synthetic direct normal irradiation time series (Larrañeta et al., 2018-b).

	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

For the deterministic component, we calculated the fourth order polynomial fit to the point cloud of the hourly clearness index versus the hourly direct fraction index for solar elevations higher than 10° and k_b 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).

For the stochastic component, we used the procedure described in paragraph 2.2.3-ii to reproduce the ECDF of the training dataset. For each k_t, the stochastic component was fitted to the ECDF of the difference between the measured k_b and the deterministic k_b. In order to have a statistically significant ECDF, we divide the k_t values 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 behaviour of the solar radiation varies significantly depending on this parameter (Tovar-Pescador 2008).

The synthetic hourly DNI data was generated by adding the stochastic component

to the deterministic component. Figure 2-11 shows the scatter plots of the hourly direct fraction index versus the hourly clearness index of the measured (left) and modelled (right) data for the three locations used for the application of the procedure (a-c).

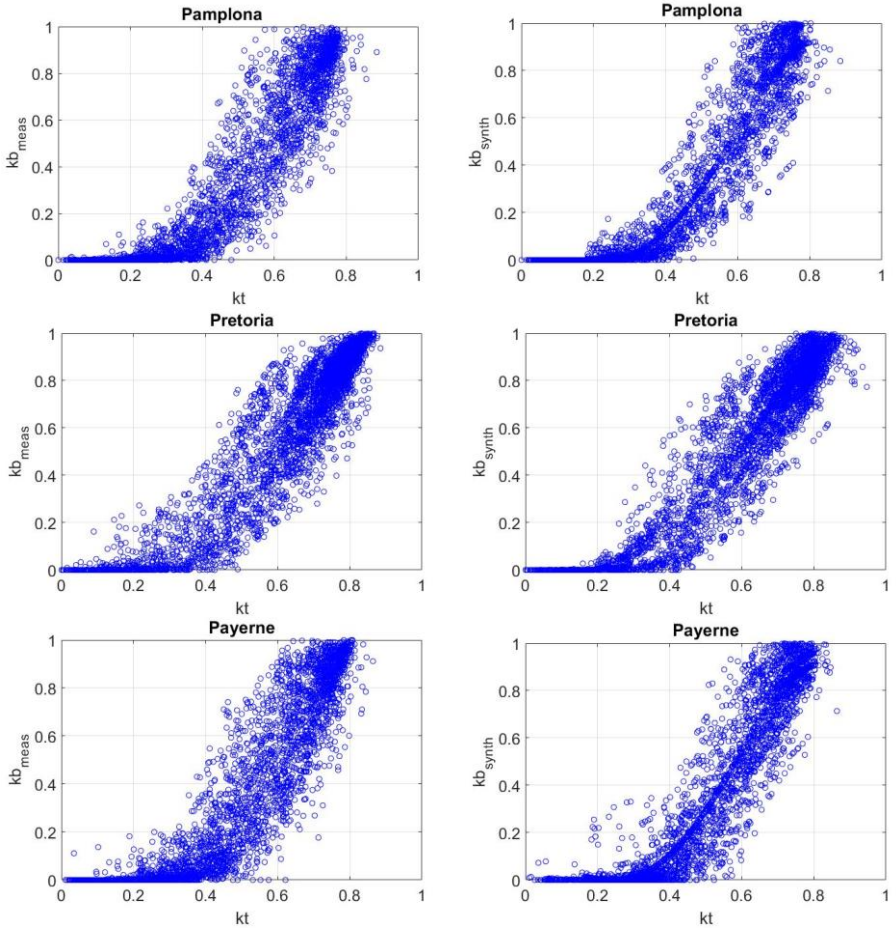


Figure 2-11. Scatter plots of the k_t versus the measured (left) and synthetic (right) k_b for Pamplona, Pretoria, and Payerne.

2.3. Non-Dimensional (ND) model

The original method (Fernández-Peruchena et al., 2015) consists in the normalization of the daily DNI curve by the clear-sky envelope, creating daily dynamic paths from observed DNI data. The method transforms each daily 1-min DNI curve into a dimensionless curve where the normalized time and the DNI range from 0 to 1. In Figure 2-12 (a), we show the dimensionless daily shape obtained from a randomly selected day Figure 2-12 (b). 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, the two parameters defining the ASHRAE exponential model (MacPhee, 1972) are adjusted 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. The non-dimensionalization of the temporal scale is performed day by day by dividing the elapsed Universal Time (UT) since sunrise by day length (time between sunrise and sunset). This way, the x-axis value of 0 corresponds to sunrise and the x-axis value of 1 corresponds to sunset. The non-dimensionalization of the solar DNI scale is performed by dividing each actual solar DNI value by the corresponding DNI value of the clear-sky DNI envelope. This way, the y-axis value of 0 corresponds to totally overcast conditions and the y-axis value of 1 corresponds to clear sky conditions.
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. At least one year of 1-min measured solar radiation data in the site under study is required.

