Trabajo Fin de Máster

Máster en Sistemas de Energía Térmica

Synthetic generation of Plausible Solar Years for long-term forecasting of solar radiation.

Generación sintética de Años Solares Plausibles para la predicción a largo plazo de la radiación solar.

Autor: Paola Jiménez Valero

Tutores: Miguel Larrañeta Gómez-Caminero

Sara Moreno Tejera

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

Sevilla, 2021

Trabajo Fin de Máster
Máster en Sistemas de Energía Térmica

Synthetic generation of Plausible Solar Years for long-term forecasting of solar radiation.

Autor:
Paola Jiménez Valero

Tutores:
Miguel Larrañeta Gómez-Caminero
Sara Moreno Tejera

Dpto. de Ingeniería Energética
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla
Sevilla, 2021

Trabajo Fin de Máster: Synthetic generation of Plausible Solar Years for long-term forecasting of solar radiation.

Autor: Paola Jiménez Valero

Tutor: Miguel Larrañeta Gómez-Caminero

Sara Moreno Tejera

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2021

El Secretario del Tribunal

A mi familia

A mis amigos

A mis tutores

Resumen

En este trabajo se mejora y se valida un algoritmo propuesto en estudios previos para generar series sintéticas de años solares plausibles (PSY). La metodología proporciona 100 años sintéticos en paso minutal de radiación global horizontal (GHI) y radiación directa normal (DNI) acoplados. El algoritmo requiere un conjunto de datos de 10 a 20 años de GHI y DNI acoplados en paso horario que se pueden obtener para la mayoría de las ubicaciones del mundo a partir de estimaciones de satélite. El algoritmo se evalúa en cinco emplazamientos seleccionados por su diferente tipo de clima según la clasificación de Koppen-Geiger. La evaluación se realiza en diferentes escalas temporales: anual, mensual, diaria y minutal. En todos los casos, las series sintéticas de años solares plausibles cubren una gama más amplia de escenarios que las series observadas, pero manteniendo su distribución. Los resultados sugieren que los PSY generados sintéticamente son capaces de reproducir la variabilidad natural del recurso solar en cualquier emplazamiento facilitando la simulación estocástica de los sistemas de aprovechamiento solar.

Abstract

In this work we test and improve an algorithm proposed in previous studies to generate synthetic series of plausible solar years (PSY). The method provides 100 synthetic years coupled 1-min global horizontal irradiance (GHI) and direct normal solar irradiance (DNI). The algorithm requires 10-20 years of hourly coupled GHI and DNI datasets that can be retrieved for most of the locations of the world from satellite estimates. The algorithm is evaluated at five s locations selected because of their different type of climate according to the Koppen-Geiger classification and at different temporal scales: annual, monthly, daily, and 1-minute resolution. In all cases, synthetic PSYs series cover a wider range of scenarios than the observed series but maintaining their distribution. Results suggest that the synthetically generated PSYs are capable to reproduce the natural variability of the solar resource at any location facilitating the stochastic simulation of solar harnessing systems.

Index

INDEX OF TABLES

INDEX OF FIGURES

Notation

CAR_{GHI}	Autoregressive coefficient of GHI
$CARDN_I$	Autoregressive coefficient of DNI
DNI_o^D	Daily observed direct normal irradiance
DNI_s^D	Daily synthetic direct normal irradiance
DNI_o^H	Hourly observed direct normal irradiance
DNI_o^M	Monthly observed direct normal irradiance
DNI_s^M	Monthly synthetic direct normal irradiance
DNI_s^{1min}	1 min synthetic direct normal irradiance
DS_{GHI}^D	Daily smoothed standard deviation of GHI
$DSDNI^D$	Daily smoothed standard deviation of DNI
F_{GHI}^D	Daily Fourier series for GHI
F_{DNI}^D	Daily Fourier series for DNI
Fm_o^D	Daily observed morning fraction index
Fm_s^D	Daily synthetic morning fraction index
GHI_o^D	Daily observed global horizontal irradiance
GHI_s^D	Daily synthetic global horizontal irradiance
GHI_o^H	Hourly observed global horizontal irradiance
GHI_o^M	Monthly observed global horizontal irradiance
GHI_s^M	Monthly synthetic global horizontal irradiance
GHI_s^{1min}	1min synthetic global horizontal irradiance
Kb_o^D	Daily observed transmittance index
Kb_s^D	Daily synthetic transmittance index
Kt_o^D	Daily observed clearness index
Kt_s^D	Daily synthetic clearness index

VI_o^D Daily observed variability index

VI_s^D Daily synthetic variability index

1 INTRODUCTION

The design and development of solar energy systems requires an approach of the performance of the plant and its economic viability under normal operating conditions. For this purpose, time series of solar irradiance and meteorological variables are used to represent normal operation conditions. In general, these data sets are based on the definition of representative annual sets [1] suitable for the detailed performance evaluation of solar systems. Currently, most of the methods used to represent normal operating conditions in simulation systems are based on the selection of the most representative year. It is a 1- year meteorological set representative of "typical" conditions, defined as Typical Meteorological Year (TMY) (Hall et al. 1978) which represents a year of typical climatic conditions for a given location. The TMY has been modified from its initial concept to respond to the needs of other technologies. Several modifications to the TMY have been proposed over the years, but the initial procedure has not changed significantly[2]. The main weak point of the TMY method is that a single representative year does not represent extreme conditions [3] and does not allow characterizing the variability of the solar resource in detail. This is a problem since the lifetime of these plants are 25 to 30 years and a typical year does not represent all possible futures scenarios.

Given the solar power plants have a considerable cost, it is important to know the energy that a certain system will generate depending on its location [4]. Therefore, the characterization of the solar resource in a specific location is key to energy policy decisions, engineering design and development of solar energy systems The use of representative series of solar irradiation and meteorological variables [5] is not sufficient because these series of averaged values provide an incomplete characterization of the climate. To characterize the solar resource a detailed study of frequency distribution and variability at different time scales is required [6].

The uncertainty related to the variability of the solar resource plays an important role in probability estimation of the future energy performance of a solar power plant and in its financial conditions [7]. Developers and lenders need information related to the annual amortization of the project in extremes years, in addition to the annual TMY series [8]. For this reason, TMY may not be suitable enough for designing systems that meet the worst climatic conditions that could occur in a location. It has been observed that the magnitude of the observed variability changes clearly depends on the geographical area. The evaluation of long-term space-time variability can have substantial impacts on the design, financial capacity and operation of solar energy conversion systems that use technologies such as photovoltaics or CSP. Moreover, the study of the variability of the solar resource in different types of climates provides valuable information on the useful life of solar materials in different locations and on the selection of the site during the pre-feasibility phase of solar projects that use these technologies [9].

The solar resource assessment of Model or measured solar resource data are important for understanding long-term solar variability for a region or location. Solar radiation data sets must be accurate and reliable to minimize the risk related to the uncertainty of available data sets. These data series enable resource planning, improve production cost modeling, as well as support the solar industry by reducing solar project development costs, thus the ideal is to have a long serie of radiation data measured with precision in the location [10] but this is not the usual. The existence and availability of time series of solar radiation derived from satellites allows understanding the dynamics of solar radiation on different time scales to model project capacity and reduce financial risk [7].

At the present time, it is becoming increasingly important the probabilistic approach to model the variability and uncertainties of the solar energy system. These probabilistic methods allow modelling the changes and evaluate likelihoods of the simulated metric's probability of lying above or below a particular value or within a specific range [11].In this scenario, it takes importance the generation of synthetic solar datasets in order to extend the information related to the variability of the solar radiation. The advantage of the synthetic data series is that allow to represent the behavior of solar radiation in a great range of scenarios and facilitates the risk evaluation of solar projects [12].

For this reason, the generation of Plausible Meteorological Year becomes more important. The Plausible Meteorological Years (PMYs) are defined as high-frequency yearly series of Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI) and other relevant meteorological variables (temperature, relative humidity, and wind speed, among other), which preserve the natural variability of the resource in a specific location [13]. The plausible solar years (PSYs) is a simplification of PMYs where only solar data are generated.

The methods for the synthetic generation PMYs use large datasets that could be provided by satellite data providers for most of the world locations. These series take into account the observed variability of the solar resource at the site, assuming the future trends may be characterized by past trends. PMYs generation allows the stochastic simulation of a solar harnessing system considering the uncertainty and variability inherent of the solar resource [14].

The PMYs have great advantages over TMY, since these datasets consider a wider range of solar radiation scenarios preserving its inherent annual, monthly, daily and intra-daily variability [13]. Using these PMYs as input in the performance evaluation of solar systems allows obtaining a probabilistic description of the system performance and a valuable information for the project finance phase.

Larrañeta et al., (2019) present a model for the synthetic generation of high-resolution plausible solar years. The algorithm is suitable for any location where satellite derived datasets are available. The synthetic generation process has been automatized and an open tool has been recently developed [15]. The methodology does not need any local adaptation or calibration, but a lower uncertainty in the dataset will allow to obtain better results.

In this work we improve the method presented by Larrañeta et al. (2019) and test it on 5 locations with different climates We generate 100 annual sets of coupled 1-min DNI and GHI solar data for these locations using SolarGIS satellite derived data and compare them with the behavior of the data series observed from each location. SolarGIS professional data sets provide 20 to 25 annual sets of 15-min of coupled DNI and GHI and other relevant meteorological variables. This method allows to obtain 100 PSYs that maintain the characteristics of the original series but provides a wide range of solar radiation scenarios suitable for the probabilistic description of the performance evaluation of solar systems. This method provides deeper characterization associated to the variability of solar resource, allowing to obtain improvements in the financial conditions and encouraging investment in solar plant.

2 OBJECTIVES

The first and main object of this project is the development of a daily model for the synthetic generation of long-term daily values of global horizontal solar irradiance (GHI) and direct normal solar irradiance (DNI). The model is constructed taking into account three components based on [16], a seasonal component (Fourier model), an autoregressive component and a random component.

The second objective is the implementation of the daily model into an algorithm that intends to generate 100 annual sets of one-minute coupled GHI and DNI data. The algorithm would generate first monthly, then daily and finally one-minute synthetic data automatically without any local adaptation

The last of the objectives of this work is the evaluation of the algorithm performance in five locations with different climates.

3 METHODOLOGY

In this section we describe the method used for synthetic generation of multiple annual solar radiation time series at different locations. This method requires a minimum of 10 non-consecutive annual time series of hourly-coupled DNI and GHI values as input and provides 100 annual Plausible Solar Years (PSYs) of 1-min coupled DNI and GHI dataset whose annual cumulative values correspond to the annual Probability of Exceedance from 1 to 100. Input data can be obtained from observations, satellite estimates or reanalysis irradiance databases. The size of the sample of the observations should be sufficient to statistically characterize de annual and monthly distribution of the solar radiation. For this reason, it's advisable a minimum period of 10 years, consecutive or non-consecutive, of observed data sets.

3.1 Previous methodology

The method for the synthetic generation that is update in this work consists on the conjunction of three steps (Larrañeta et al., 2019).

In the first step, we obtain 100 annual series at monthly scale. We first integrate the observed hourly data into the monthly resolution and calculate the cumulative distribution functions (CDFs) of each month. Then we apply the probability integral transform method to generate the monthly synthetic solar data. We limit the synthetic generation to maintain the relation between GHI and DNI in a given location (for more details see Larrañeta et al., 2019).

The method is applied 10000 times for each month of the year. The monthly values can be concatenated to obtain 10000 annual cumulative sets of synthetic data in a monthly basis.

The observed GHI and the DNI hourly data are also integrated into annual cumulative sums and then fitted to a Normal distribution in order to estimate the theoretical probabilities of exceedance. In this step, we make a modification to the original algorithm that assumed that the annual GHI follows a Normal distribution but the annual DNI follows a Weibull distribution. It is assumed that hourly and sub-hourly DNI follows a Weibull distribution, but it is not yet clear if the annual DNI still follows a Weibull distribution.

At the end of step 1, we seek and select the 100 closest sets (from P1 to P100) within the 10000 synthetically generated to those estimated fitting a normal distribution to the observed sets. In Figure 3-1 we present the results of the first step calculations for the location of Seville.

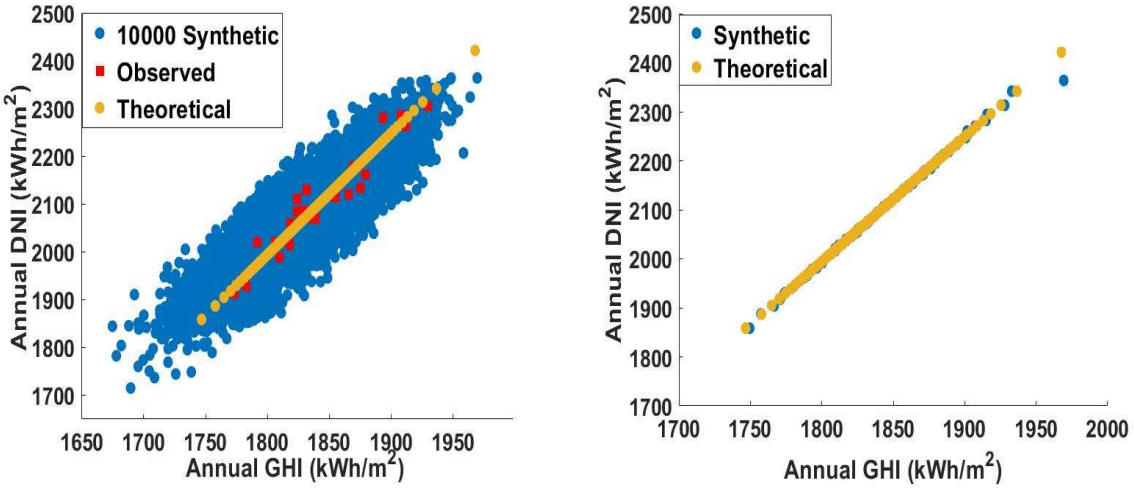


Figure 3-1 Result of the first step for the location of Seville.

In the second step, we downscale the synthetic solar irradiance sets from monthly to daily time resolution. Based on [16], the daily model consists of three components: a seasonal, an autoregressive (AR) and a random component. The development of the daily model for the synthetic generation of long-term daily values of GHI and DNI is one of the objectives of our work. This model and its implementation in the algorithm will be explained in detail in t sections 3.2 and 3.3.

In the step 3 we generate 1 min synthetic values from the synthetic daily quartets of kb, kt, VI and Fm obtained in step 2. In this step, we use the ND model tool [17] for the synthetic generation of 1-min coupled GHI and DNI data. The ND model uses non-dimensional databases from a location with similar climate to the input data; therefore, it can be used in at location without local adaptation [18]. The non-dimensional database daily profiles are labeled in terms of energy (kt, kb), variability (VI) and distribution (Fm). We generate synthetic daily quartets of kb, kt, VI and Fm in step 2 that are used to seek and find the most similar day among the dimensionless databases. The selected dimensionless profiles are "opened" into real values by multiplying the 1-min kt and kb values to the extraterrestrial irradiance and the clear sky DNI of each corresponding day.

In Table 1 we present the dimensionless databases implemented in the algorithm. We have selected seven locations with different climates. We use the 1-min observed GHI and DNI data obtained from the Australian Bureau of Meteorology website [19] for six locations and the 1-min observed GHI and DNI datasets from GTER meteorological database [20] for the location of Seville, Spain.

Table 3-1 Dimensionless databases available sites and their Köppen-Geiger classification climatecation

Location (ID)	**Country**	**Latitude (ºN)**	**Longitude (ºE)**	**Altitude (m)**	**Köppen-Geiger classification climate**
Sevilla	Spain	37.22	5.58	16	Csa
Darwin	Australia	-12.27	130.53	17	Aw
Broome	Australia	-17.57	122.14	12	Bsh
Alice Springs	Australia	-23.41	133.53	583	Bwh
Rockhampton	Australia	-23.37	150.51	20	Cfa
Melbourne	Australia	-37.48	144.57	25	Cfb
Adelaide	Australia	-34.55	138.35	59	Csb

3.2 Daily model

In this section we explain the method for the synthetic generation of long-term daily values of GHI and DNI. This method will be implemented into an algorithm to generate 100 annual sets of one-minute coupled GHI and DNI data. First, we will explain the development of the daily model and later the procedure for the synthetic generation of daily values of GHI and DNI with this model.

3.2.1 Development of daily model

The daily model that is constructed in this section is based on [16] The model consists of three components: a seasonal (F_t), an autoregressive component (AR) and a random component (Z_t). The calculation of each component of the model is performed following the mathematical model proposed by Boland [21]

Whith the aim of clarifying the development of the daily model, figures presented in this section correspond to an annual set of daily data for the location of Seville (2002) and we follow the same procedure for GHI and DNI.

However, for the synthetic generation we use a database compose of daily values of GHI and DNI recorded during 14 consecutive years (2002-2015) for Seville (Spain) This database is provide by the group of Thermodynamics and Renewable Energy of the University of Seville (GTER).

First, we estimate the Fourier component, that correspond to the seasonal component to the model (F_t). Based on previous studies (Boland) the series of daily radiation values has significant cycles that are identified with the Fourier Transforms, Boland identified two significant frequencies: once a year and twice a year. Therefore, for calculate the seasonal component we use a sum of Fourier terms with two frequencies. The Fourier model is given as:

$$F_t = \alpha_0 + \alpha_1 x cos\frac{2\pi t}{365} + \beta_1 x sin\frac{2\pi t}{365} + \alpha_2 x cos\frac{4\pi t}{365} + \beta_2 x sin\frac{4\pi t}{365} \quad (3\text{-}1)$$

Where α_0 is the mean, $\alpha_1$and β_1, are the coefficients of the once-a year cycle, and α_2 and β_2 are the coefficients of the twice a year cycle. We can see in Figure 3-2 for GHI and Figure 3-3 for DNI how well the curve fits the general trend of the data for both GHI and DNI.

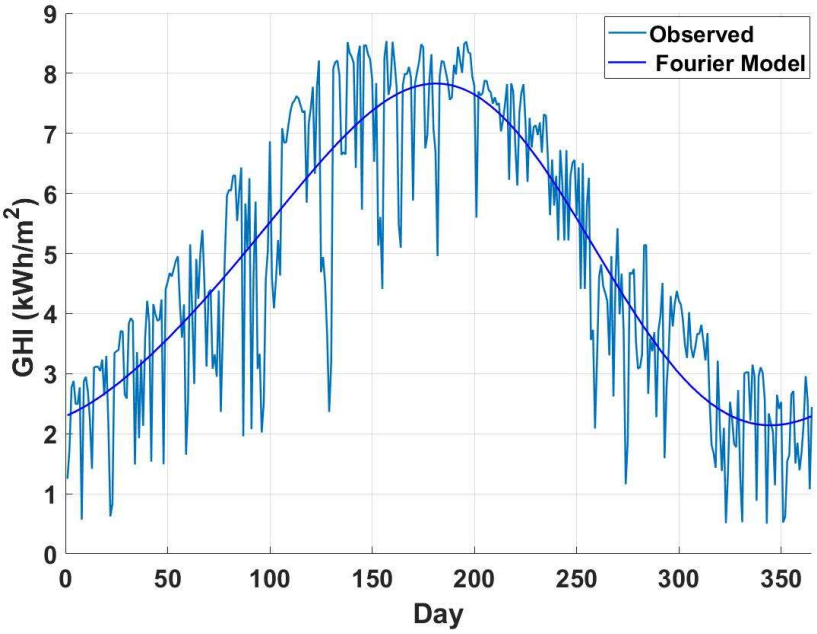


Figure 3-2 Daily GHI data setand Fourier model.

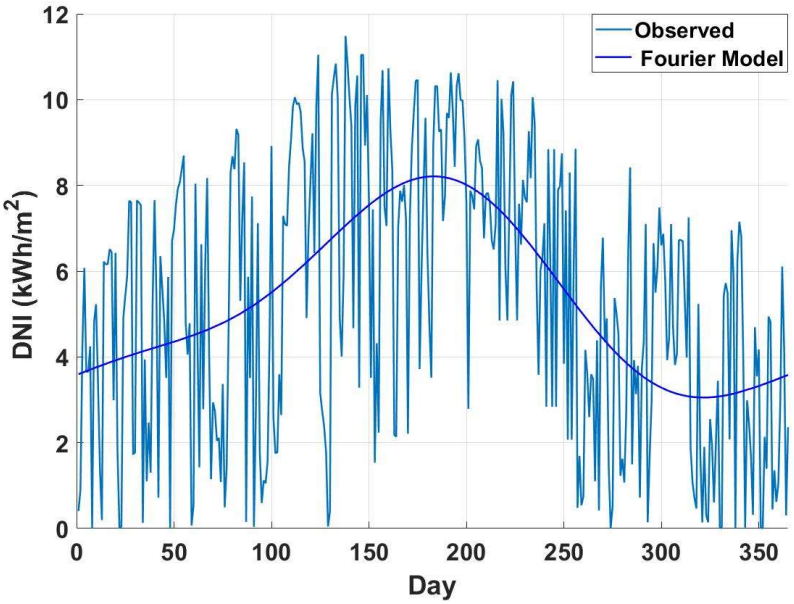


Figure 3-3 Daily DNI data set and Fourier model.