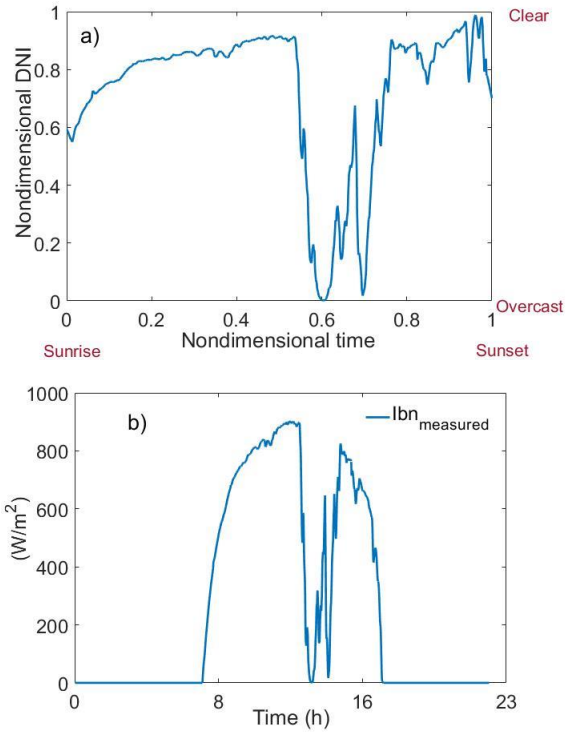


Figure 2-12. Dimensionless daily profile (a) of a randomly selected daily DNI curve (b).

In the case of the ND model, we have published so far only one approach for the improvement of the original method (Larrañeta et al., 2018-b) for the synthetic generation of 1-min DNI data from hourly means. However, a few more improvements have been developed but haven´t yet been published. They are described in paragraph (2.3.3.). The improvements already published can be divided in two main blocks.

- Clear sky envelope calculation. We found that the third step of the original methodology 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 obtained from one

location, to generate synthetic high temporal resolution DNI data in any location.

- Selection of the most similar day. The original method seeks for the most similar day between the normalized database and the input hourly DNI data in terms of energy. Since variability and temporal distribution of the irradiation have a great impact on CSP plant production, we propose to implement a method to seek for the most similar day in terms of energy, variability and distribution.

2.3.1. Clear sky envelope calculation

For the calculation of the dimensionless 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 ($I_{bn_{cs}}$). We implement the clear sky A-B model (paragraph 2.2.1-i), Ec (2-4).

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 Ibn/E_0 for every value of solar elevation higher than 5°. We use the relationship Ibn/E_0 to make 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 Figure 2-13 because the hourly resolution trends to smooth the amplitude of the cloud transients. In Table 2-13, we present the annual A-B parameters in the 1-min and the hourly resolution for 14 years (2002-2015) from the training database.

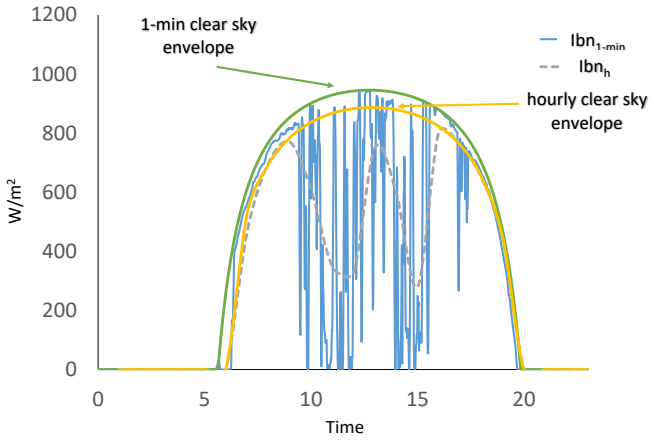


Figure 2-13. Daily clear sky curve fitted to measured 1-min and hourly DNI data.

Table 2–13. Annual 1-min and hourly A-B clear sky parameters for the years 2002-2015 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

For the normalization of the training dataset, we use the annual empirically calculated A-B parameters that define the clear sky envelope in the 1-min resolution year by year.

2.3.2. Selection of the most similar day

In the original model, 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-Tejera 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 (k_b^d)to characterize the daily energy of a given day following next equation:

$$k_b^d = \frac{Ibn^d}{Ibn_{cs}{}^d} \ , \qquad \text{Ec. (2–18)}$$

where Ibn^d is the daily DNI and $Ibn_{cs}{}^d$ 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.

$$VI = \frac{\sum_{i=2}^{n}\sqrt{\left(I_{bn}{}^{i} - I_{bn}{}^{i-1}\right)^2 + \Delta t^2}}{\sum_{i=2}^{n}\sqrt{\left(I_{bn_{cs}}{}^{i} - I_{bn_{cs}}{}^{i-1}\right)^2 + \Delta t^2}} \qquad \text{Ec. (2–19)}$$

$I_{bn_{cs}}$ is the hourly enveloping clear sky direct normal irradiance, the subscript i represents the time instant, Δt refers to an interval of one hour, 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.

$$F_m = \frac{Ibn_{md}^d}{Ibn^d} \qquad \text{Ec. (2–20)}$$

Ibn_{md}^d is the DNI recorded from the sunshine to the solar noon.

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 to each other are similar. We train

the algorithm with 14 years (2002-2015) of the training database from Seville. We use the daily triplets of k_b^d, 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 triplets 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 to be downscaled.

Once the most similar days of the normalized database to the input DNI are selected, to generate the synthetic 1-min DNI data, 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.