Then we subtract the trend from the data. We can see in Figure 3-4 and Figure 3-5 that the residuals are not stationary, there is a great difference between the summer and winter.

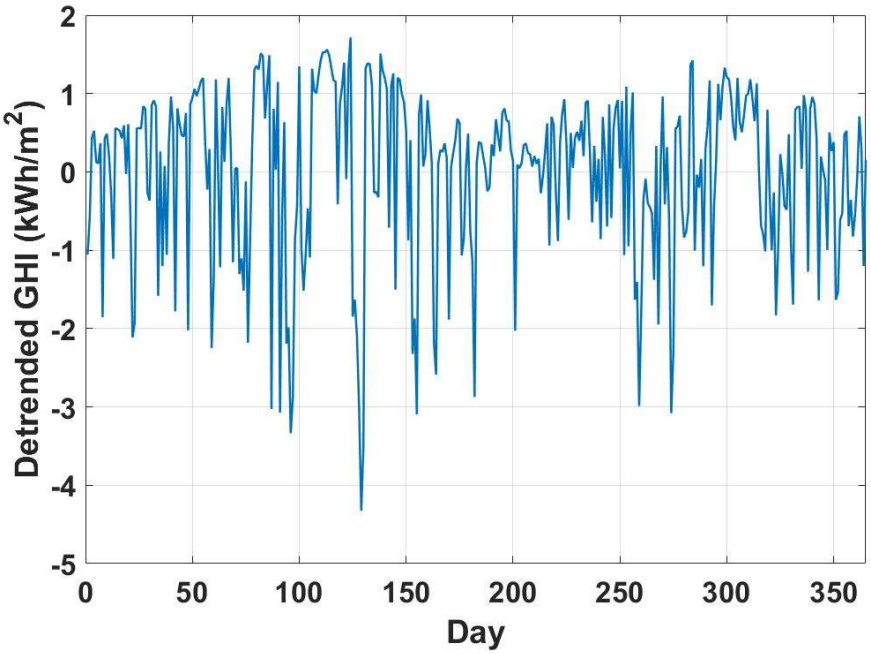


Figure 3-4 Daily GHI data set with Fourier function removed.

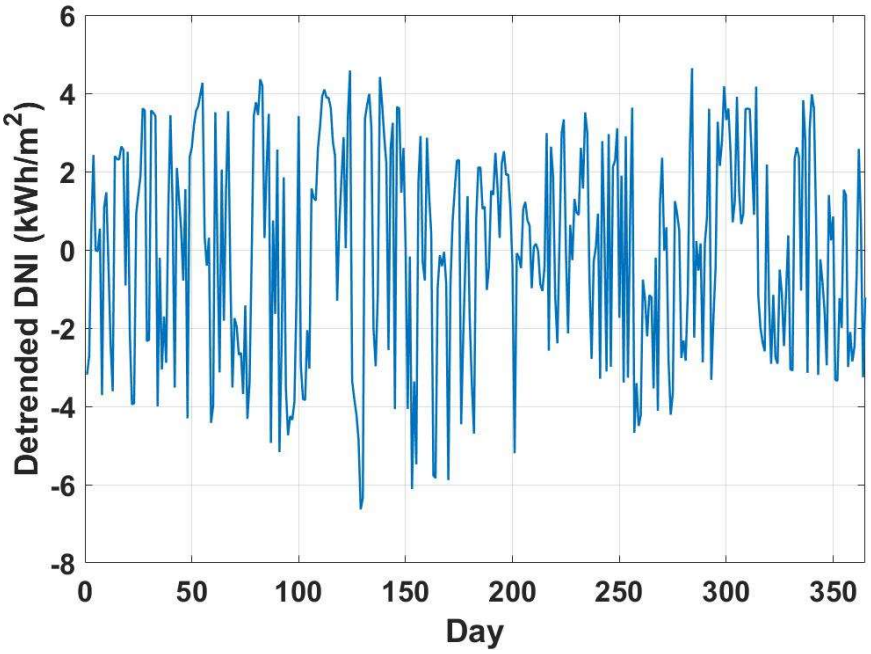


Figure 3-5 Daily DNI data set with Fourier function removed.

With the objective of standardize the series, we divide by the standard deviation to obtain a set that we can assume that it is stationary data (Figure 3-6 and Figure 3-7)

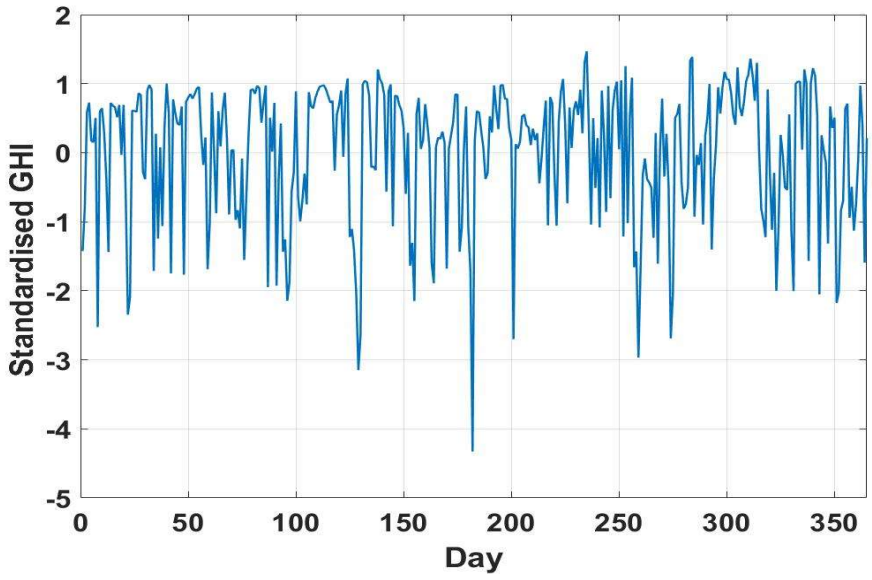


Figure 3-6 Standardised residuals of daily GHI.

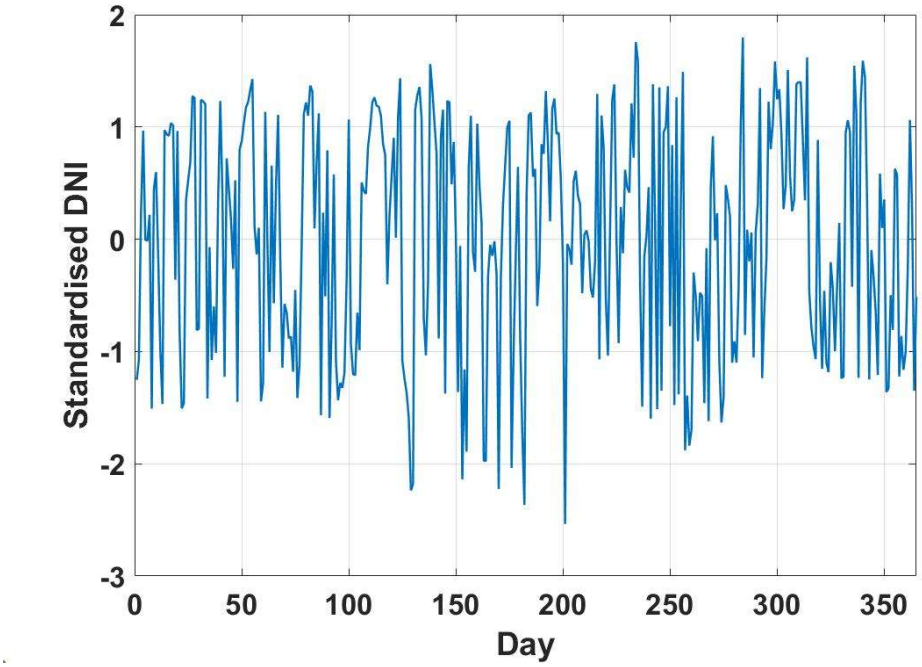


Figure 3-7 Standardised residuals of daily DNI.

Once the trend is removed and assuming that the serie is stationary, we can study the correlations of the sets to prove if it is an ARMA model. We use the Econometric Modeler application to analyze the correlations, this app is an interactive tool of Matlab for analyze time series data. In Figure 3-8 and Figure 3-9 we can see the Sample Autocorrelation Function (SACF) and Sample Partial Autocorrelation Function (SPACF) for daily standardize residuals for GHI and DNI. We can observe in both cases that in SACF the lags decrease gradually and in the SPACF there is only a significate lag. This indicate that the data can be modeled with an AR (1) process and consequently, each day of the series depend on the previous day.

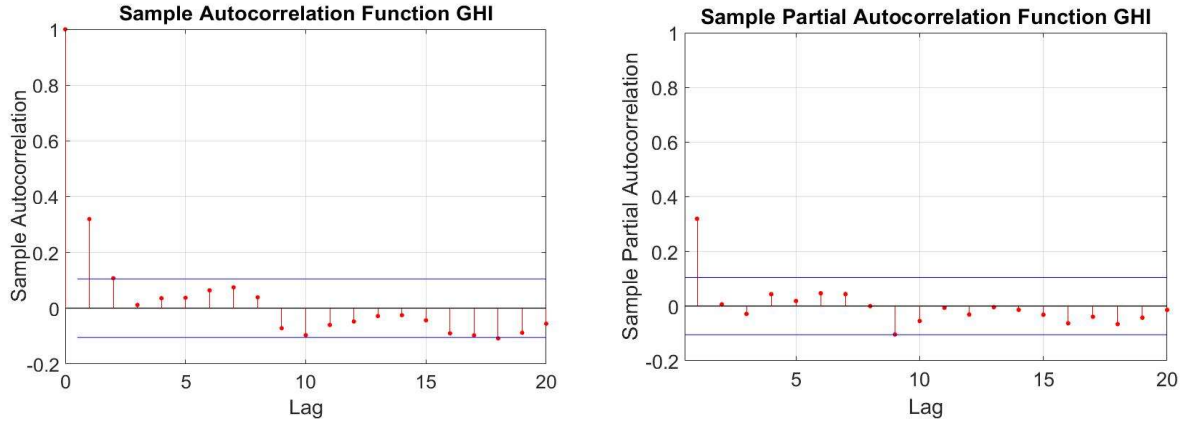


Figure 3-8 SACF (left) and SPACF (right) for the daily standardised residuals of GHI.

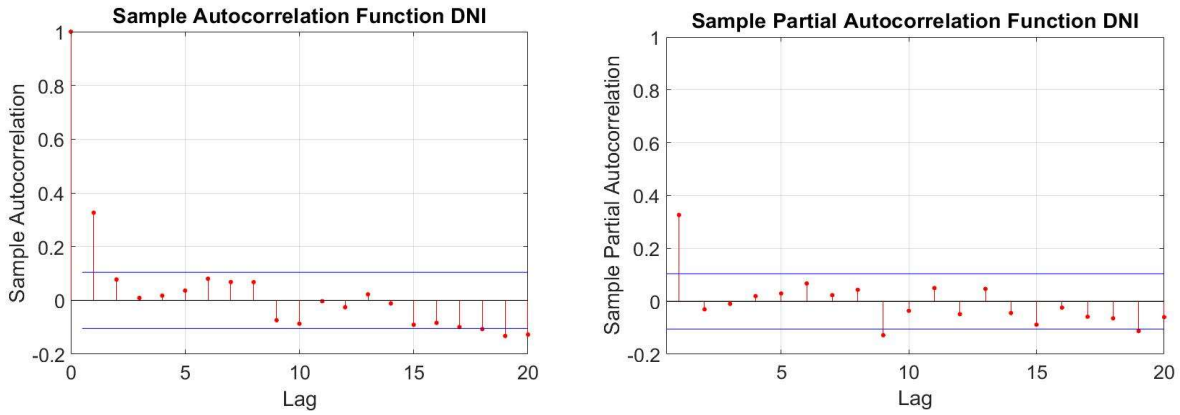


Figure 3-9 SACF (left) and SPACF (right) for the daily standardised residuals of DNI.

To verify the hypothesis, we overfit the model with an AR(2) and we obtain the coefficients and the value of the p parameter of both adjustments. Table 3-2 show the value of the parameters for GHI and Table 3-3 for DNI. The p value indicates that the only significant coefficient is for an AR(1)in both cases being the AR(1) coefficient 0.319 for GHI and 0.326 for DNI for the studied year in Seville.

Table 3-2 Final Estimates of parameters for the overfitting exercise for an AR (1) in the case of GHI.

Parameter	**Value**	**Standard Error**	**t Statistic**	**P-Value**
Constant	0.0035	0.0631	0.055	0.9561
AR{1}	0.319	0.0498	6.4396	1.20E-10
AR{2}	0.007	0.0576	0.1223	0.9027

Table 3-3 Final Estimates of parameters for the overfitting exercise for an AR (1) in the case of DNI.

Parameter	**Value**	**Standard Error**	**t Statistic**	**P-Value**
Constant	0.0069	0.134	0.0518	0.9586
AR{1}	0.326	0.052	6.5417	6.08E-11
AR 2}	-0.0292	0.0516	-0.566	0.5714

The hypothesis has been tested with Ljung-Box test (REF) and with the study of the SACF of the white noise series. We obtain the white noise series subtract the AR (1) model to the standardize series, this series will be the random component of the daily model (Z_t). In Figure 3-10 we can see the white noise series of GHI and the SACF in Figure 3-11for GHI In Figure 3-12 and Figure 3-13 we show the white noise series for DNI and the SACF in this case in both figures we can verify that the lags in SACF are not significant, this indicate that the model is appropriate.

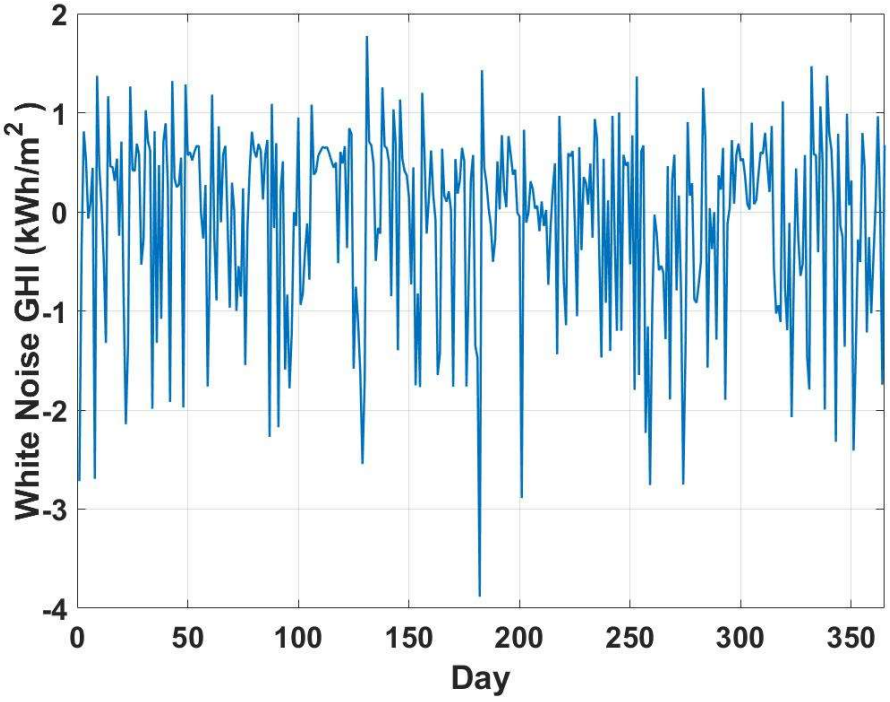


Figure 3-10 White noise of GHI serie

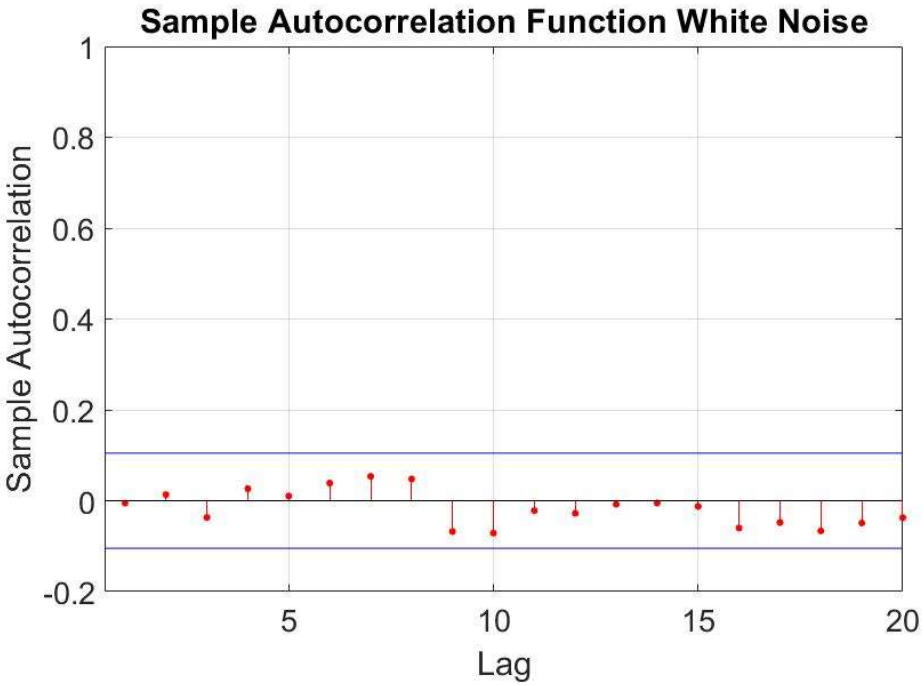


Figure 3-11 Autocorrelation of the white noise series for GHI.

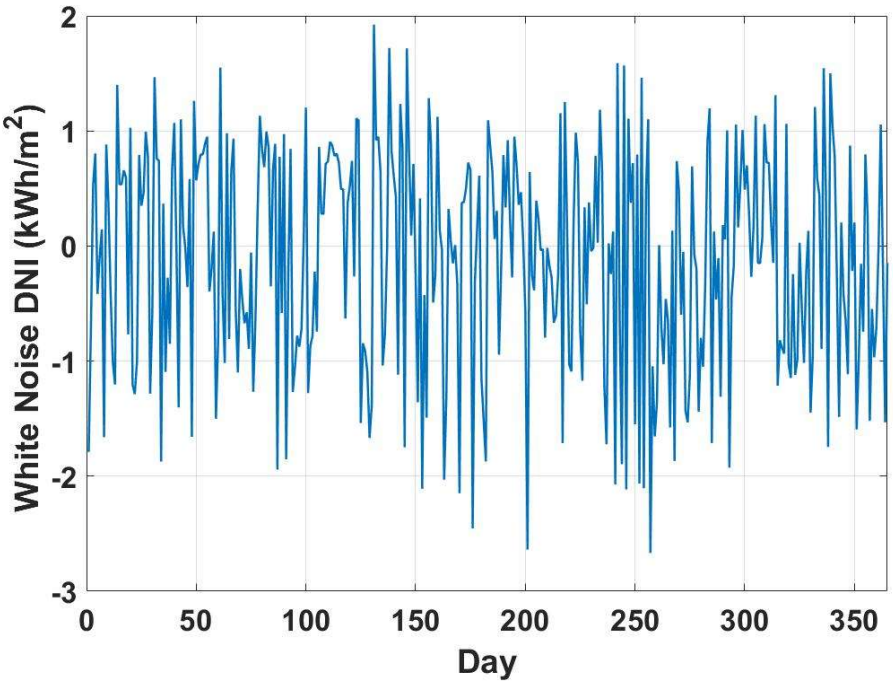


Figure 3-12 White noise of DNI serie

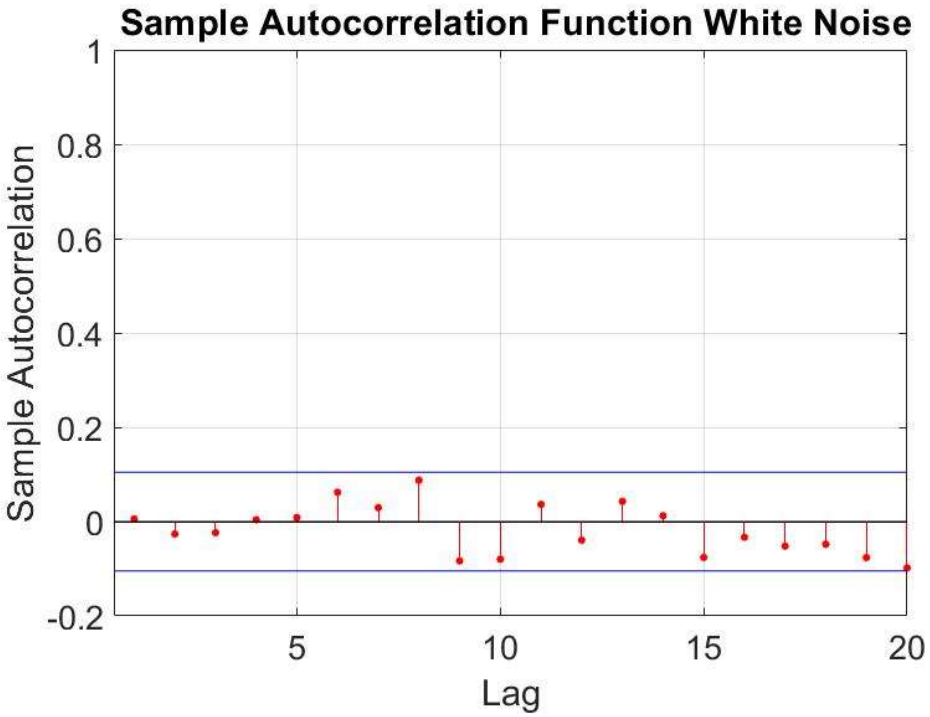


Figure 3-13 Autocorrelationof the white noise series of DNI.