2.3.3. Non-published improvements

We have also already implemented the synthetic generation of consistent GHI and DNI in the ND model. To that end, we have firstly normalized the GHI of the training database. In this case the X-axis is also the normalized time and the Y-axis is the instant clearness index defined as:

$$k_t{}^i = I_{g0}{}^i / I_0{}^i \qquad \text{Ec. (2–21)}$$

Where $Ig0^i$ is the global horizontal irradiance and $I0^i$ is the 1-min extraterrestrial irradiance on a horizontal surface.

Then, in the selection of the most similar day, we use the daily k_t among the k_b, VI and F_m to seek for the most similar day to be unpacked. Once the day is selected, we use again the extra-terrestrial curve to generate the synthetic 1-min GHI.

We have also implemented an improvement related to the *VI* index. When quantifying the short-term variability through the daily *VI* index, the time step of the data series plays a significant role. Since the input data to be downscaled comes in the hourly resolution, the dimensionless daily profiles from the training database have also been labelled in terms of variability in the hourly resolution but this decision implies a great loss of information and has a significant impact in the generated data.

To tackle that weakness, we have implemented an algorithm to obtain VI values in the 1-min resolution from the hourly values. For predicting the high temporal resolution *VI* time series, we implement four supervised Machine Learning techniques; Lineal regression, nonlinear regression, k Nearest Neighbour (kNN) and decision tree. The algorithms have been trained with 14 years of measurements of DNI in the period 2002-2015 and tested with the measures of the entire 2016. Figure 2-14 summarizes the approach.

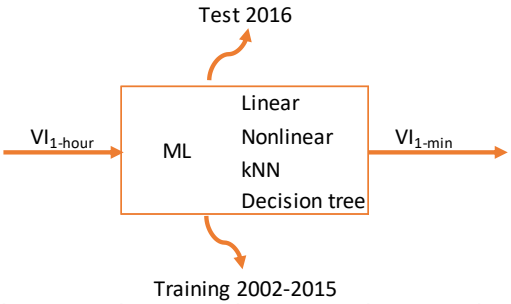
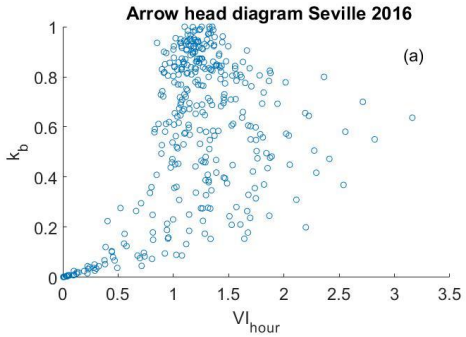


Figure 2-14. Schematic summary of the methodology

In Figure 2-15 we show an "arrow head" diagram (k_b versus VI) for one entire year (2016) of the measured set in the hourly (a) and 1-min (b) resolution and the synthetic *VI* set in the 1-min resolution (c).



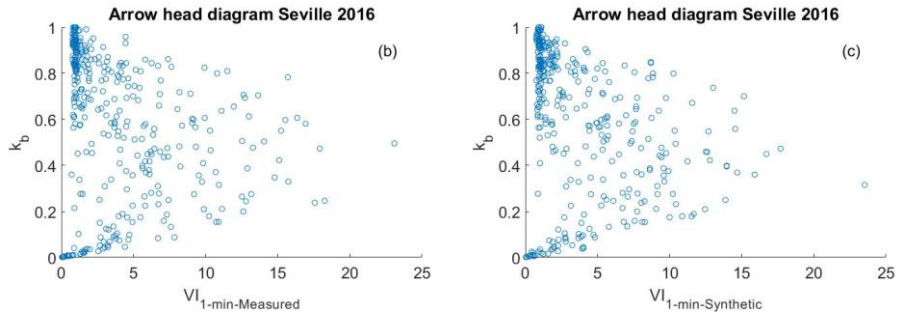


Figure 2-15. Arrow head diagram or the year 2016 for the location of Seville in the measured hourly resolution (a), the measured 1-min resolution (b) and the synthetic 1-min resolution calculated with the kNN algorithm.

It can be observed from figures a and b, that the arrow head diagram of the measured data is strongly dependent on the time scale of the data. Note that the VI values in the 1-min time scale range from 0 to 25 while the VI values in the hourly time scale range from 0 to 3.5. In the hourly resolution (a) most of the data have *VI* daily values close to 1 while in the 1-min resolution (b) it can be observed that the *VI* daily values have greater dispersion, thus variability is characterized more precisely the greater the time resolution of the available data. The arrow head diagram of the synthetic *VI* daily data in the 1-min resolution (c) presents a similar shape to the measured *VI* daily data in the 1-min resolution (b).

3. RESULTS AND DISCUSSION

In the analysis of the performance of the methods we only compare the SA model in the second approach and the ND model in its single approach because of the different resolution of the synthetically generated data with the first approach of the SA model (10-min) to the rest of the approaches (1-min). We could integrate the 1-min values to the 10-min resolution in order to compare the results of the three approaches but we understand it losses relevance since we have reached a greater resolution for the synthetic generation. We neither compare the results with the original methods since in that case, the models were trained with local data for the location under study and in the approaches here presented we generate synthetic high resolution data in anywhere using data from Seville to train the models, therefore, results are not comparable.