Once explained how each component of the model is estimated, the daily model (DM) would be defined as follows:

$$DM = F_t + AR + Z_t \quad (3\text{-}2)$$

Where F_t is the component that represents the seasonality of the series, AR is the autoregressive component that makes each day dependent on the previous day and finally, Z_t is the white noise, that represets the random component of the model.

The same methodology is applied to both components, GHI and DNI, of the complete series (14 years). We estimate the seasonal component from the Fourier model for the complete series and then we standardize the series subtracting the Fourier series to the observed data and dividing by the standard deviation of the annual set (the standard deviation of each day of the year). Later we construct the AR(1) model for the standardized series, to this end we calculate the AR(1) coefficient for each year and we finally select the average of all the coefficients as the final coefficient. In this case, the AR(1) coefficient is 0.378 and 0.45 for GHI and DNI respectively for the location of Seville. Finally, the random component or white noise is calculated by subtracting the AR(1) model to the standardized series. Table 3-4 shows the AR coefficient for each year and the average of all the coefficients. We select this value as the final coefficients.

Table 3-4 AR(1) coefficients for GHI and DNI for each year and average of all the coeffcients.

	AR coefficient	
Year	GHI	DNI
1	0.319	0.326
2	0.330	0.442
3	0.390	0.518
4	0.360	0.423
5	0.369	0.402
6	0.411	0.494
7	0.454	0.474
8	0.358	0.422
9	0.251	0.369
10	0.480	0.569
11	0.476	0.537
12	0.323	0.438
13	0.392	0.430
14	0.391	0.473
Φ	0.378	0.451

Once obtained the Fourier series, the AR(1) coefficient and the white noise series, we can carry out the synthetic generation procedure of daily values of GHI and DNI.

3.2.2 Synthetic generation of the daily values of GHI and DNI

The synthetic generation of the daily values is carried out with the inverse procedure to the calculation of the components of the model. So, we first apply a bootstrapping technique to generate the synthetic white noise series. This technique uses random numbers assumed to be probabilities and calculate the estimated value of that probability from the daily observed white noise CDFs of each month. We maintain the relation between the GHI and the DNI by using the same random number for each couple of daily solar radiation values.

Then we calculate the autoregressive component. The AR component depends on the previous day solar radiation value. We calculate this component by multiplying the previous value of a given day to the estimated AR(1) coefficient for a given location and adding a selected value from the white noise series (WN). For the

autoregressive series the first component is estimated by subtracting to the observed value the Fourier component corresponding to this day and dividing by the standard deviation. The following components of the autoregressive series are constructed in the same way as shown in Eq. 3-3,

$$AR(t) = \Phi AR(t-1) + WN(t) \quad (3\text{-}3)$$

Where t is the day of the year. In this way, the first day of each year depends on the last day of the previous year.

In the last step, we procure the seasonality to the synthetic set. We add both random and autoregressive components for each day and then undo the standardization. To that end, we multiply each day by its corresponding standard deviation value and finally add the contribution from the Fourier model.

$$X(t) = Ft + St(\Phi AR(t-1) + WN(t)) \quad (3\text{-}4)$$

Where $X(t)$ is the synthetic generated day, St is the standard deviation and Ft is de Fourier component. This process is continued to generate a sequence that will make up a synthetic month of daily values. We can see in Figure 3-14 for GHI and Figure 3-15 for DNI that we obtain synthetic series that present the same general characteristic as the observed series.

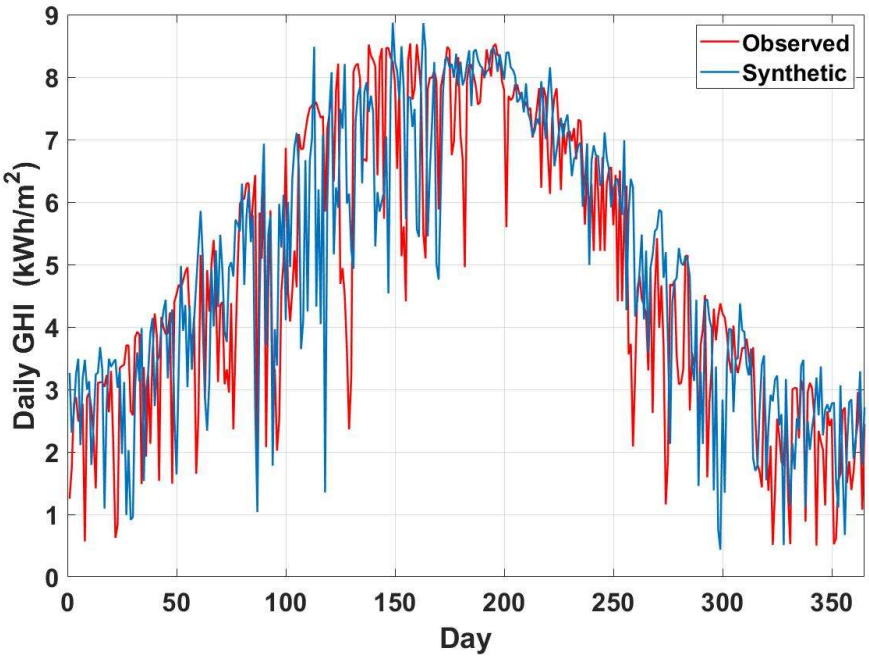


Figure 3-14 Observed and synthetic daily GHI data sets for the year2002

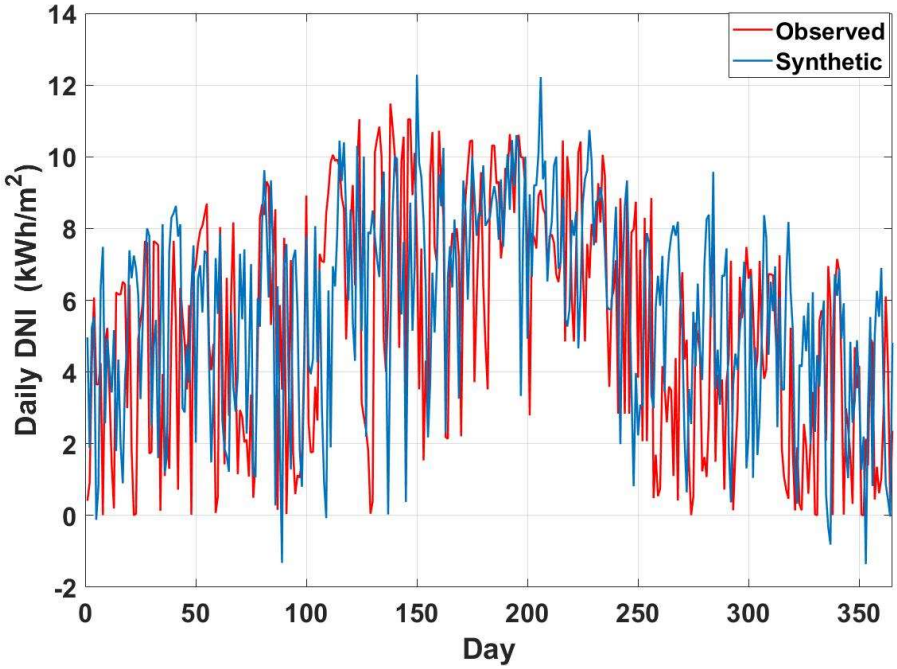


Figure 3-15 Observed and synthetic daily GHI data sets for the year 2002

Finally, for each synthetic day, we search the observed day with the closest value in terms of energy (k_t and k_b) and allocate it the same information in terms of variability and distribution by means of the variability index (VI) and morning fraction index (Fm) [22]

3.3 Implementation of the daily model into the complete algorithm

In this section we will explain the implementation of the daily model into the complete algorithm for the syntetic generation of PSYs [23]. This methodology for the generation of synthetic series of multiple solar radiation time series follows the procedure shown in Figure 3-16.

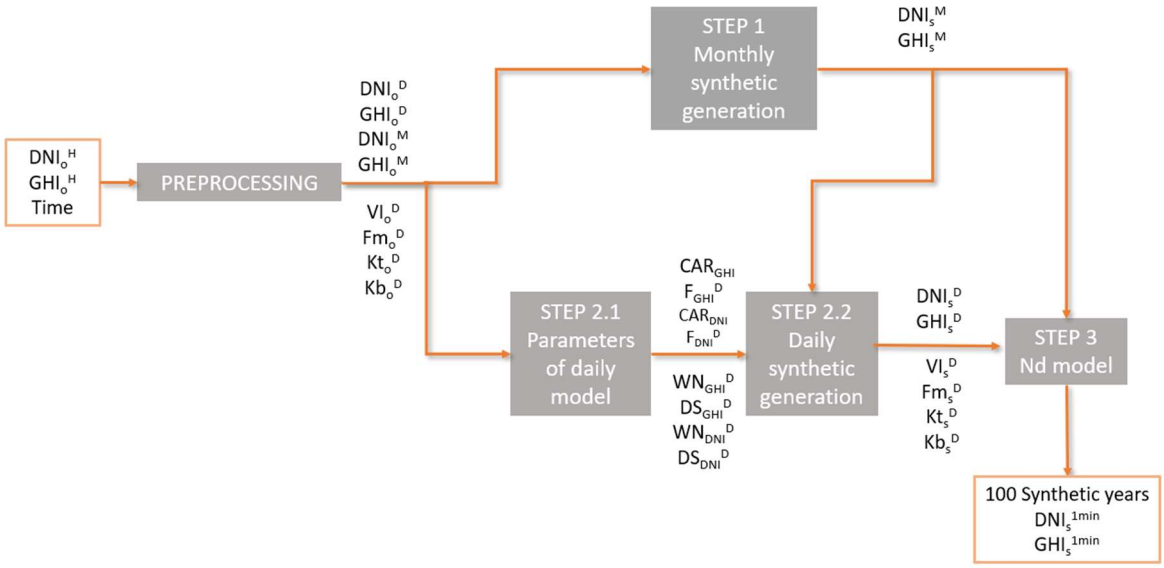


Figure 3-16 Diagram of the procedure implemented in the complete algorithm

The algorithm requires a minimum of 10 non-consecutive annual time series of hourly-coupled DNI and GHI values. Before the synthetic generation, the algorithm performs a preprocessing of the input data set to validate the GHI, DNI and data time to avoid errors with time reference. This is a common source of error when dealing with hourly solar radiation data sets. The output of the preprocessing is the monthly and daily GHI and DNI observed radiation data in addition to the k_t, k_b, VI and Fm indexes calculated from observed data.

After processing, the algorithm generates in the first step monthly synthetic data from the observed monthly GHI and DNI data. Then, we implemented the daily model in the second step. This step generates daily synthetic values from the observed data and from the synthetic monthly values obtained in step 1. To this end, we follow the procedure explained in section 3.2. We obtain the coefficients of the daily model in the step 2.1 ($CAR_{GHI}, CAR_{DNI}, F_{GHI}$ and F_{DNI}), the white noise series and the annual standard deviation for GHI and DNI. In the step 2.2 we performed the procedure of synthetic generation and we run this procedure for the generation of 1000 annual sets of daily values. We aggregate the daily synthetic sets into monthly cumulates to select those months than more closely match the ones obtained in step 1. Finally, in the step 3 we use the synthetic daily quartets of k_t, k_b, VI and Fm obtained in step 2 to generate 1 min synthetic values. The final output of the algorithm is the 100 syntehtic years 1-min GHI+DNI.

4 RESULTS AND DISCUSSION

In this section, we evaluate the performance of the algorithm in five locations with different climates. The results are evaluated in each step of the algorithm: annual, monthly, daily, and 1-min resolution. We compare observed data to synthetic PSYs and to the solar representative year (SRY) for each location. SRY is a simplification of TMY that uses only solar data for the calculation of the most representative long term annual set. SRYs have been calculated following AENOR methodology [24]

4.1 Input database

We use extensive databases of satellite-derived data for five locations as input of the multiyear algorithm for the results evaluation. The locations have been selected with the aim of covering most of the significate climates for solar harnessing systems. We have used Solargis climData professional time series that consist on 15-min time series of coupled solar and meteorological data for a period of 20-25 consecutive years. Solargis calculate solar resource data by a suit of solar models receiving as input data derived from geostationary meteorological satellites and global meteorological models. All Solargis parameters are validated by quality-controlled ground measurements acquired by high-accuracy meteorological equipment worldwide. In Table 4-1 we present the main geographical and climatological information of the selected sites.

Table 4-1 Characterization of selected sites, period of years available and their Köppen-Geiger classification climate.

Location (ID)	Country	Latitude (ºN)	Longitude (ºE)	Altitude (m)	Years	Köppen classification climate
Brasilia (BRB)	Brazil	-15.60	-47.71	1023	1999-2020	Aw
Boulder (BOU)	United States	40.13	-105.24	1689	1999-2020	Bsk
Tamanrasset (TAM)	Algeria	22.79	5.53	1385	1994-2020	Bwh
Goodwin Creek (GCR)	United States	34.25	-89.87	98	1999-2020	Cfa
Toravere (TOR)	Estonia	58.26	26.46	70	1994-2020	Dfb

4.2 Annual evaluation

In this section, we evaluate the performance of the algorithm in the annual resolution. We compare the 100 synthetic annual values to the 20-25 observed annual values for GHI (left) and DNI (right) in a boxplot. We have included the SRY annual cumulative value as a green dot.

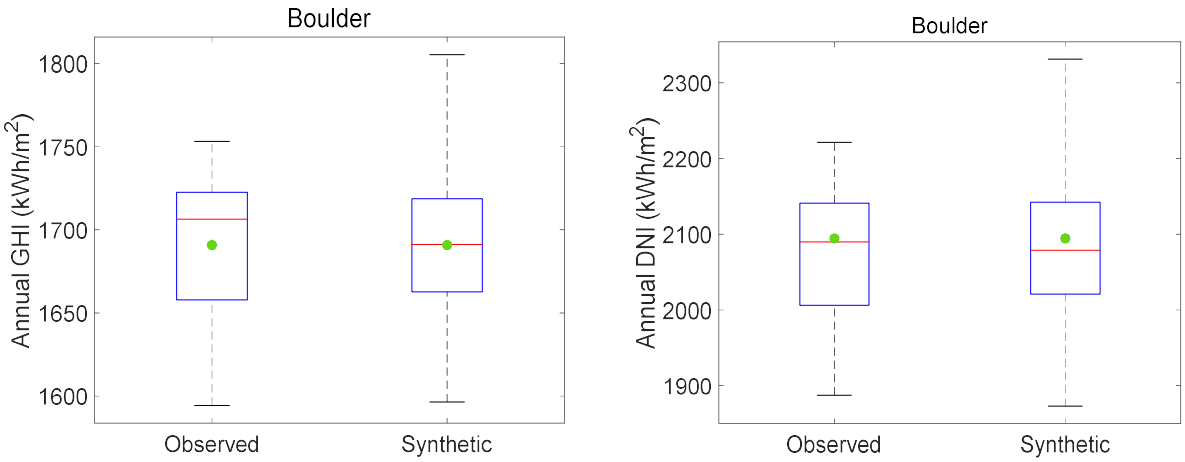


Figure 4-1 Boxplot of the observed and synthetic annual values for GHI (left) and DNI (right) with SRY in a green dot at Boulder.

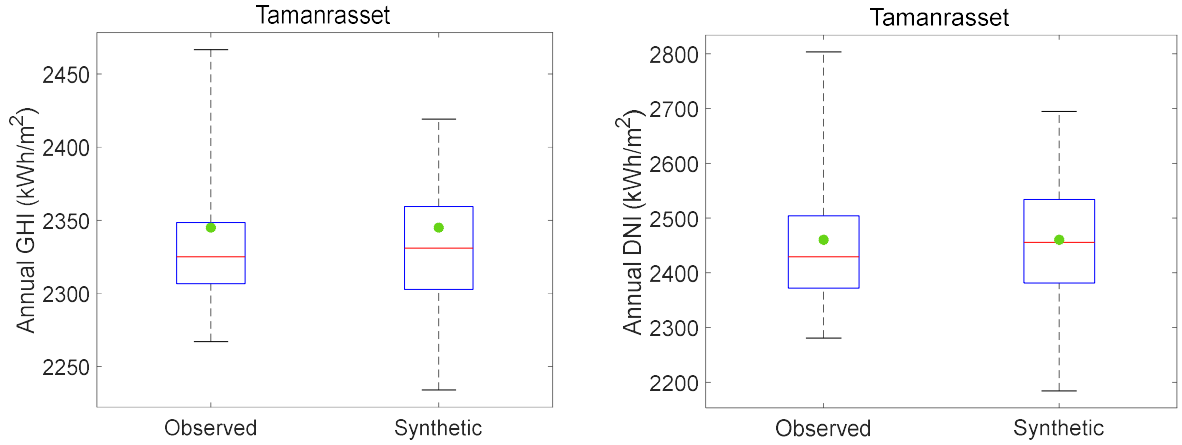


Figure 4-2 Boxplot of the observed and synthetic annual values for GHI (left) and DNI (right) with SRY in a green dot atTamanrasset.

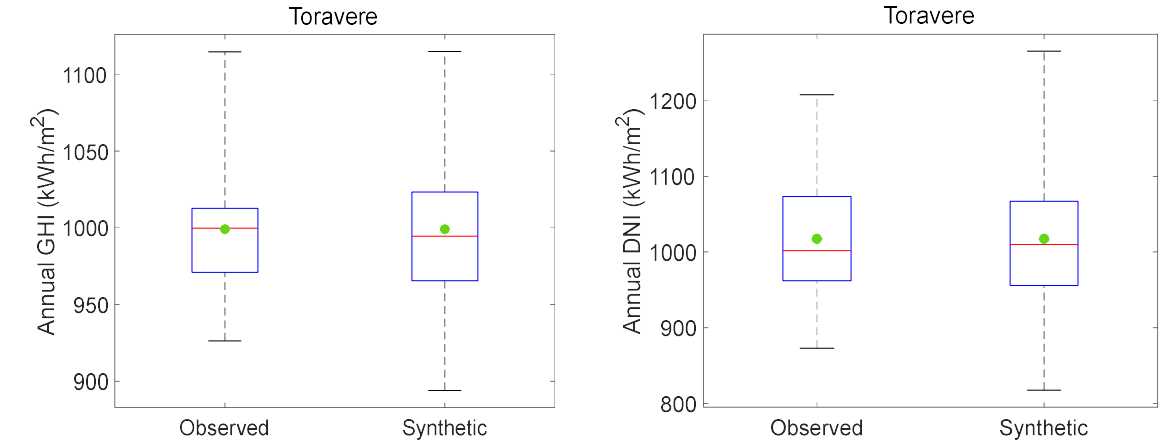


Figure 4-3 Boxplot of the observed and synthetic annual values for GHI (left) and DNI (right) with SRY in a green dot at Toravere.

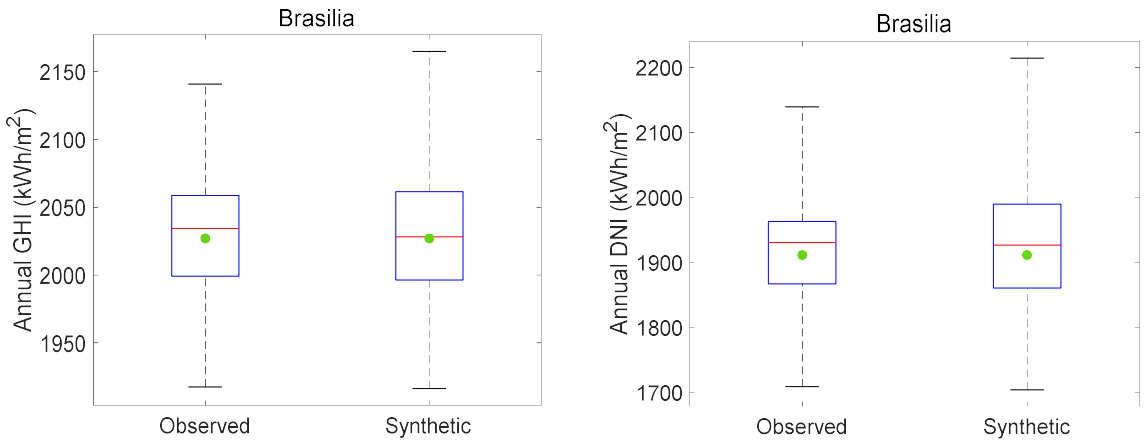


Figure 4-4 Boxplot of the observed and synthetic annual values for GHI (left) and DNI (right) with SRY in a green dot at Brasilia.

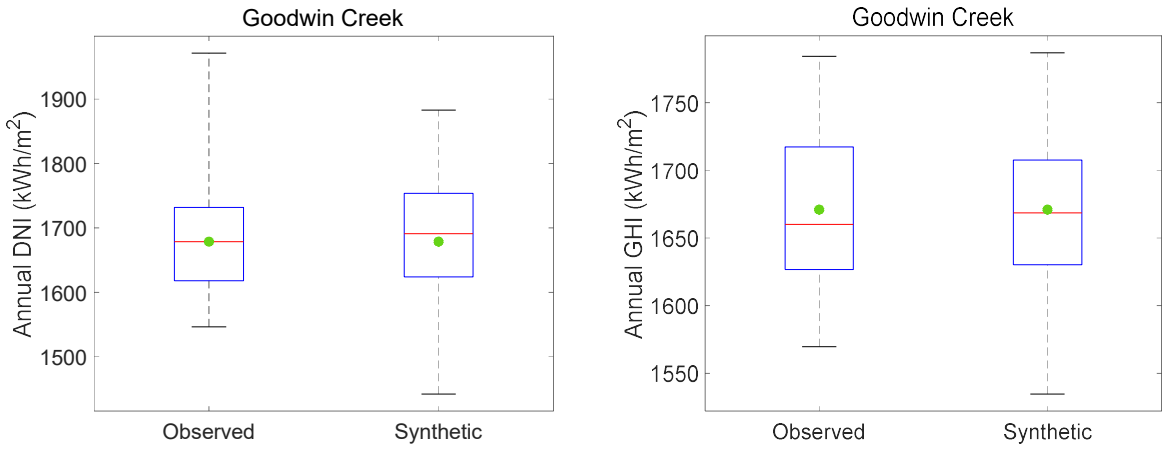


Figure 4-5 Boxplot of the observed and synthetic annual values for GHI (left) and DNI (right) with SRY in a green dot at Goodwin Creek.

In Figures from 4-1 to 4-5, we observe that annual cumulative values of the SRY are very close to the median in the observed and synthetic datasets for both DNI and GHI at all locations. In general, the synthetic values cover a wider range of scenarios than the observed values, meaning that the PSYs subscribe extreme scenarios that have not been observed preserving the natural variability of the resource in each selected location.

Figure 4-6 shows the annual GHI versus the annual DNI of the observed, PSYs and SRYs sets in a scatter plot at all the selected locations.

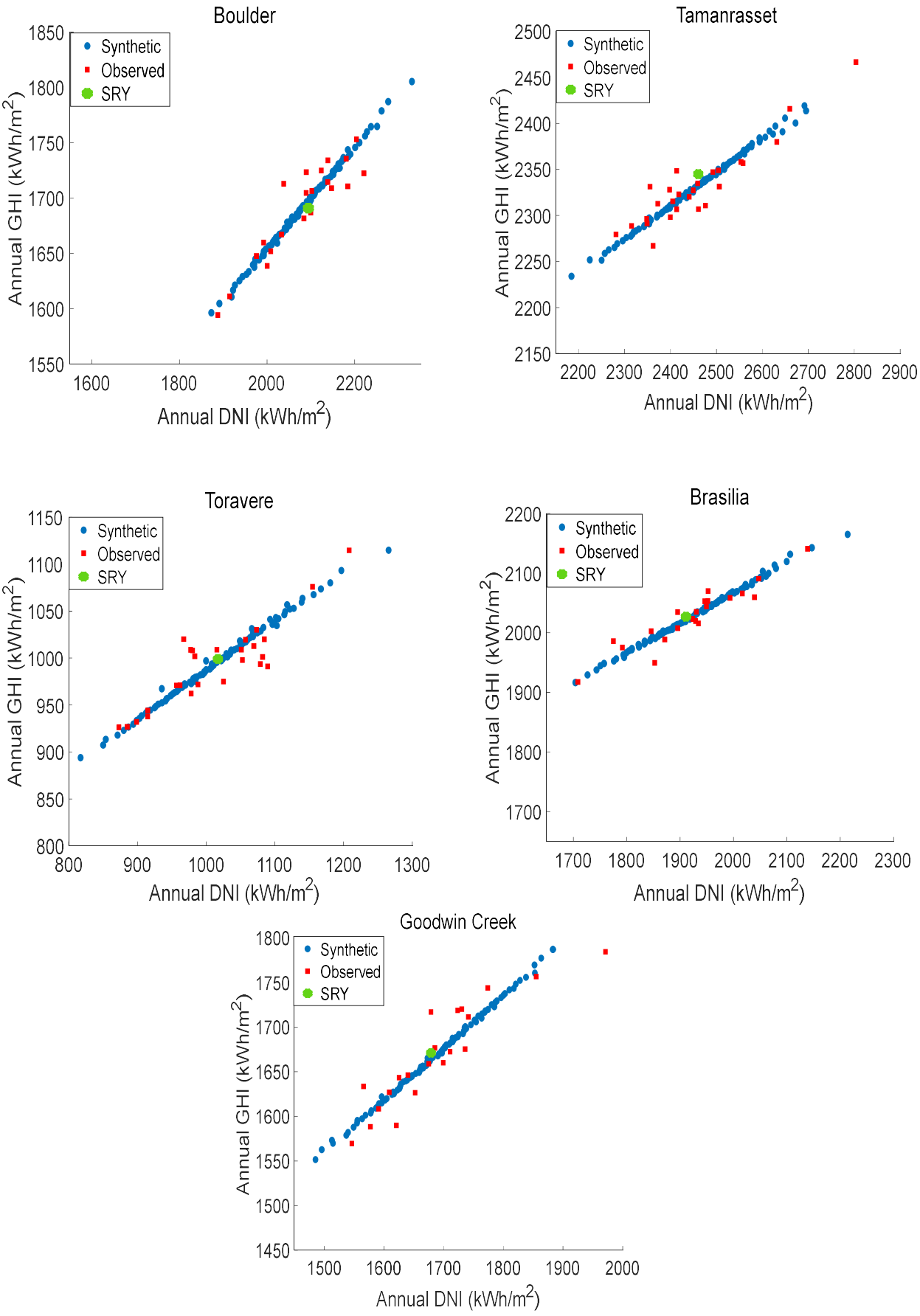


Figure 4-6 Scatter plot of the annual GHI versus the annual DNI. Observed sets are presented in red dots, synthetics PSYs in blue and SRY in green.

In Figure 4-6 we observe that the synthetic values are overlapped to the observed ones with a lower dispersion and covering a wider range of scenarios. For Boulder, Toravere and Brasilia, the algorithm generates extreme annual sets not yet observed among bad case, average and good case annual sets. In Tamanrasset and Goodwin Creek we observe an annual set extremely high that we assume that could be an anomaly derived from the satellite model since we have observed, in these years, an atypical sequence of clear sky consecutive days. The SRYs are located in the middle of the point cloud for all the locations, representing a long-term representative annual set.

4.3 Monthly evaluation

In this section, we compare the monthly values of the synthetic, observed and SRY sets. We calculate the clearness index (k_t) and the direct fraction index (k_b) in a monthly basis in order to exclude seasonality in the evaluation of GHI vs DNI. In Figure 4-7, we face both datasets in a scatter plot for the observed (red), synthetic (blue) and SRY (green).

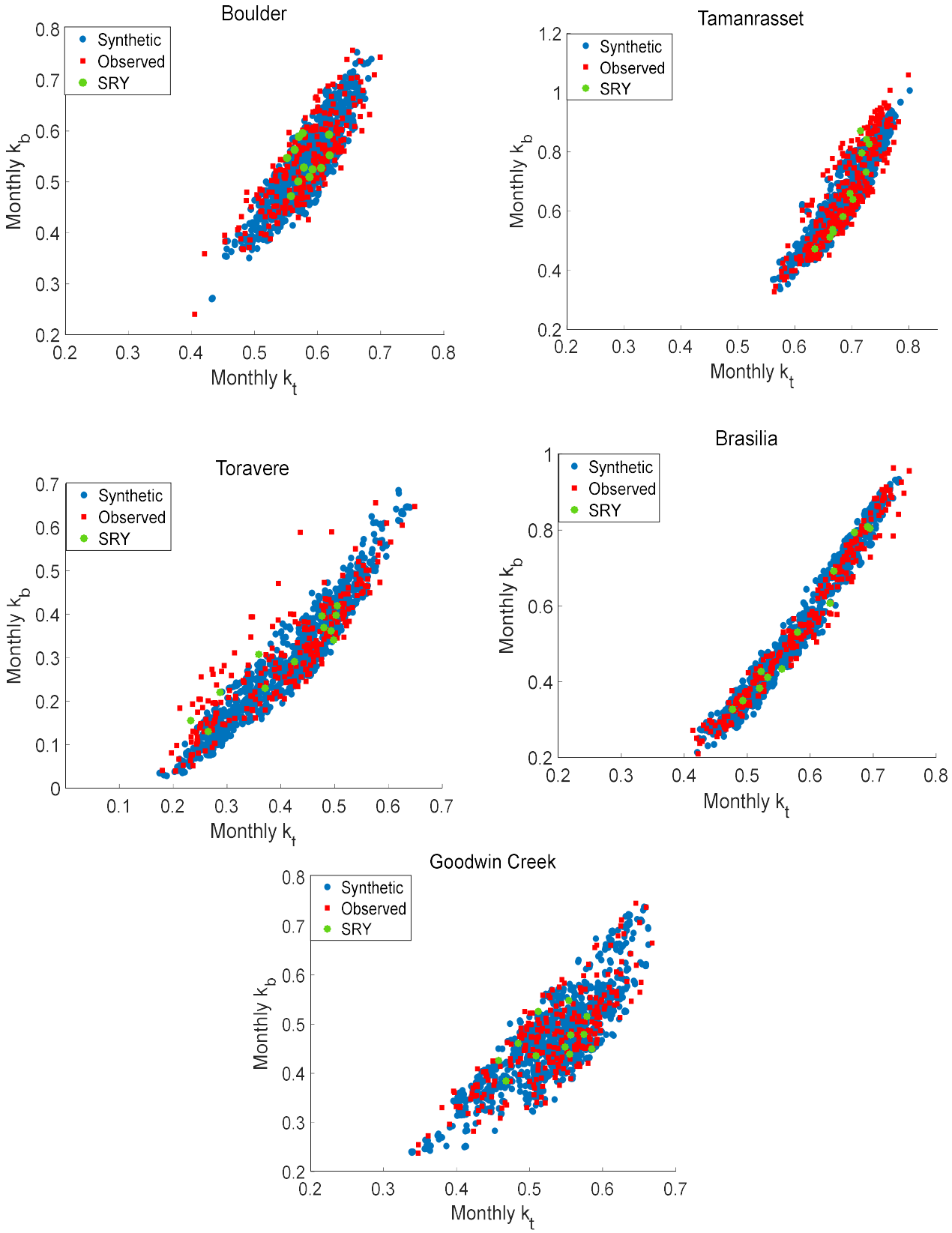


Figure 4-7 Scatter plot of the monthly clearness versus the direct fraction index. Observed sets are presented in red dots, synthetics PSYs in blue and SRY in green.

We observe that we maintain the relation of the GHI and the DNI for all the locations under evaluation because the synthetic data point cloud is overlapped to the observed data point cloud. We maintain the relation between k_t and k_b either in locations with great dispersion like Goodwin Creek or in locations with low dispersion like

Brasilia. There is an exception in Toravere where we find an extremely disperse observed data point cloud that the algorithm is unable to reproduce. At Toravere we can find points where the is close to 0.4 and $k_b$is greater than 0.6.

It is remarkable how the algorithm reproduces significantly clear and significantly cloudy months. For example, in Boulder we find an observed point detached from the main point cloud that presents a pair of k_t and $k_b$values of 0.4 and 0.25 respectively and we find a close value in the synthetic sets. The same performance is observed in Tamanrasset where we find a pair of k_t and k_b of 0.82 and 1.05 respectively that, despite seeming an irregularity, the algorithm is capable to reproduce it.

The SRY cloud points are, as expected, inside the range of the observed sets in the center of the point cloud.

4.4 Daily evaluation

In the daily resolution, we evaluate the distributions of the observed and synthetic sets and the relation between the GHI and the DNI. In all the figures, we maintain the color code where synthetic PMYs are represented in blue, observed sets in red, and TMY in green. In Figures from 4-8 to 4-12, we present the CDFs of the daily cumulative GHI (left) and DNI (right) for the five locations under study.

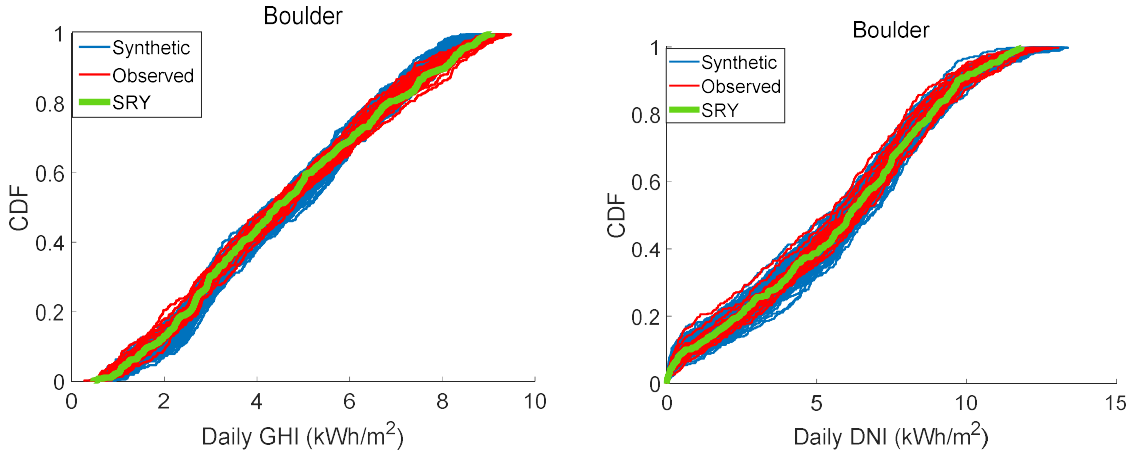


Figure 4-8 CDFs of the daily cumulative GHI (left) and DNI (right) at Boulder. Observed sets are presented in red, synthetic PSYs blue and SRY in green.

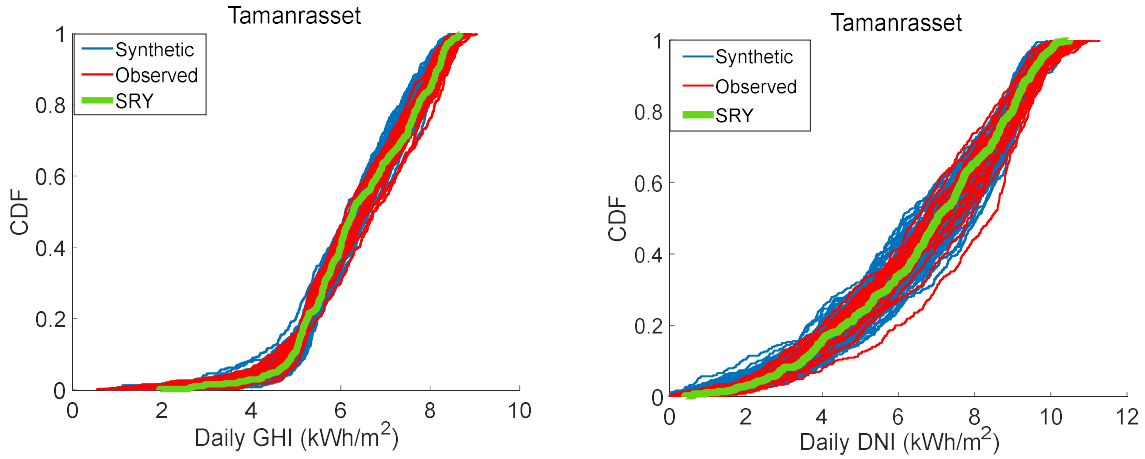


Figure 4-9 CDFs of the daily cumulative GHI (left) and DNI (right) atTamanrasset. Observed sets are presented in red, synthetic PSYs blue and SRY in green.

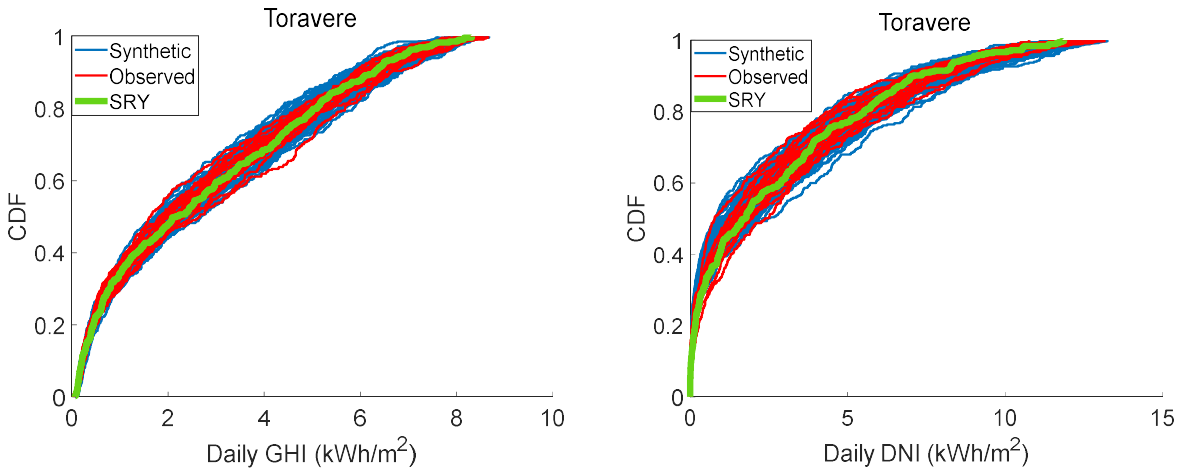


Figure 4-10 CDFs of the daily cumulative GHI (left) and DNI (right) atToravere. Observed sets are presented in red, synthetic PSYs blue and SRY in green.

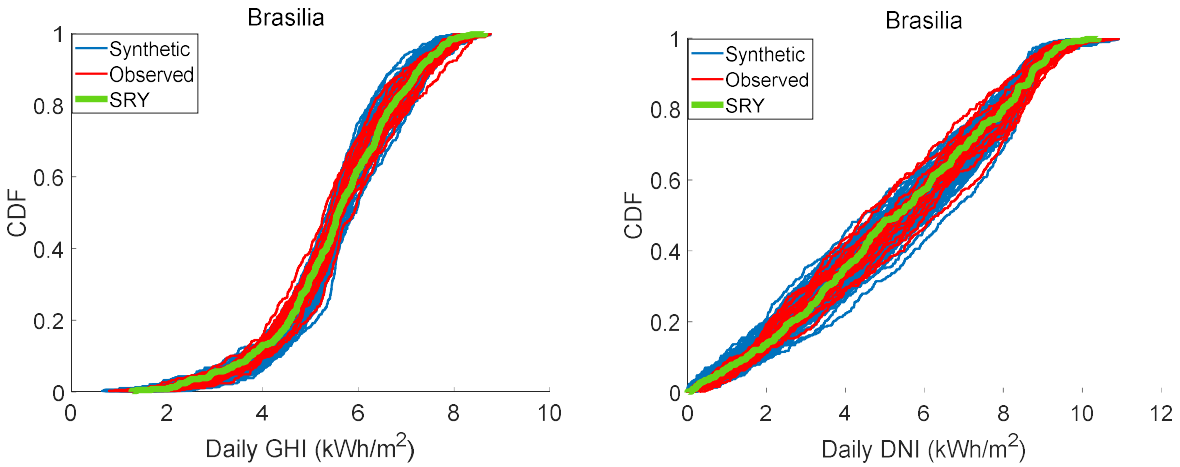


Figure 4-11 CDFs of the daily cumulative GHI (left) and DNI (right) at Brasilia. Observed sets are presented in red, synthetic PSYs blue and SRY in green.