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.

In Figure 3-1 we present some examples of the daily profiles 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, red dotted lines), and ND modelled data (right side, green dotted line).

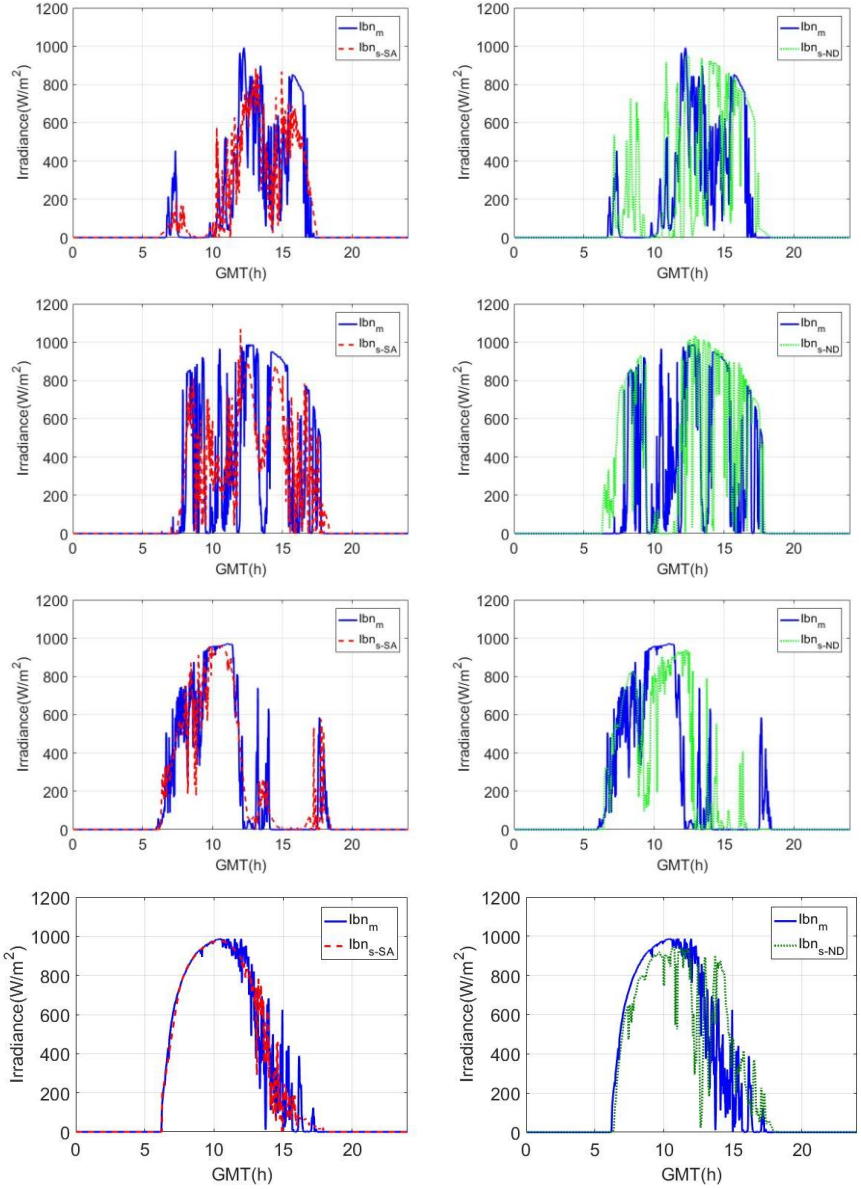


Figure 3-1. Illustrative examples of the results of the synthetic generation with the stochastic adaptation model (Ibns-SA) and non-dimensional model (Ibns-ND) compared to the measured dataset (Ibnm) for the location of Almeria.

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

3.1. Evaluation of the mean

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%.

3.2. Evaluation of the distribution

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). This index has the same units as the corresponding magnitude

$$KSI = \int_{x_{min}}^{x_{max}} D_n dx, \qquad \text{Ec. (3–1)}$$

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 $a_{critical}$.

$$KSI(\%) = 100 \cdot \frac{\int_{x_{min}}^{x_{max}} D_n dx}{a_{critical}} \qquad \text{Ec. (3–2)}$$

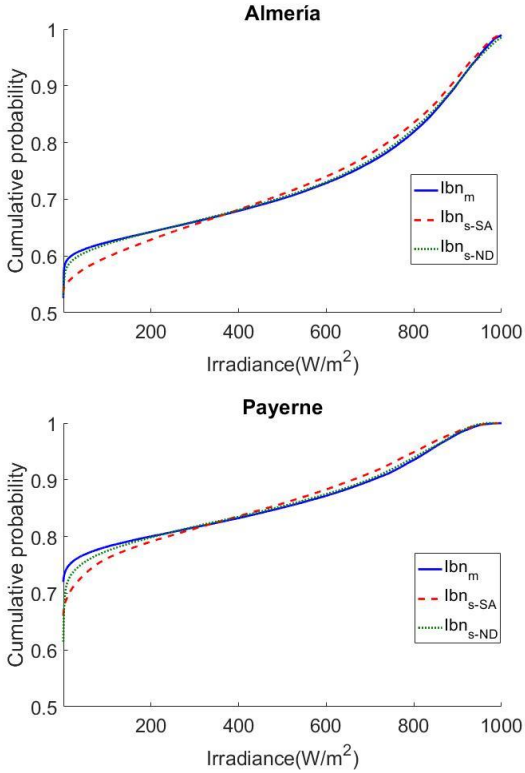
$$a_{critical} = V_c \cdot (x_{max} - x_{min})$$

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

$$V_c = {}^{1.63}/_{\sqrt{N}} \quad N \geq 35 \qquad \text{Ec. (3–3)}$$

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.