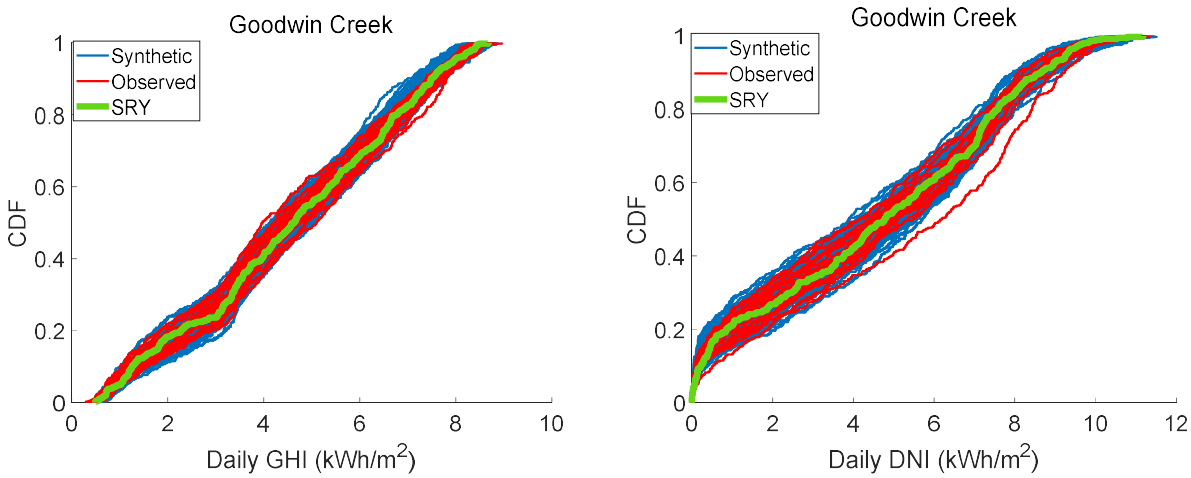


Figure 4-12 CDFs of the daily cumulative GHI (left) and DNI (right)at Goodwin Creek. Observed sets are presented in red, synthetic PSYs blue and SRY in green.

CDF plots provide relevant qualitative information. The CDF profiles with higher slope around high values represent data sets with higher levels of solar radiation (high number of clear days). On the other hand, the CDF

profiles with higher slope around low values (high number of cloudy days) represent data sets with lower levels of solar radiation. The CDF profiles of locations with averaged solar radiation behavior (equilibrated clear and cloudy days) are located in the diagonal of the figure, as it is the case of Boulder (Figure 4-8).

The PMY CDFs present a greater range of scenarios than the observed sets, although, as an exception in the locations of Tamanrasset (Figure 4-9) and Goodwin Creek (Figure 4-12) we can observe a year with high observed DNI daily values that presents a CDF tilted towards the bottom and separated form the rest. These CDFs correspond to the extreme annual values found in Figure 4-7. We straightforward see that the TMY presents an average CDF with respect to the CDFs of the observed data sets, since it is located in the middle of the space occupied by the CDFs of the observed sets.

The shape of the CDFs denotes significant variations depending on the location. Tamanrasset presents the steeper slope and the lower width mainly for GHI. This shape describes a location with high occurrence of clear sky solar radiation daily profiles. Toravere (Figure 4-10) shows just the opposite performance presenting the gentler slope describing a location with high occurrence of cover sky daily profiles. Brasilia also presents a pronounced slope in the CDFs but more centred around 5-6 kWh/m2 describing a location with a high occurrence of partially cover sky daily profiles. Goodwing Creek and Boulder show great similitudes in their CDFs that are fairly linear and close to the slope equal to 1 representing a climate with similar occurrence of clear, partially cover and covered daily solar radiation profiles.

In Figure 4-13, we present the daily clearness index versus de daily direct fraction index of the observed and synthetic sets in a scatter plot for all the locations under evaluation.

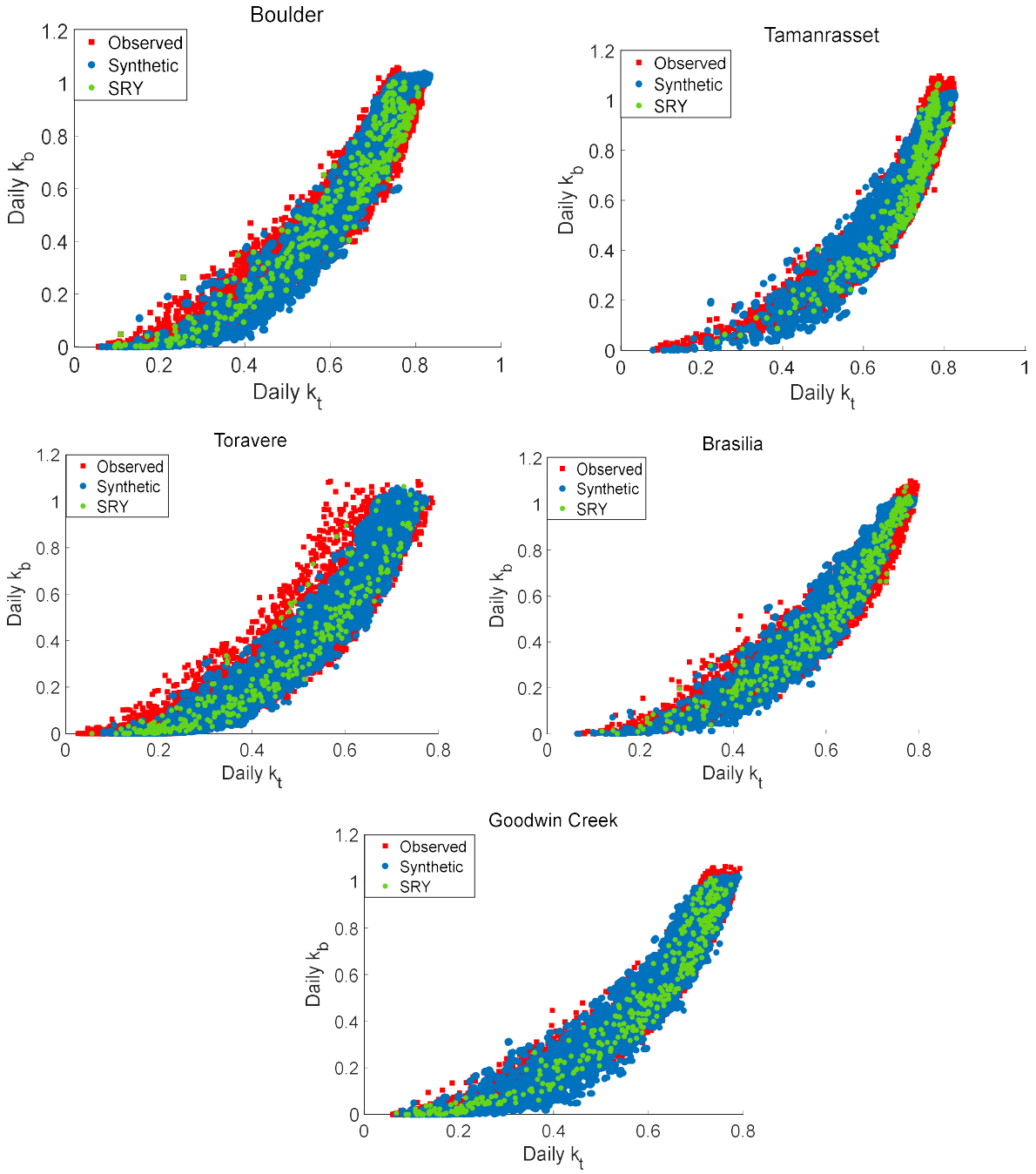


Figure 4-13 Scatter plot of the daily clearness index versus the direct fraction index. Observed sets are presented in red dots, synthetic PSYs in blue and SRY in green.

We observe a similar shape of the synthetic data (PSYs) point cloud to the observed data point cloud. However, the observed point cloud shows greater dispersion especially in Toravere and Boulder. It is worth highlighting that Toravere showed a particular dispersion in the monthly (Figure 4-7) and annual (Figure 4-6) resolution evaluation.

We also observe greater k_t-k_b pairs in the observed sets. These differences shall be associated to differences between the satellite derived and synthetic fit of the DNI Clear sky model. Clear sky characterization of satellite-derived data is one of the most significant source of errors of satellite-derived data [25].

SRYs point cloud stays in the middle of the point cloud of the observed sets reproducing an average performance; nevertheless, Tamanrasset shows a double tendency for high values again justified for possible deviations in satellite data estimations for clear sky periods in this location.

4.5 One-minute evaluation

In this section, we evaluate the performance of the algorithm in 1-minute resolution. For this purpose, we use the clearness index (k_t) and the direct fraction index (k_b) to compare the distributions of the observed and synthetic series in the selected locations. To this end, we use three years of one-minute measurements of direct normal irradiance (DNI) and global horizontal irradiance (GHI) registered by the Baseline Surface Radiation Network (BSRN) of the five selected site to compare with the synthetic dataset generated.

4.5.1 Frequency distribution of k_t and k_b values.

In Figures from 4-14 to 4-18 we show the probability density distribution of 1-minute k_t values (left) and 1-minute k_b values (right) of the observed and synthetic sets at all the selected locations.

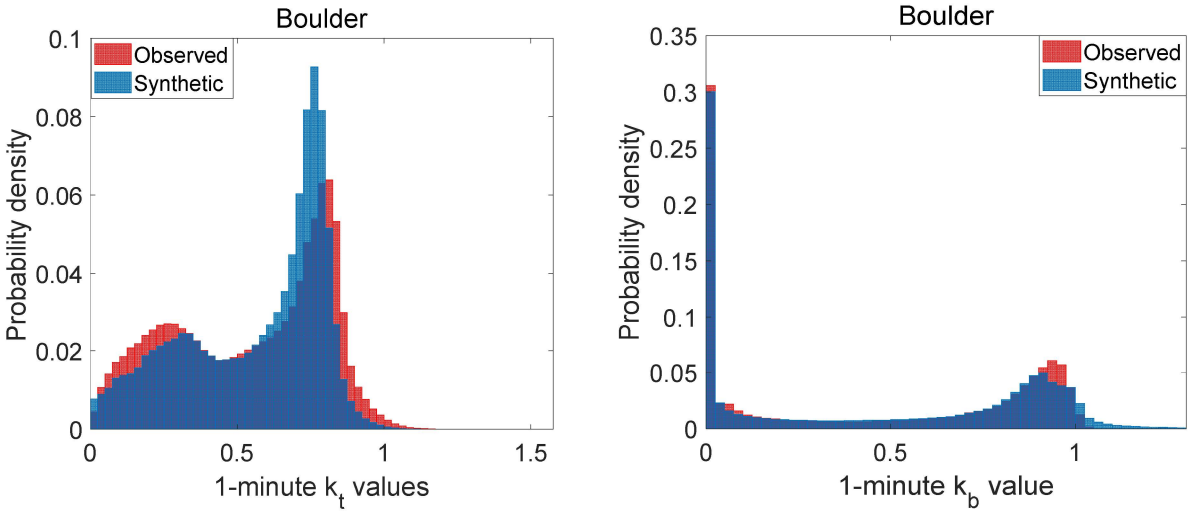


Figure 4-14 Probability density distribution of 1-minute k_t and k_b values at Boulder. Observed sets are represented in red and synthetic PSYs in blue.

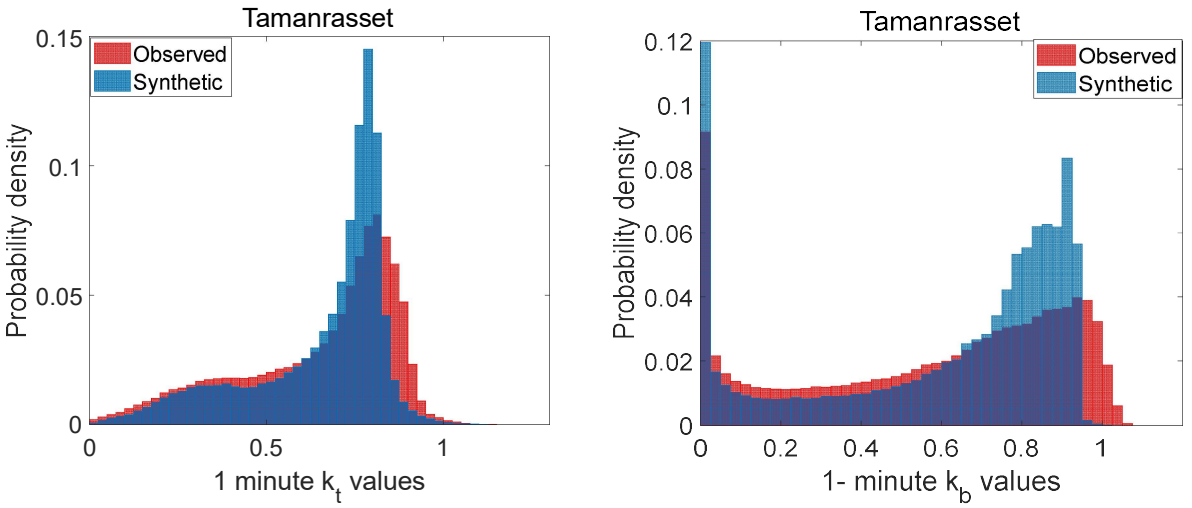


Figure 4-15 Probability density distribution of 1-minute k_t and k_b values at Tamanrasset. Observed sets are represented in red and synthetic PSYs in blue.

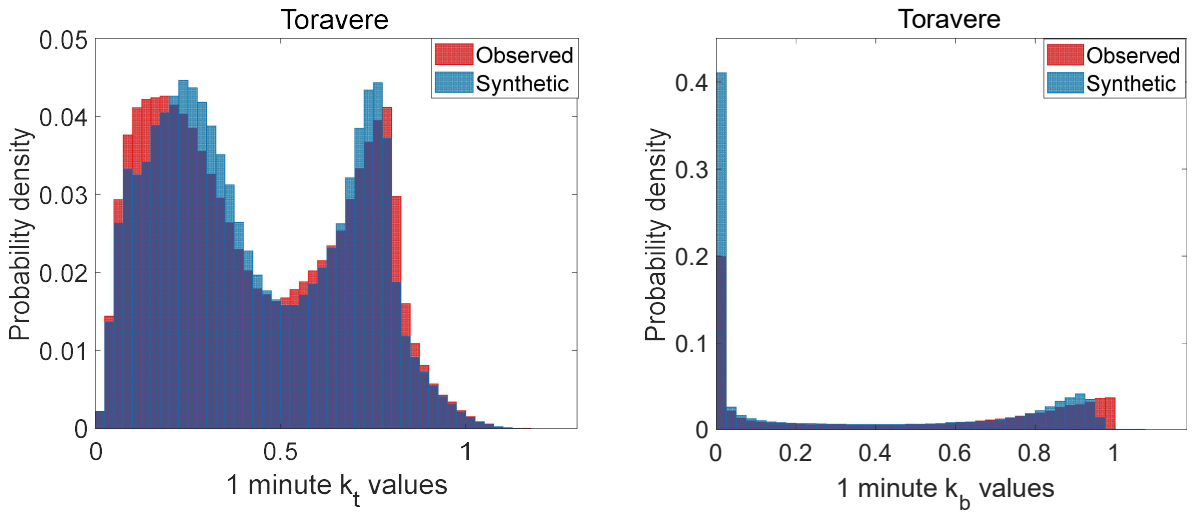


Figure 4-16 Probability density distribution of 1-minute k_t and k_b values at Toravere. Observed sets are represented in red and synthetic PSYs in blue.

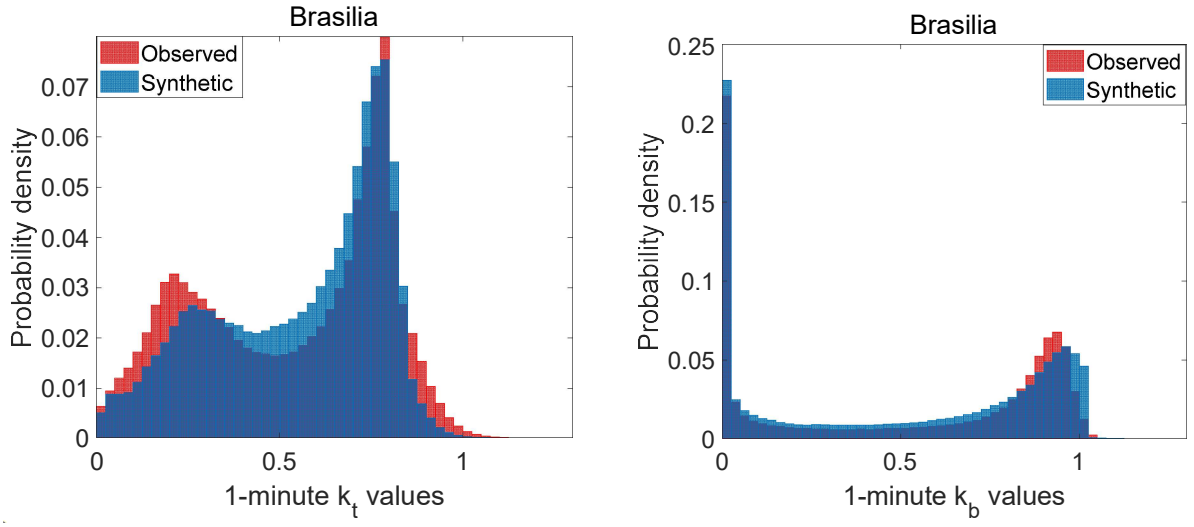


Figure 4-17 Probability density distribution of 1-minute k_t and k_b values at Brasilia. Observed sets are represented in red and synthetic PSYs in blue.

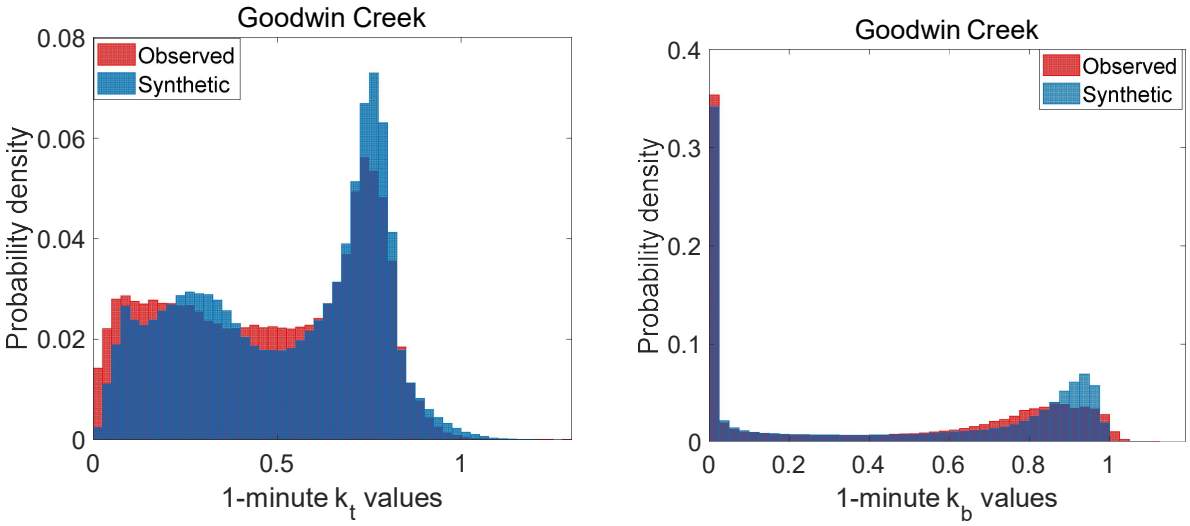


Figure 4-18 Probability density distribution of 1-minute k_t and k_b values at Goodwin Creek. Observed sets are represented in red and synthetic PSYs in blue.

In these figures, we can observe some degree of bimodality in the 1-minute kt distributions of all locations, this is due to the presence of cloudy (peak at low) and cloud-less (peak at high k_t) [26] sky conditions.When we compare observed and synthetic distributions, we can see that the synthetic 1-minute k_t values reproduce the distribution and trend of the observed series and maintain the bimodal character of the curves at all locations. The peaks corresponding to high k_t are centered in similar positions. In the case of Boulder (Figure 4-14), Tamanrasset (Figure 4-15) and Goodwin Creek (Figure 4-18), the synthetic series present a greater probability for the peak at high k_t than the observed data. Although, in the case of Tamanrasset and Boulder, the area associated to high k_t values is very similar. In the case of Goodwin Creek the difference of area at high k_t values is balanced with a higher frequency of partially cloudy days. In general, the k_t distributions maintain dispersion and amplitude between observed and synthetic sets. In the case of 1-minute observed and synthetic k_b distributions, the shape of the probability density profile is maintained for low k_b values. However, for high kb values we find greater differences, especially remarkable for Tamanrasset location. Although, as in the case of k_t distributions, these differences seem to be balanced.So, they present similar areas for k_b values higher than 0.7.