Figure 3-2 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. We can observe that the greatest differences are found for low levels of irradiance in the case of the SA model.



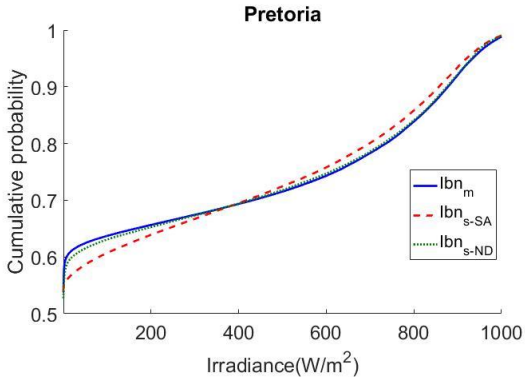


Figure 3-2. ECDFs of the measured (Ibn_m) and synthetic DNI datasets generated with the stochastic adaptation model ($Ibn_{s\text{-}SA}$) and the non-dimensional model ($Ibn_{s\text{-}ND}$).

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. Table 3-1 shows the *KSI* value of each model in the selected locations.

Table 3–1. 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/m²)	Almería (2013)	12.7	2.8
	Pretoria (2016)	15.4	3.4
	Payerne (2014)	10.6	3.6

Figure 3-3 and Figure 3-4 show a bar plot of the monthly *KSI* (W/m²) and *KSI* (%) of the DNI time series for the three locations under evaluation.

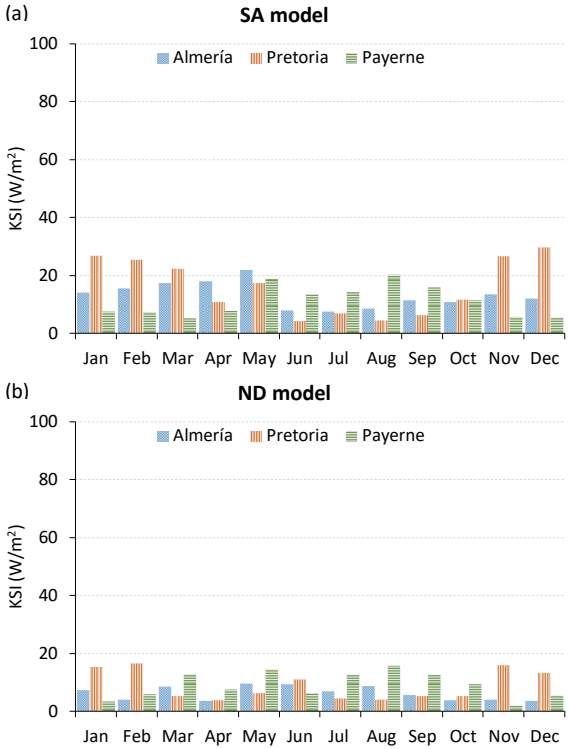


*Figure 3-3. **KSI** (W/m²) values obtained in the comparison of 1-min synthetic DNI data compared to the measured data for the locations of Almería, Pretoria and Payerne with the SA model (a) and the ND Model (b).*

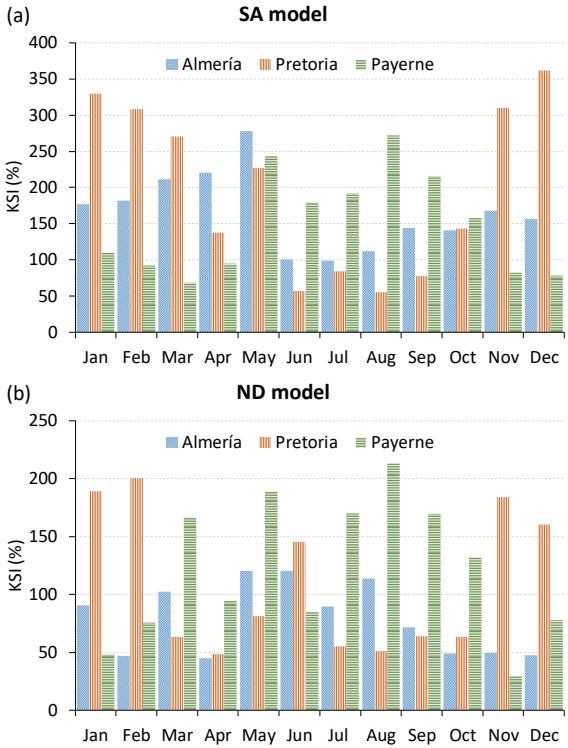


Figure 3-4. KSI (%) values obtained in the comparison of 1-min synthetic DNI data compared to the measured data for the locations of Almería, Pretoria and Payerne with the SA model (a) and the ND Model (b).

The SA method provides average monthly *KSI* values of 11.2 W/m² in Payerne, 16.0 W/m² in Pretoria and a maximum value of 29.6 W/m² in Pretoria. The ND model provides average monthly *KSI* values of 6.2 W/m² in Almería, 9.0 W/m² in Payerne and a maximum value of 16.5 W/m² in Pretoria.

There is a seasonal bias on the *KSI* strongly related to the intra-daily variability characterization of the solar radiation. The months with greater *KSI* values are found in those seasons having a large number of high variability days. Table 3-2 shows monthly *KSI* values calculated from the implemented models and ground measurements for Pretoria site, along with the number of measured days with a high intra-daily variability (*VI* index > 15). Figure 3-5 shows the histogram of the VI index

calculated with the measured and synthetic 1-min datasets for Pretoria site.

Table 3–2. Monthly KSI (W/m2) of the implemented models for the measured and synthetic DNI annual sets for the location of Pretoria and the number of days with VI index > 15 (high intra-daily variability).

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
KSI SA model (W/m^2)	26.8	25.4	22.4	10.7	17.3	4.3	6.8	4.3	6.3	11.7	26.6	29.6
KSI ND model (W/m^2)	15.3	16.5	5.2	3.8	6.2	10.9	4.4	4.0	5.2	5.2	15.8	13.1
Nº days VI>15	12	15	10	6	7	2	4	2	6	9	15	17

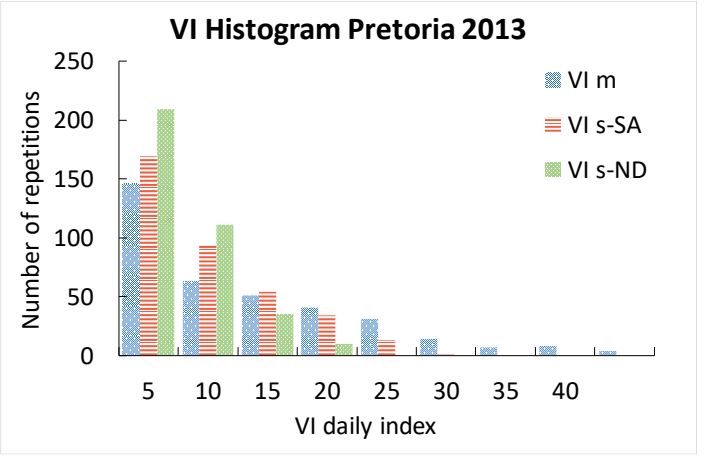


Figure 3-5. Histogram of the VI index calculated with the measured (VIm) and synthetic 1-min datasets with the stochastic adaptation model (VIs-SA) and the non-dimensional model (VIs-ND) for the location of Pretoria.

When quantifying the short-term variability through the daily *VI* index, the time step of the data series plays a significant role. Since the input data to be downscaled comes in the hourly resolution, the dimensionless daily profiles from the training database have also been labelled in terms of variability in the hourly resolution but this decision implies a great loss of information and has a significant impact in the generated data. In figure 11 we can observe that the synthetic datasets present higher

repetitions for lower VI daily values.

The greater presence of high intra-daily variability days is found from November to March, which coincides with the greater KSI values. Moreover, synthetically generated 1-min DNI series show a higher occurrence of low daily VI index values with respect to the measured one. These results suggest a weakness in the characterization of the intra-daily variability through the *VI* index in the case of the ND model. In the case of the SA model, the stochastic component reproduction (section 2.2.3-ii) may not be accurate enough since it is independent of the intra-daily variability. Additionally, the use of hourly values in the calculation of the variability index entails a loss of information.

The KSI statistic evaluates the differences of the CDFs of solar radiation measured and synthetic datasets; these differences are assumed to have an impact on plant production. Thus, 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 on its impact on plant production. In this doctoral thesis, 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 Andasol 3 plant (NREL, 2013) currently in operation. The main characteristics are summarized in Table 3-3.

Table 3–3. 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:

$$NRMSD\ (\%) = 100 \cdot \left({}^{RMSD}/_{P_{max} - P_{min}} \right), \qquad \text{Ec. (3–4)}$$

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.

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_m^i - P_s^i)^2}\ , \qquad \text{Ec. (3–5)}$$

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.

The results for the NRMSD of the thermal power produced in the three analysed locations are presented in Table 3-4.

Table 3–4. 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
NRMSDhourly (%)	Almería (2013)	2.3	12.4
	Pretoria (2016)	2.7	12.8
	Payerne (2014)	2.3	14.7
NRMSDdaily (%)	Almería (2013)	0.8	3.0
	Pretoria (2016)	1.2	3.1
	Payerne (2014)	0.8	4.1

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. The differences in the performance of both models are reduced in the daily NRMSD.

3.3. Evaluation of the autocorrelation

For the evaluation of the autocorrelation, we calculate the ramp rates (RRs) as the difference between successive data points over one minute using eq. 13:

$$RR = \left(\left({I_{bn}}^{i} - {I_{bn}}^{i-1}\right) - \left({I_{bn_{cs}}}^{i} - {I_{bn_{cs}}}^{i-1}\right)\right)\Big/_{\Delta t} \quad , \qquad \text{Ec. (3–6)}$$

where Δt refers to an interval of one minute. The units are given in W/m²·min. We calculate the absolute RRs values for the annual datasets taking into account only daytime observations for solar elevations above 5°.