To perform a detailed analysis, we have studied the probability density distribution of 1-minute k_t and k_b values conditioned to different air optical mass range following the same methodology used by Fernández-Peruchena [27] and Tovar et al. [3]. The ranges of air optical mass (m) considered are shown in Table 4-2. These ranges are considered since the effects of the clouds have a clear dependence with the zenith angle. The objective of this analysis is to evaluate if the synthetic series reproduce the behaviour of the observed data for each range.

Table 4-2 Optical air mass intervals considered in our study.

m	1.0	1.5	2.0	2.5	3	3.5
Interval	1-1.2	1.4-1.6	1.8-2.2	2.3-2.7	2.75-3.25	>3.25

In Figures from 4-19 to 4-23, we can see the probability density distribution of 1-minute k_t (left) and k_b values (right) of the observed and synthetic sets at all the selected locations for m equal to 2. The rest of the results obtained for all ranges of m and for all locations are shown in the Annex A.

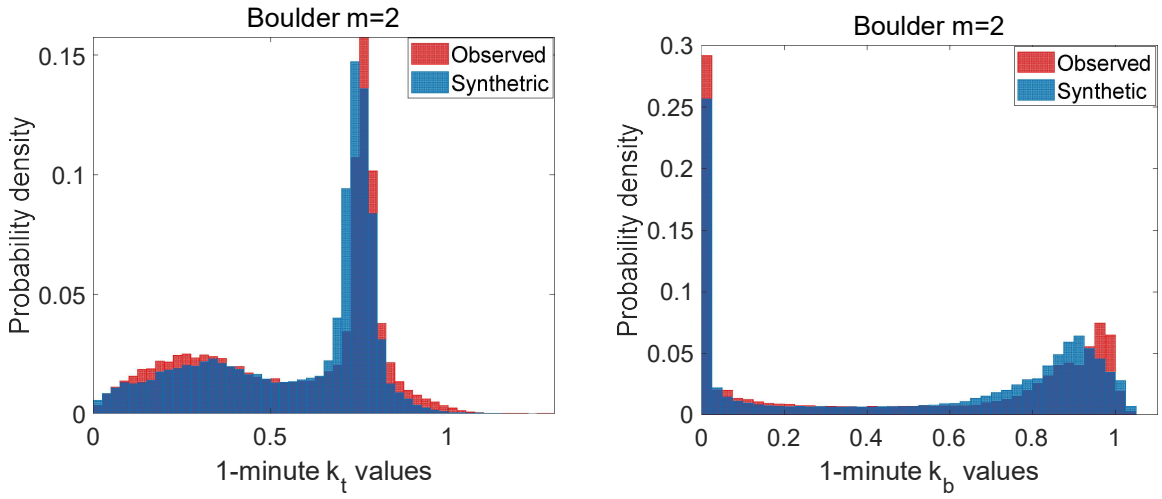


Figure 4-19 Probability density distribution of 1-minute k_t and k_b values at Boulder for m=2. Observed sets are represented in red and synthetic PSYs in blue.

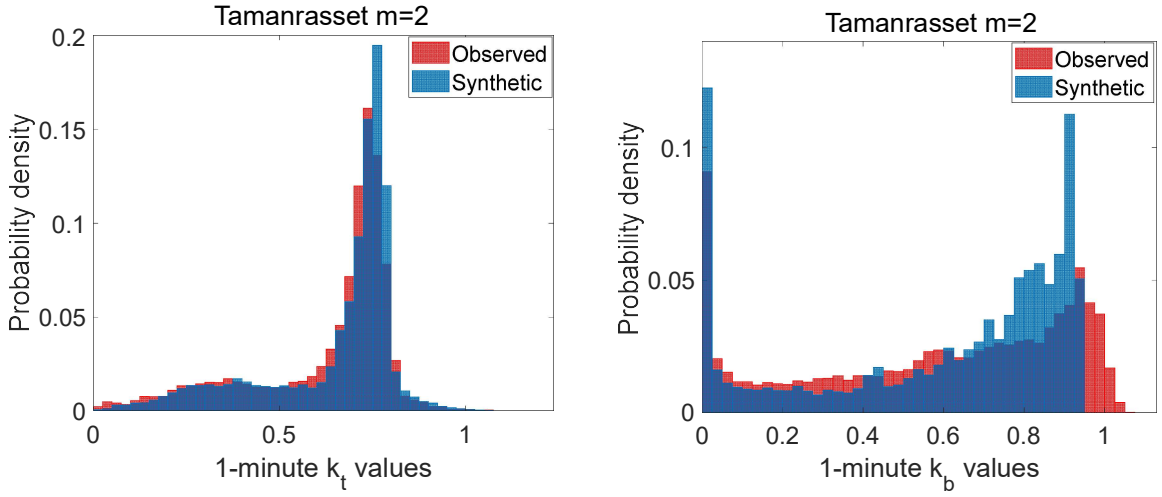


Figure 4-20 Probability density distribution of 1-minute k_t and $k_b$values at Tamanrasset for m=2. Observed sets are represented in red and synthetic PSYs in blue.

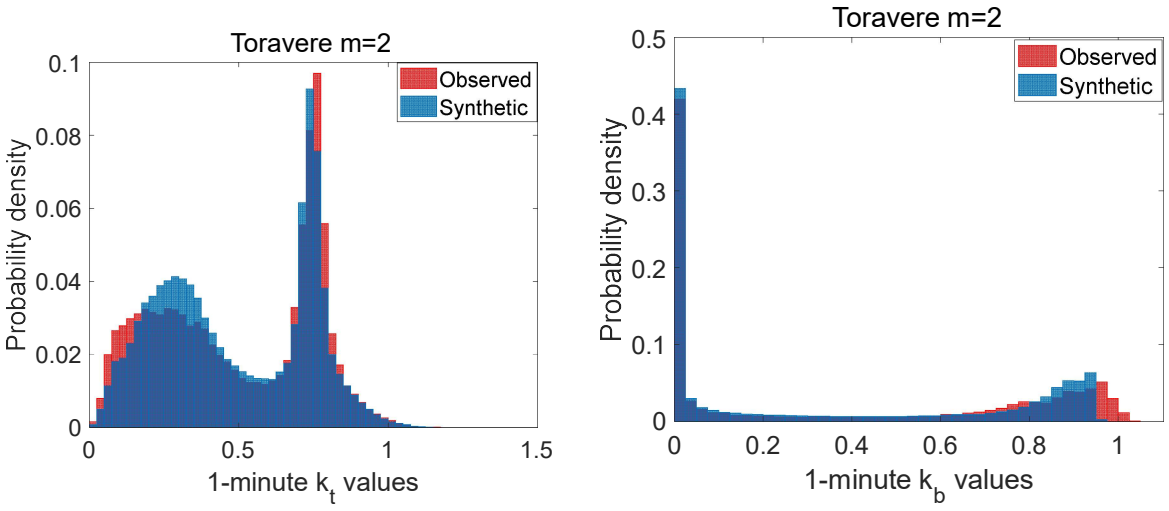


Figure 4-21 Probability density distribution of 1-minute k_t and k_b values at Toravere for m=2. Observed sets are represented in red and synthetic PSYs in blue.

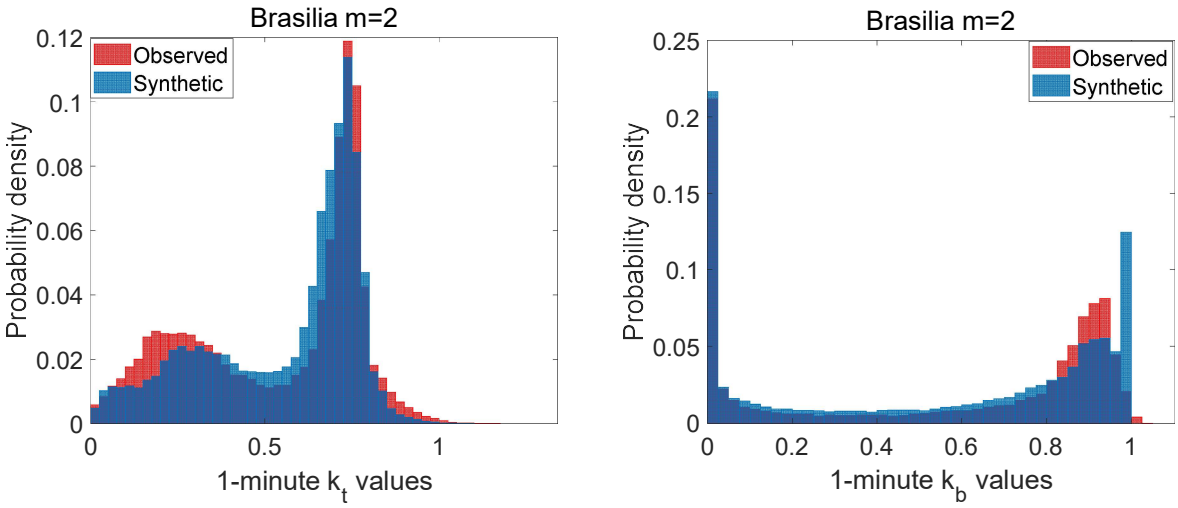


Figure 4-22 Probability density distribution of 1-minute k_t and k_b values at Brasilia for m=2. Observed sets are represented in red and synthetic PSYs in blue.

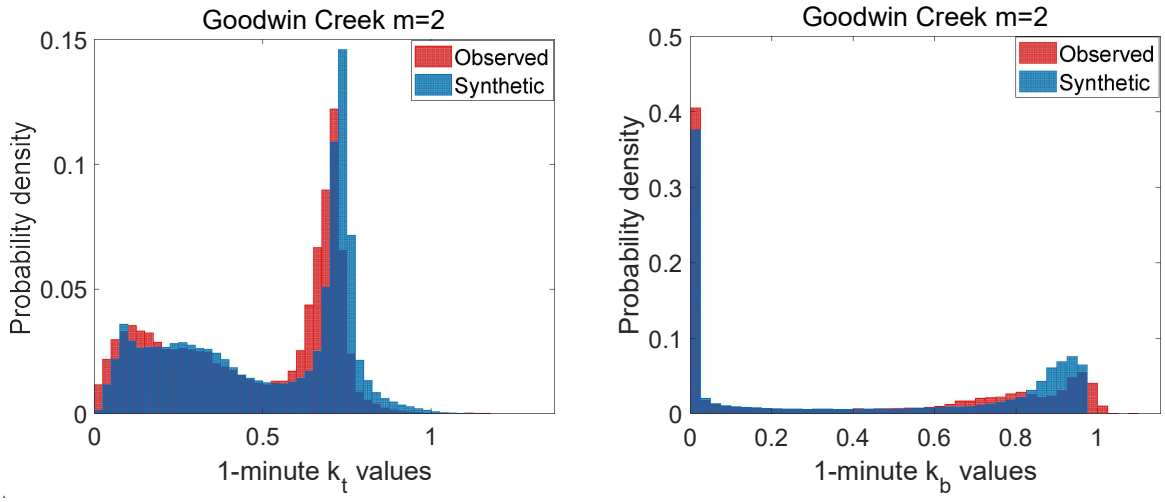


Figure 4-23 Probability density distribution of 1-minute k_t and k_b values at Goodwin Creek for m=2. Observed sets are represented in red and synthetic PSYs in blue.

In the case of m equal to 2, we can observe similar behaviour to the previously studied curves of all k_t and k_b. In general, the synthetic 1-minute k_t values reproduce the distribution and trend of the observed sets. As in the

previous case, the peaks corresponding to high k_t of the synthetic and observed values are centered, except in the case of Goodwin Creek (Figure 4-23) that seems to be more displaced. We also highlight the case of Toravere (Figure 4-21), where the synthetic values of low k_t have a greater probability than the observed set. For the 1-minute observed and synthetic k_b distributions, the shape of the probability density profile is maintained for low $k_b$values. However, we observe greater differences between k_b curves for elevated k_b values. In this case, the syntethic sets of Goodwin Creek (Figure 4-23), Boulder (Figure 4-19) and Toravere (Figure 4-21) reproduce well the behaviour of the observed sets since the differences seem to be compensated. In the case of Tamanrasset (Figure 4-20), for high k_b values the results are not as good as previously observed. The shape of the synthetic values reaches higher values than the observed values for a range ok k_b from 0.7 to 0.9. We can also observe that in the case of Brasilia (Figure 4-22), the probability values of the synthetic series for k_b values from 0.8 to 0.9 are lower, however, in the synthetic series for values of k_b equal to 1, the probability is greater than the observed values. So, the probabilities of k_b values on clear sky conditions are blanced. The differences found between synthetic and observed k_b distributionmay be a consequence of the different clear sky model used to calculate this index for both series.

4.5.2 Distribution analysis

To compare quantitatively the observed and synthetic distributions of k_t , we have fitted the probability density distribution curves by a function previously used by other authors, Tovar et al. (1997) propose to approach the bimodal character of these distributions by the sum of two functions based on the Boltzman statistics:

$$f(k_t|m) = f_1(k_{t1}) + f_2(k_{t2}) \quad (4\text{-}1)$$

$$f_i(k_t) = A_i \frac{\lambda_i e^{(k_t - k_{t0i})\lambda i}}{(1 + e^{(k_t - k_{t0i})(\lambda i)})^2} \quad (4\text{-}2)$$

This function is symmetric around k_{t0i}, where the function reaches its maximum. The parameters $A_i\lambda i$ determine the probability of this maximum and λi is also related to the width of the function for eachk_t. In this case, the parameters of f_1 is associated with cloudy sky conditios and parameters of $f_2$is associated with clear sky conditions. We only use this function to analyze the 1-minute k_t distribution. The 1-minute k_b distributions are not analyzed because they do not presesent a bimodal curve and we have not found any study in the bibliography that allows to compare the distributions.

The parameters of the function let us compare the shape of the distributions of both types of data for all the ranges of m. In Figures from 4-24 to 4-28 we represent the values of k_{th1} and k_{th2} of both series for all the locations.

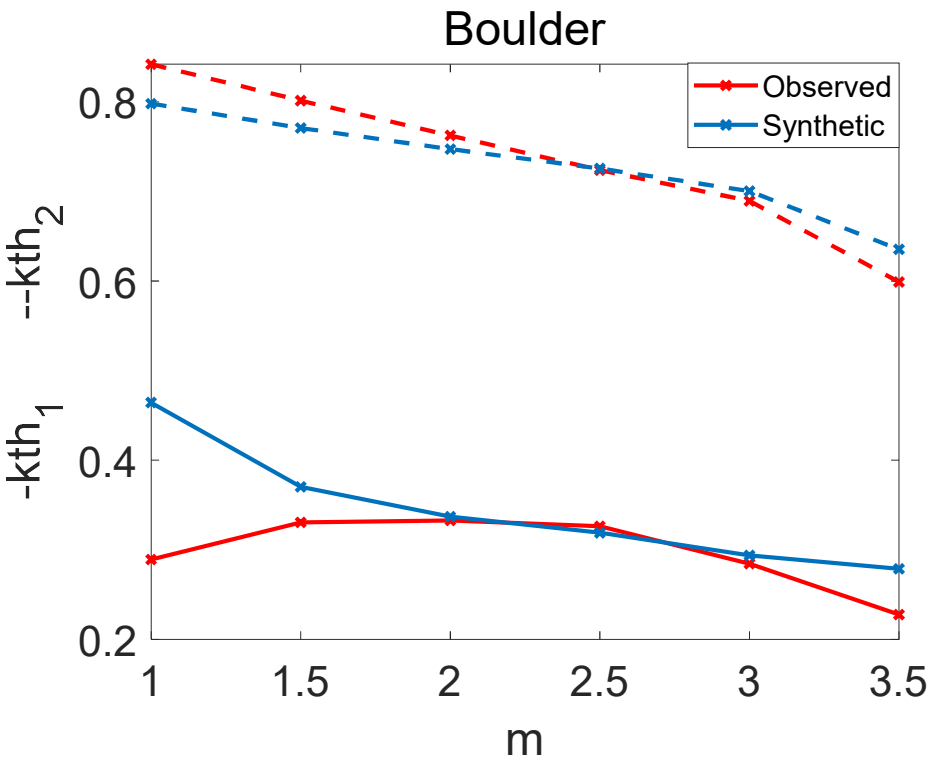


Figure 4-24 Comparision of the k_{th1} and k_{th2} parameters between observed (red) and synthetic series (blue) for all the ranges of m in Boulder.

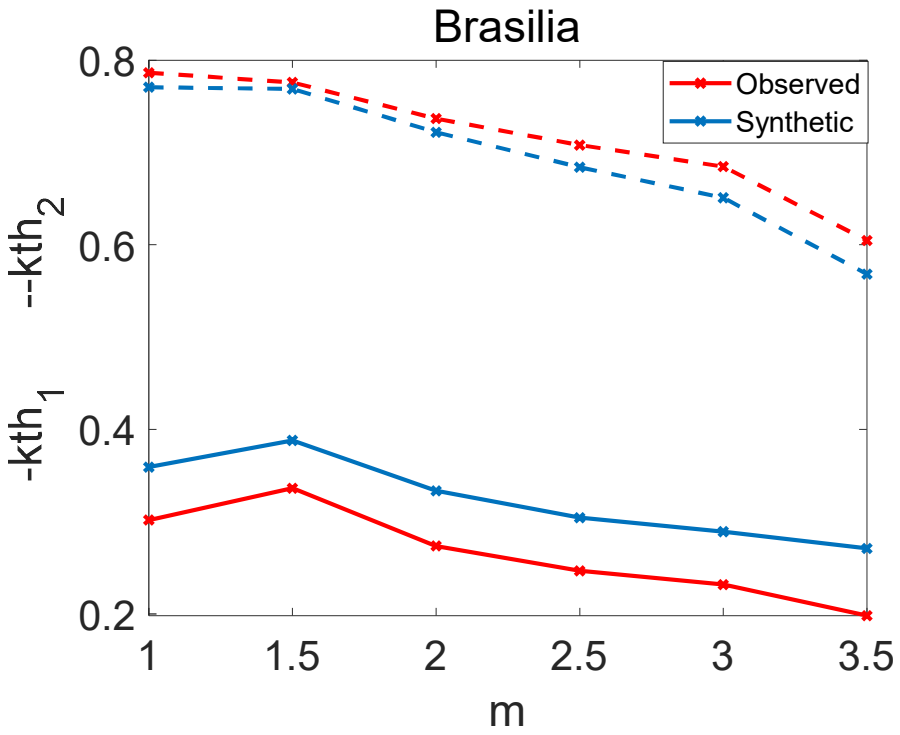


Figure 4-25 Comparision of the k_{th1} and k_{th2} parameters between observed (red) and synthetic series (blue) for all the ranges of m in Brasilia

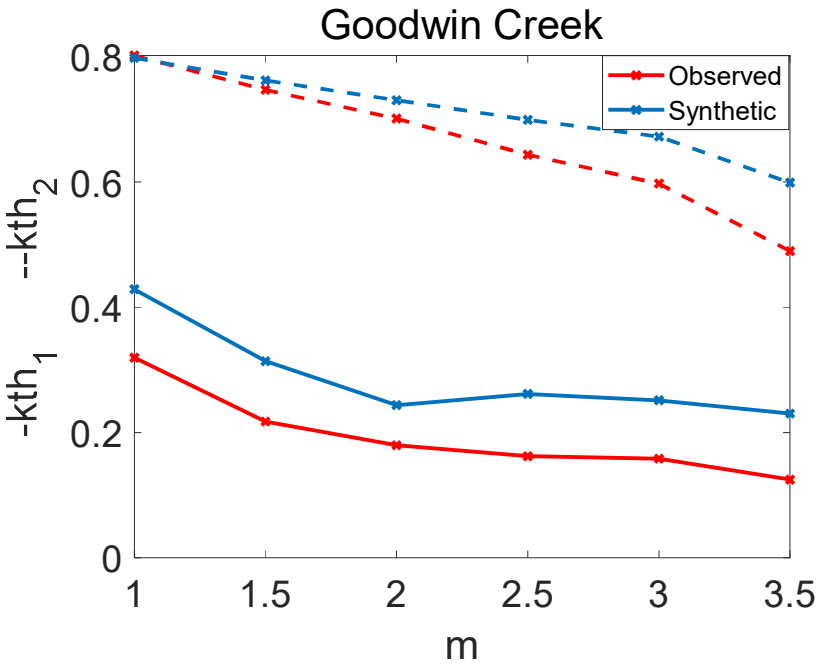


Figure 4-26 Comparision of the k_{th1} and k_{th2} parameters between observed (red) and synthetic series (blue) for all the ranges of m in Goodwin Creek