Figure 3-6 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. Figure 3-7 shows a bar plot of 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 (> 500 W/m²min), the differences in the measured and synthetic ECDFs values are almost negligible. The differences found at low RR indicate a lower variability of both modelled datasets with respect to the measured dataset.

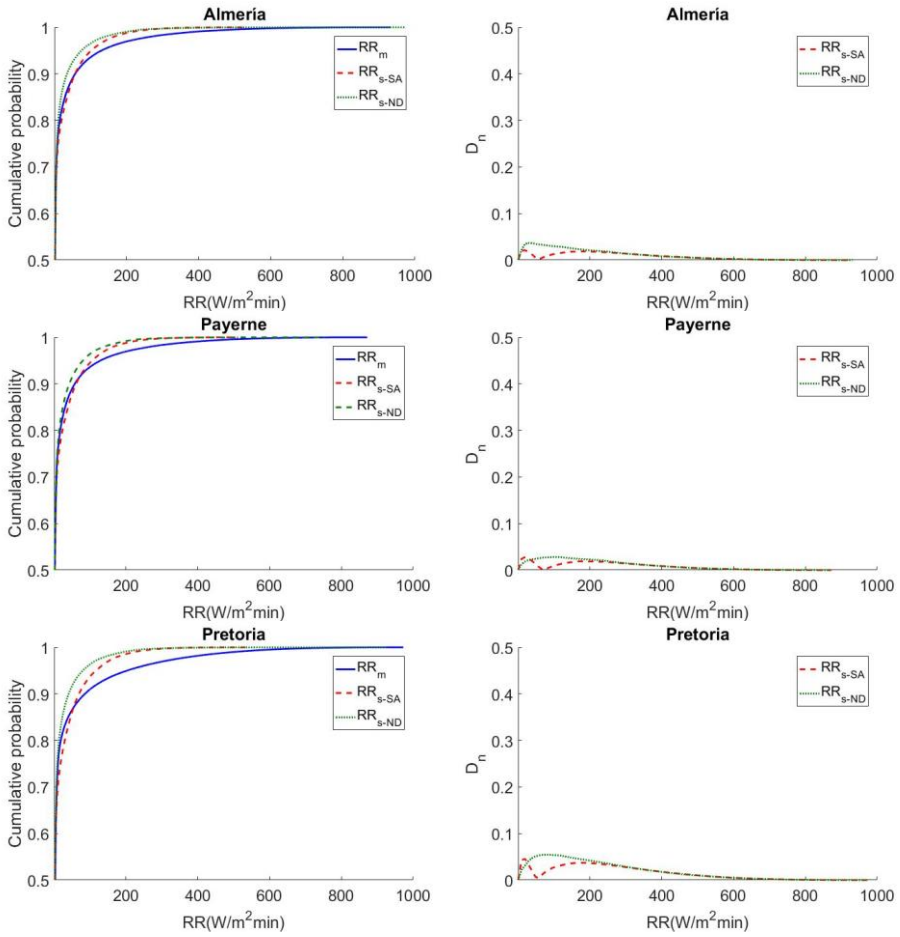


Figure 3-6. ECDFs of the measured (RRm) and synthetic absolute RR datasets generated with the stochastic adaptation model (RRs-SA) and non-dimensional model (RRs-ND) (left) and their differences (right).

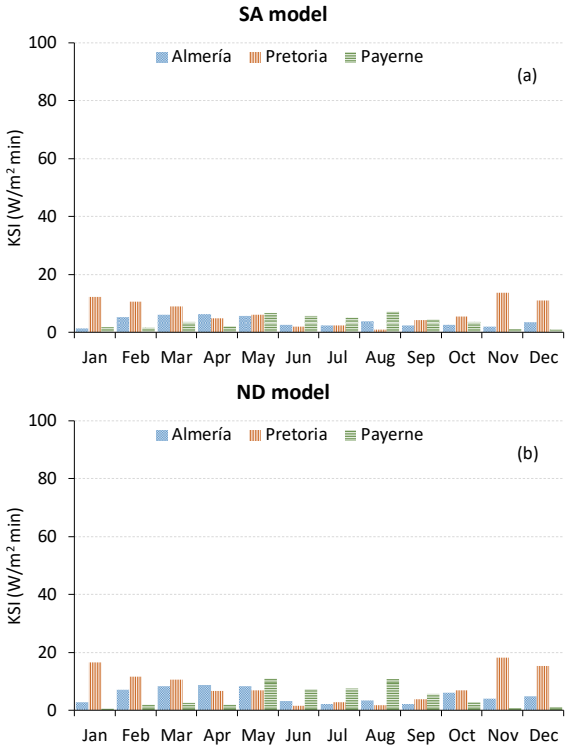


Figure 3-7. Monthly KSI values for the absolute RR time series of each model in the selected locations obtained with the SA model (a) and the ND Model (b).

4. CONCLUSIONS

In this doctoral thesis, 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.

The SA and ND models 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 have compared 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 estimation of the thermal power produced in the solar field of a typical PT plant. In this case, we observe that the SA methodology provides a lower hourly NRMSD (~0.9 %) compared with the ND methodology (~3.4 %).

These methodologies are based on previous works (Polo et al., 2011; Fernández-Peruchena et al., 2015).

The improvements to the Polo et al (2011) methodology (SA model), that have been implemented in two approaches, are focused on the characterization and reproduction of the stochastic component of the solar radiation and the identification of clear sky equivalent DNI periods.

The stochastic component of the solar radiation is reproduced based on the ECDF of a sufficiently large database from one location. This way, only real situations will be generated and there is no need of having an extensive database in the location under study. This concept can be applied to other phenomena related to the solar radiation as presented in Larrañeta et al (2018-b), where we developed a methodology to generate decomposition models that emulate the non-linear relation between the

global and direct component of the solar radiation.

The identification of the clear sky equivalent DNI periods from hourly values of DNI in (Larrañeta et al 2017-a) is based on the fit of a clear sky envelope to the hourly values day by day and then, by the comparison of the differences between the means, the slopes and the lengths of the measured and theoretical clear sky curves. Identifying clear sky equivalent DNI periods from hourly values of DNI could be useful for other applications related to the solar radiation further than the synthetic generation of solar data, such as the on-site adaptation of satellite derived data and the evaluation of solar systems performance degradation.

The modifications of the Fernández-Peruchena et al. (2015) methodology (ND model) are focused on the characterization of the daily profiles in terms of energy, variability and distribution.

The ND model presents better results in terms of frequency distribution of the 1-min DNI synthetic data than the SA model. With respect to the RRs and the NRMSD of the thermal power produced in a PT plant with a common configuration, the SA model performs better. These results suggest that the similarity in terms of frequency distribution of the synthetic solar radiation dataset with the measured dataset alone does not provide more accurate results in terms of the estimation of the power produced by a CSP plant, but taking into consideration other features like a proper characterization and reproduction of the intra-daily variability may lead to more accurate results in this terms.

Both models show satisfactory results for CSP plants performance evaluation considering that the uncertainties involved in this subject (DNI measurements uncertainties, assumptions taken in CSP modelling, etc.) are greater than the average daily NRMSD value obtained for the locations under study (0.9% for the SA model and 3.4% for the ND model). The SA model could be used when accurate hourly DNI values are available at the site. Alternatively, the ND model could be used when the cloud transients are roughly characterized at hourly scale, even if the hourly DNI series are not precise at the site. This may be the case of the evaluation of operating strategies or the optimization of storage systems.

This research shows, for the first time, the generation of 1-min DNI series in different locations without any local adaptation or calibration. The sites selected are located at different climates and latitudes, supporting the feasibility of a global applicability of the methodologies.

Future improvements are focused in the ND model since we understand it has a greater field of applicability. It can be used when daily information of the DNI profiles is available and it reproduces cloud transients that have actually happened. Future improvements of this method should focus on the following issues:

- Improvement in the characterization of the variability of the solar radiation. A possible approach is to use two indexes: one giving information about the number of fluctuations and the other giving information about the amplitude of the fluctuations.
- The model performance should also improve by training the models with data from different climates. We could generate a database of normalized profiles for each of the Köppen classification climates. When a downscale were required, we could use the normalized profiles of the same Köppen climate to generate high temporal resolution solar radiation data.
- Clear sky assessment on a daily basis: Once the most similar days to be downscaled have been selected, we should perform a daily assessment of the clear sky envelope that minimizes the differences between the available 1-h set and the synthetically generated 1-min set in terms of energy. To that end we could solve an optimization problem where the cost function to be minimized should be the difference between the cumulative daily value of the measured and synthetic set depending on the A-B parameters of the clear sky model for a given day.

REFERENCES

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., 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.

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.

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.

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

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.

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.

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., 2013. 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

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.

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.

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

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.

Kondratyev, K.Y. Radiation in the Atmosphere. Academic Press 1969, New York.

Larrañeta, M., Moreno-Tejera, S., Lillo-Bravo, I., Silva-Pérez, M.A. 2018 (a). A methodology for the stochastic generation of hourly synthetic direct normal irradiation time series. Theorical and Applied Climatology 131, 3-4 1517-1527.

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

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). 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.

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.

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

M. Martínez-Chico, F.J. Batlles, J.L. Bosch., 2011. Cloud classification in a Mediterranean location using radiation data and sky images. Energy 36, 4055-4062.

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

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).

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.

Moreno-Tejera, S., Silva-Pérez, M.A., Ramírez-Santigosa, L., Lillo-Bravo, I. 2017. Classification of days according to DNI profiles using clustering techniques. Solar Energy, 146, pp. 319-333. Cited 2 times.

Moreno-Tejera, S., Silva-Pérez, M.A., Lillo-Bravo, I., Ramírez-Santigosa, L. 2016.

Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions. Solar Energy, 132, pp. 430-441. Cited 12 times.

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.

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.

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

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.

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.

Reno, M. J., Hansen, C.W. 2016. Identification of periods of clear sky irradiance in

time series of GHI measurements. Renewable Energy 90, pp. 520 – 531.

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. Estimación del recurso solar para sistemas termosolares de concentración. Doctoral thesis 2002. University of Seville.

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.

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.

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

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

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.

ANNEX A

ANNEX B

ANNEX C

ANNEX D

ANNEX E