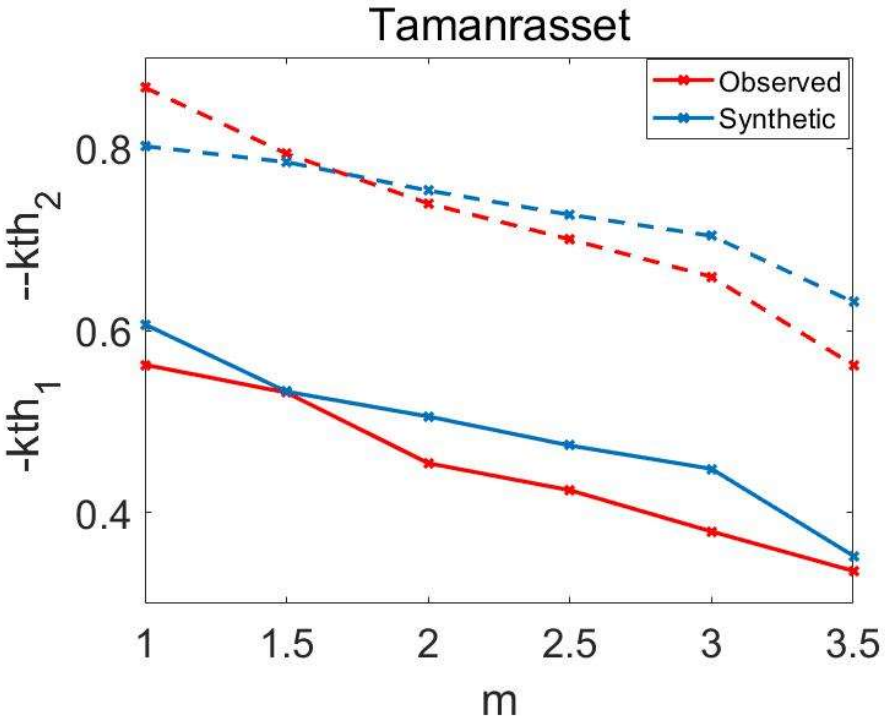


Figure 4-27 Comparision of the k_{th1} and k_{th2} parameters between observed (red) and synthetic series (blue) for all the ranges of m in Tamanrasset

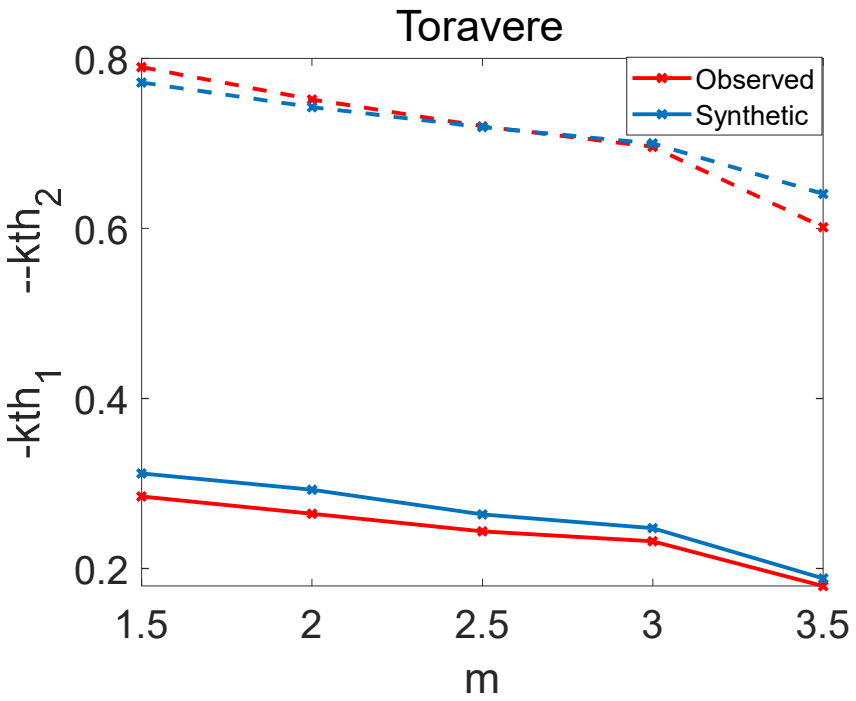


Figure 4-28 Comparision of the k_{th1} and k_{th2} parameters between observed (red) and synthetic series (blue) for all the ranges of m in Toravere

In these Figures, we can observe that for the ranges of m from 2 to 3 the values of $k_{th1}$and k_{th2} of the synthetic series follow the trend of the observed series, being the values of these parameters very similar in locations such as Brasilia, Boulder and Toravere. This means that the peaks of the synthetic and observes curves in these locations are centered in very similar values.

Next, we evaluate the parameters for the most significant ranges of m. The range of m=1 has been discarded because in Toravere there are not $k_t$values for this range. In table 4-3, 4-4 and 4-5 we show the values of these parameters for ranges of m=1.5, m=2 and m=3.

Table 4-3 Comparison of the parameters at all locations of observed and synthetic 1-minute k_t distribution for m=1.5

	Goodwin Creek		**Brasilia**		**Boulder**		**Tamanrasset**		**Toravere**	
m=1.5	**Obs**	**Synth**	**Obs**	**Synth**	**Obs**	**Synt**	**Obs**	**Synth**	**Obs**	**Synth**
$A_1\lambda_1/A_2\lambda_2$	0.198	0.145	0.158	0.173	0.112	0.121	0.067	0.050	0.273	0.377
Kth1	0.217	0.314	0.336	0.388	0.331	0.370	0.532	0.533	0.285	0.312
Kth2	0.747	0.762	0.776	0.769	0.802	0.771	0.794	0.785	0.790	0.772

Table 4-4 Comparison of the parameters at all locations of observed and synthetic 1-minute k_t distribution for m=2

	Goodwin Creek		**Brasilia**		**Boulder**		**Tamanrasset**		**Toravere**	
m=2	**Obs**	**Synth**	**Obs**	**Synth**	**Obs**	**Synt**	**Obs**	**Synth**	**Obs**	**Synth**
$A_1\lambda_1/A_2\lambda_2$	0.276	0.195	0.233	0.224	0.146	0.140	0.112	0.097	0.368	0.463
Kth1	0.180	0.244	0.274	0.333	0.333	0.337	0.454	0.506	0.265	0.293
Kth2	0.701	0.731	0.737	0.722	0.763	0.747	0.739	0.754	0.752	0.743

Table 4-5 Comparison of the parameters at all locations of observed and synthetic 1-minute k_t distribution for m=3

	Goodwin Creek		**Brasilia**		**Boulder**		**Tamanrasset**		**Toravere**	
m=3	**Obs**	**Synth**	**Obs**	**Synth**	**Obs**	**Synt**	**Obs**	**Synth**	**Obs**	**Synth**
$A_1\lambda_1/A_2\lambda_2$	0.466	0.339	0.343	0.291	0.221	0.219	0.208	0.174	0.471	0.694
Kth1	0.158	0.251	0.232	0.289	0.285	0.294	0.379	0.448	0.232	0.248
Kth2	0.598	0.673	0.685	0.651	0.690	0.700	0.659	0.704	0.696	0.700

In tables Tables 4-3,4-4 and 4-5, we can see the ratio between the product $A_1\lambda_1$ and $A_2\lambda_2$. This ratio shows the bimodal behaviour of the distribution since quantify the ratio between the probability of both peaks. We can observe in these tables that the best results for this product are shown is in Brasilia and Boulder, wich differences are approximately 10% for m=1.5 and 4% for m=2. The site with the highest differences in this product of parameters is Toravere with a difference of 38% for m=1.5, 26% for m=2 and 47% for m=3. Therefore, this indicates that in the locations where the amplitude and dispersion of the curves is best reproduced is in Brasilia and Boulder for the case of m=2. Regarding the values of k_{th1} and k_{th2}, we observe that the differences between the observed and synthetic values are smaller, except in Goodwin Creek and Brasilia. That means the peaks of the observed and synthetic curves are centered on similar values as we have previously observed. In the case of Goodwin Creek the values of k_{th1} shown the greatest difference for all ranges of m. In the case of k_{th2} the differences al lower in all the locations.

5 CONCLUSIONS AND FUTURE WORKS

In this work, we present an improve algorithm for the synthetic generation of plausible solar years. The algorithm uses an extended database (10-20 years) of coupled GHI and DNI at hourly resolution as input and provides 100 synthetics coupled GHI+DNI 1-min datasets as output. The algorithm has been tested in five locations with different Koppen-Geiger climates without any local adaptation using satellite-derived data.

Synthetic data has been evaluated in different time resolutions in comparison to the observed sets. In annual resolution the PSYs subscribe extreme scenarios maintaining the natural variability of the resource in all the locations. In monthly resolution, it is remarkable how the algorithm reproduces significantly clear and cloudy months, maintaining the kt-kb relation in locations with high and low dispersion. In daily resolution, the selected locations describe several solar radiation scenarios that the algorithm is capable to reproduce and enlarge, providing a wider range of scenarios. Finally, the synthetic 1-minute kt and kb frequency distributions reproduce the trend of the observed series maintaining the dispersion and amplitude of the distribution profiles at all the evaluated locations. In the case of Tamanraset and Goodwin Creek, extremely high solar radiation dataset has been observed in all resolutions, but this anomaly seems not affect significantly to the obtained results. These anomalies are due to the uncertainty of the satellite data. This tool does not reduce the uncertainty associated to the input data.

Once the performance of the method has been assessed, we can conclude that the synthetic PSYs obtained with this methodology reproduce the behavior of the observed datasets maintaining the relation between GHI and DNI and allowing to obtain extreme scenarios that have not been yet observed. Results show that this approach can complement the common approach of the TMY bringing the opportunity to perform a stochastic assessment of solar harnessing systems.

Future works could focus on the generation of Plausible Meteorological Years instead of Plausible Solar Years, considering other relevant meteorological variables like wind speed, temperature, and relative humidity in addition to solar radiation data. In addition, the inclusion of new climates on the database used for the 1-min synthetic generation could improve the obtained results.

The approach considers the hypothesis that future trends can be characterized by past trends but the impact of climate change on long term tendencies either in solar radiation and meteorological variables shall be implemented.

REFERENCES

[1] L. Hall, R. Prairie, H. Anderson, and E. Boes, "Generation of a Typical Meteorological Year for 26 SOLMET stations.," 1978.

[2] Nielsen, K.P.; Vignola, F.; Ramírez, L.; Blanc, P.; Meyer, R.; Blanco, "M. Excerpts from the report: 'BeyondTMY - Meteorological data sets for CSP/STE performance simulations.,'" *AIP Conf. Proc.*, vol. 1850, 1400, 2017, doi: 10.1063/1.4984525.

[3] W. Wilcox, S.; Marion, "Users manual for TMY3 data sets," 2008.

[4] E. F. M. Abreu, P. Canhoto, V. Prior, and R. Melicio, "Solar resource assessment through long-term statistical analysis and typical data generation with different time resolutions using GHI measurements," *Renew. Energy*, vol. 127, pp. 398–411, 2018, doi: 10.1016/j.renene.2018.04.068.

[5] Climate Normals., "National Centers for Environmental Information (NCEI) formerly known as National Climatic Data Center (NCDC)," *Natl. Centers Environ. Inf. Former. known as Natl. Clim. Data Cent.*, [Online]. Available: https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals.

[6] K. Engeland, M. Borga, J. D. Creutin, B. François, M. H. Ramos, and J. P. Vidal, "Space-time variability of climate variables and intermittent renewable electricity production – A review," *Renew. Sustain. Energy Rev.*, vol. 79, no. February, pp. 600–617, 2017, doi: 10.1016/j.rser.2017.05.046.

[7] C. F. Peruchena, M. Larrañeta, M. Blanco, and A. Bernardos, "High frequency generation of coupled GHI and DNI based on clustered Dynamic Paths," *Sol. Energy*, vol. 159, no. August 2017, pp. 453–457, 2018, doi: 10.1016/j.solener.2017.11.024.

[8] P. Heller, *The Performance of Concentrated Solar Power (CSP) Systems: Analysis, Measurement and Assessment*. 2017.

[9] A. Habte, M. Sengupta, C. Gueymard, A. Golnas, and Y. Xie, "Long-term spatial and temporal solar resource variability over America using the NSRDB version 3 (1998–2017)," *Renew. Sustain. Energy Rev.*, vol. 134, no. March, p. 110285, 2020, doi: 10.1016/j.rser.2020.110285.

[10] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, "The National Solar Radiation Data Base (NSRDB)," *Renew. Sustain. Energy Rev.*, vol. 89, no. January 2018, pp. 51–60, 2018, doi: 10.1016/j.rser.2018.03.003.

[11] C. K. Ho, S. S. Khalsa, and G. J. Kolb, "Methods for probabilistic modeling of concentrating solar power plants," *Sol. Energy*, vol. 85, no. 4, pp. 669–675, 2011, doi: 10.1016/j.solener.2010.05.004.

[12] M. Larrañeta, C. Fernandez-Peruchena, M. A. Silva-Pérez, I. Lillo-bravo, A. Grantham, and J. Boland, "Generation of synthetic solar datasets for risk analysis," *Sol. Energy*, vol. 187, no. May, pp. 212–225, 2019, doi: 10.1016/j.solener.2019.05.042.

[13] C. M. Fernández-Peruchena, M. Gastón, M. Sánchez, J. García-Barberena, M. Blanco, and A. Bernardos, "MUS: A multiscale stochastic model for generating plausible meteorological years designed for multiyear solar energy yield simulations," *Sol. Energy*, vol. 120, pp. 244–256, 2015, doi: 10.1016/j.solener.2015.07.037.

[14] M. Larrañeta, C. Fernández-Peruchena, and M. A. Silva-Pérez, "'Industrial Application of Synthetic Irradiance: Case Study of Solar Yield,'" in *Synthetic Solar Irradiance: Modeling Solar Data, edited by Jamie M. Bright (AIP Publishing, Melville, New York, 2021), Chapter 5, pp. 5-1–5-34*, 2021.

[15] M. Larrañeta, E. López-García, J.-V. Paola, and M. A. Silva-Pérez, "Use of Multiyear Tool : An Open Tool for the Synthetic Generation of Plausible Meteorological Years of Coupled GHI and DNI with 1-min Resolution," *SolarPaces Conf.*, pp. 1–8, 2021.

[16] A. P. Grantham, P. J. Pudney, and J. W. Boland, "Generating synthetic sequences of global horizontal

irradiation," *Sol. Energy*, vol. 162, no. November 2017, pp. 500–509, 2018, doi: 10.1016/j.solener.2018.01.044.

[17] M. Larrañeta, C. Cantón-Marín, M. A. Silva-Pérez, and I. Lillo-bravo, "Use of the ND Tool: An Open Tool for the Synthetic Generation of 1-min Solar Data from Hourly Means with Geographic Flecibility," *SolarPaces Conf.*, 2019.

[18] M. Larrañeta, C. Fernandez-Peruchena, M. A. Silva-Pérez, and I. Lillo-Bravo, "Methodology to synthetically downscale DNI time series from 1-h to 1-min temporal resolution with geographic flexibility," *Sol. Energy*, vol. 162, no. January, pp. 573–584, 2018, doi: 10.1016/j.solener.2018.01.064.

[19] Bureau of Meteorology, "One minute Solar Data," *Bur. Meteorol.*, 2015, [Online]. Available: http://www.bom.gov.au/climate/data/oneminsolar/stations.shtml.

[20] S. Moreno-Tejera, M. A. Silva-Pérez, I. Lillo-Bravo, and L. Ramírez-Santigosa, "Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions," *Sol. Energy*, vol. 132, pp. 430–441, 2016, doi: 10.1016/j.solener.2016.03.032.

[21] John Boland, "Time Series Modelling of Solar Radiation," in *Modeling Solar Radiation at the Earth s Surface*, 2008, pp. 283–312.

[22] S. Moreno-Tejera, M. A. Silva-Pérez, L. Ramírez-Santigosa, and I. Lillo-Bravo, "Classification of days according to DNI profiles using clustering techniques," *Sol. Energy*, vol. 146, pp. 319–333, 2017, doi: 10.1016/j.solener.2017.02.031.

[23] "'Multiyear Open Tool.'" http://estacionmeteo.us.gter.es/multi_year/validar_datos.

[24] CTN 206/SC 117-CENTRALES TERMOSOLARES UNE 206011:2014., "Centrales termosolares. Procedimiento de generación de Año Solar Representativo.," 2014.

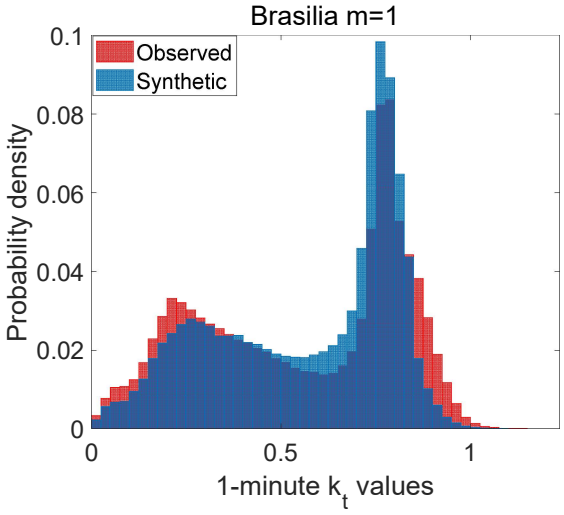
[25] J. Polo *et al.*, "Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets," *Sol. Energy*, vol. 132, pp. 25–37, 2016, doi: 10.1016/j.solener.2016.03.001.

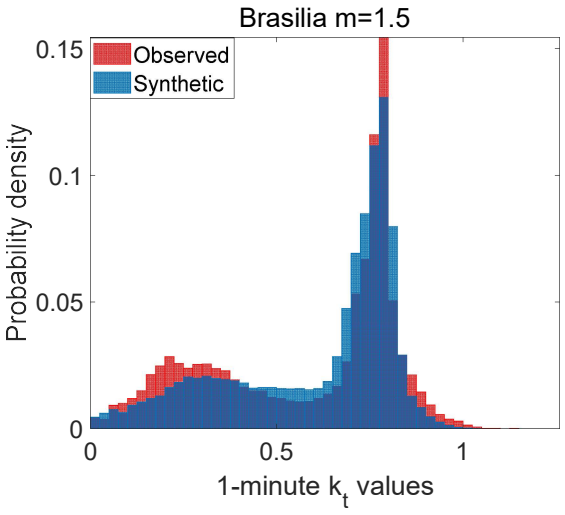
[26] J. Tovar, F. J. Olmo, and L. Alados-Arboledas, "One-minute global irradiance probability density distributions conditioned to the optical air mass," *Sol. Energy*, vol. 62, no. 6, pp. 387–393, 1998, doi: 10.1016/S0038-092X(98)00035-8.

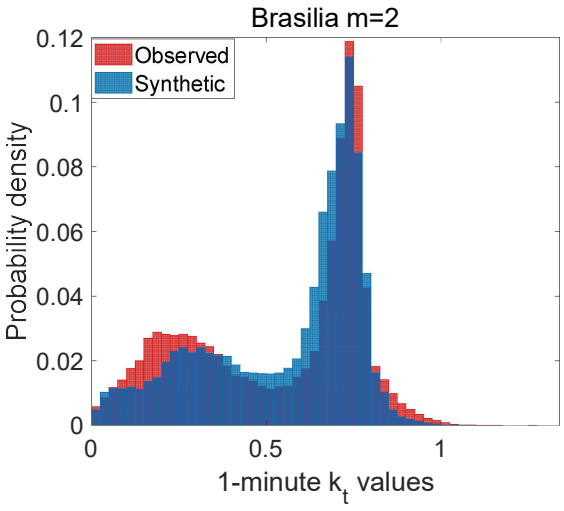
[27] C. M. Fernández-Peruchena and A. Bernardos, "A comparison of one-minute probability density distributions of global horizontal solar irradiance conditioned to the optical air mass and hourly averages in different climate zones," *Sol. Energy*, vol. 112, pp. 425–436, 2015, doi: 10.1016/j.solener.2014.11.030.

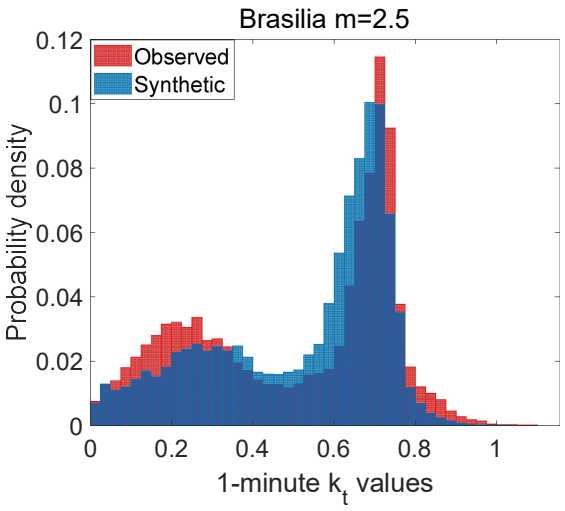
ANNEX A

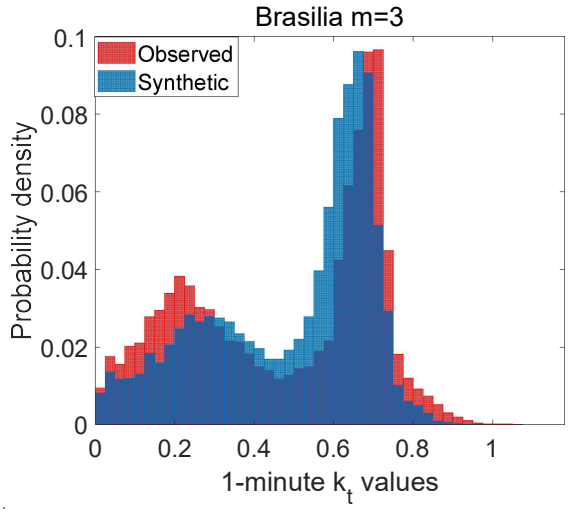
In this section we show the figures studied for the 1-minute kt distribution and 1-minute kb distribution for all m for all locations.

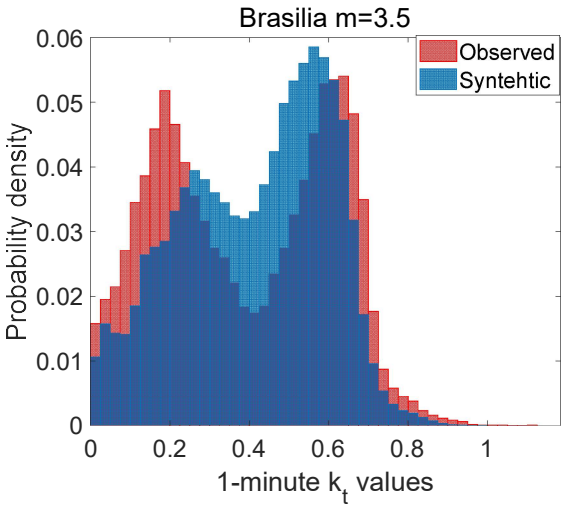


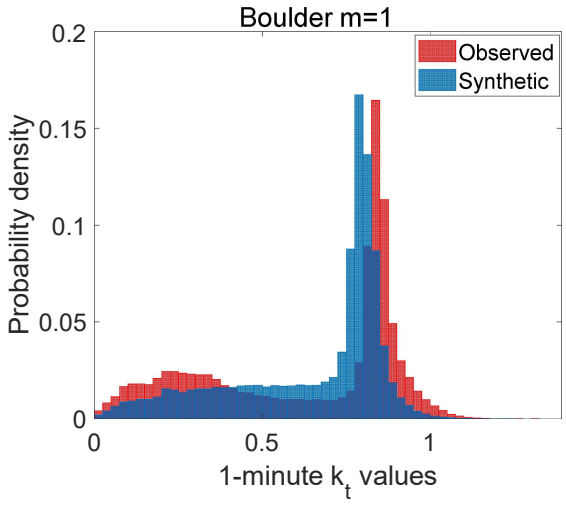


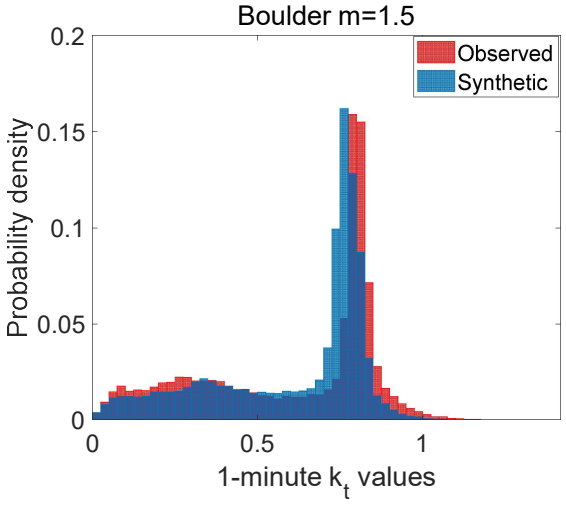


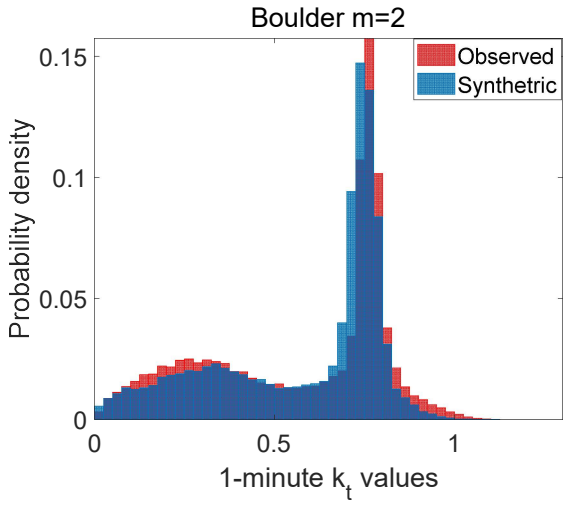


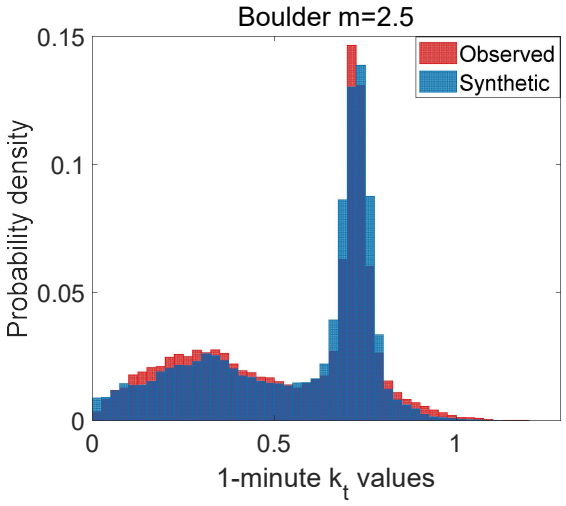


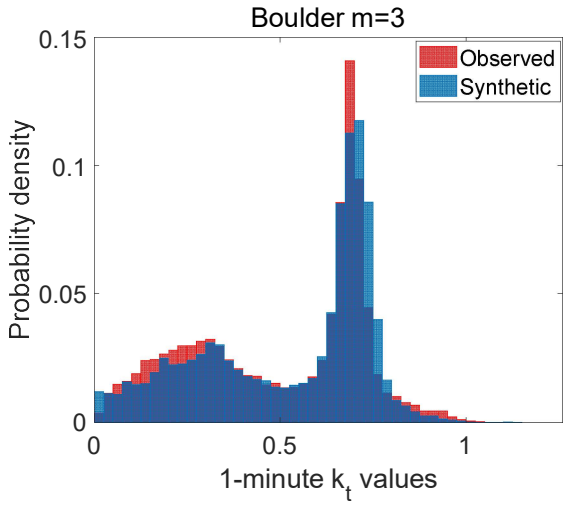


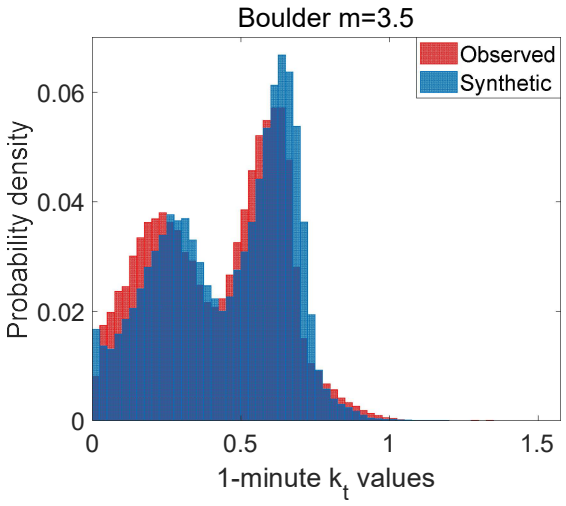

Boulder m=1
Observed
Synthetic
Probability density
0.2
0.15
0.1
0.05
0
0
0.5
1
1-minute k_t values

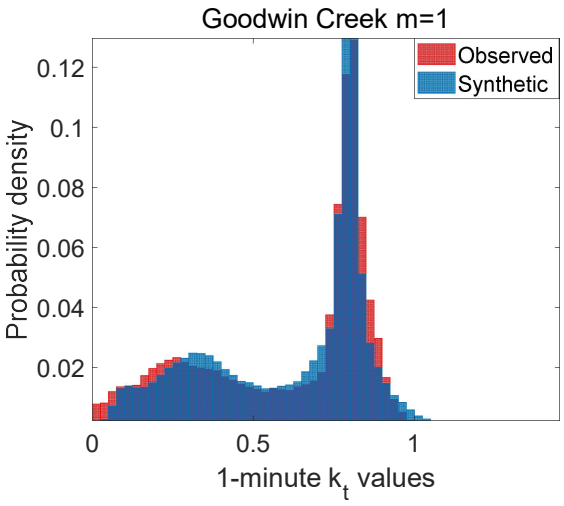

Boulder m=1.5
Observed
Synthetic
Probability density
0.2
0.15
0.1
0.05
0
0
0.5
1
1-minute k_t values

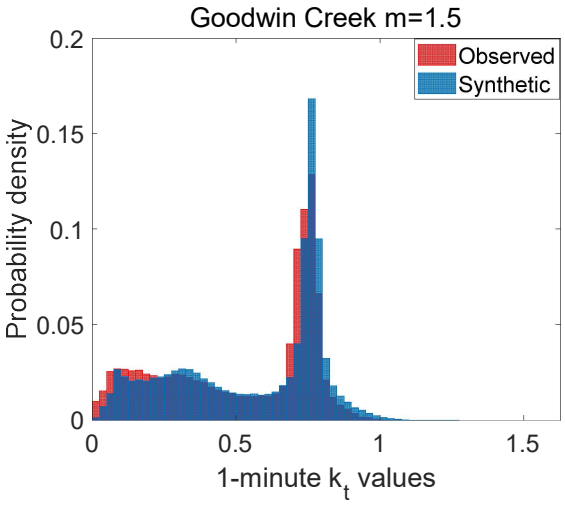

Boulder m=2
Observed
Synthetic
Probability density
0.15
0.1
0.05
0
0
0.5
1
1-minute k_t values

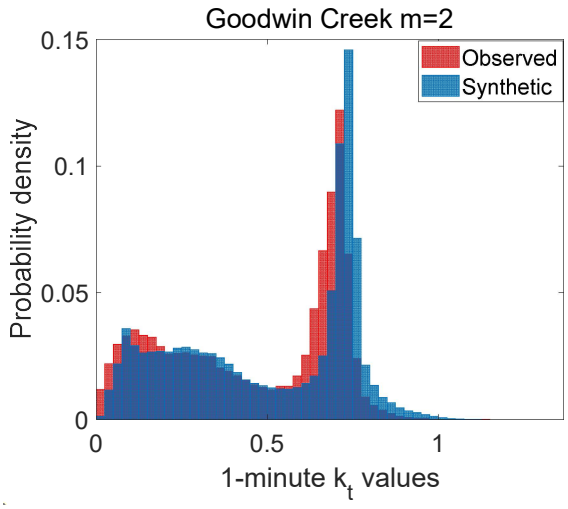

Boulder m=2.5
Observed
Synthetic
Probability density
0.15
0.1
0.05
0
0
0.5
1
1-minute k_t values

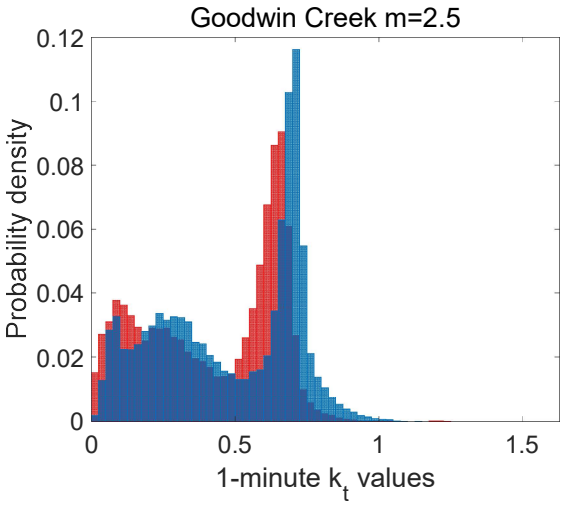

Boulder m=3
Observed
Synthetic
Probability density
0.15
0.1
0.05
0
0
0.5
1
1-minute k_t values

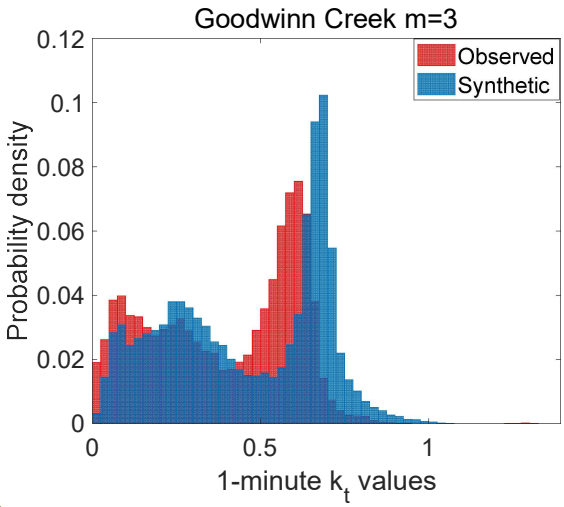

Boulder m=3.5
Observed
Synthetic
Probability density
0.06
0.04
0.02
0
0
0.5
1
1.5
1-minute k_t values

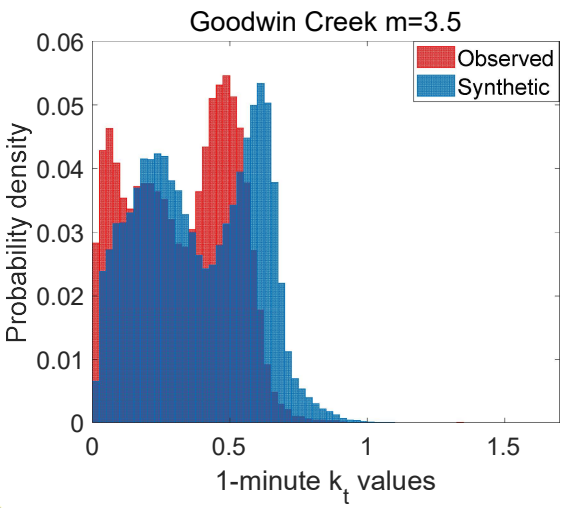

Goodwin Creek m=1
Observed
Synthetic
Probability density
0.12
0.1
0.08
0.06
0.04
0.02
0
0.5
1
1-minute k_t values

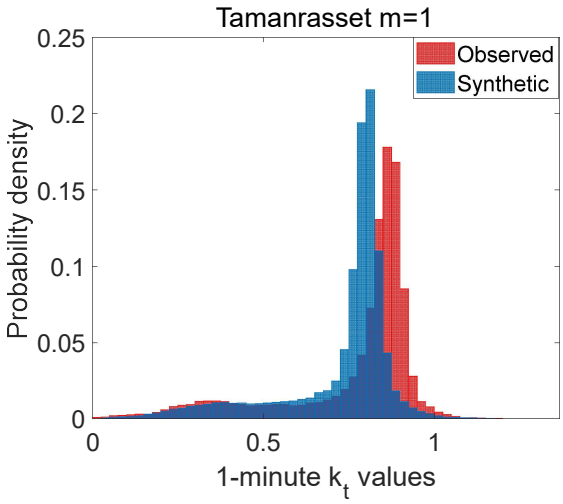

Goodwin Creek m=1.5
Observed
Synthetic
Probability density
0.2
0.15
0.1
0.05
0
0.5
1
1.5
1-minute k_t values

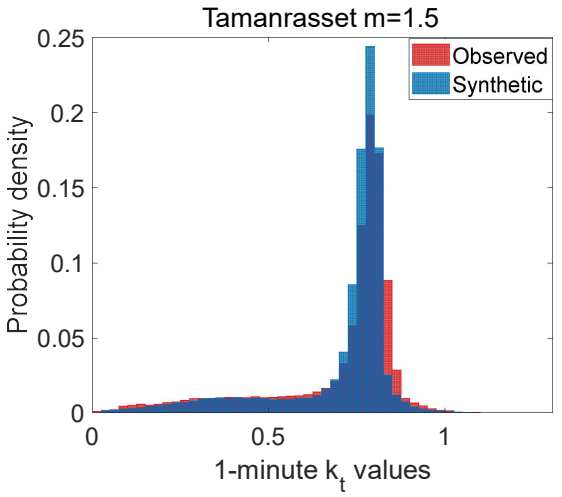

Goodwin Creek m=2
Observed
Synthetic
Probability density
0.15
0.1
0.05
0
0.5
1
1-minute k_t values

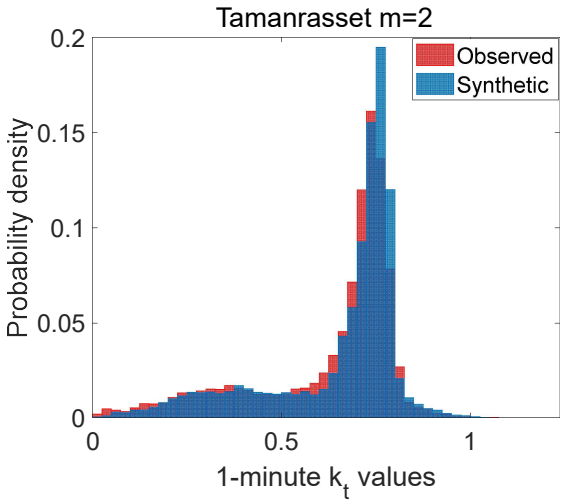

Goodwin Creek m=2.5
Observed
Synthetic
Probability density
0.12
0.1
0.08
0.06
0.04
0.02
0
0.5
1
1.5
1-minute k_t values

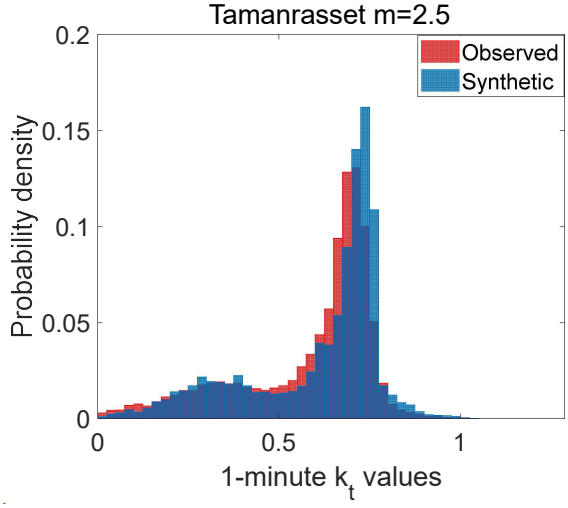

Goodwinn Creek m=3
Observed
Synthetic
Probability density
0.12
0.1
0.08
0.06
0.04
0.02
0
0.5
1
1-minute k_t values

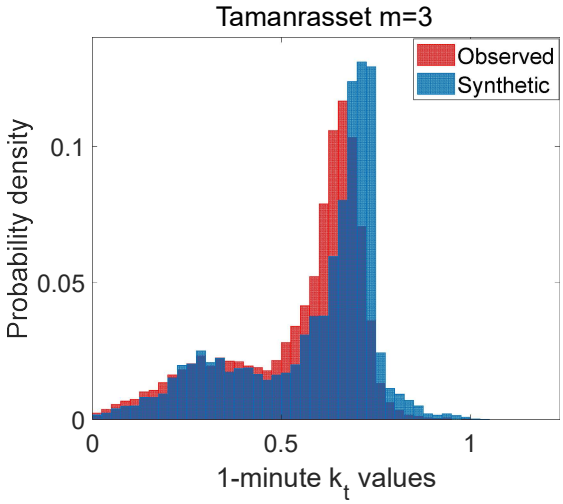

Goodwin Creek m=3.5
Observed
Synthetic
Probability density
0.06
0.05
0.04
0.03
0.02
0.01
0
0.5
1
1.5
1-minute k_t values

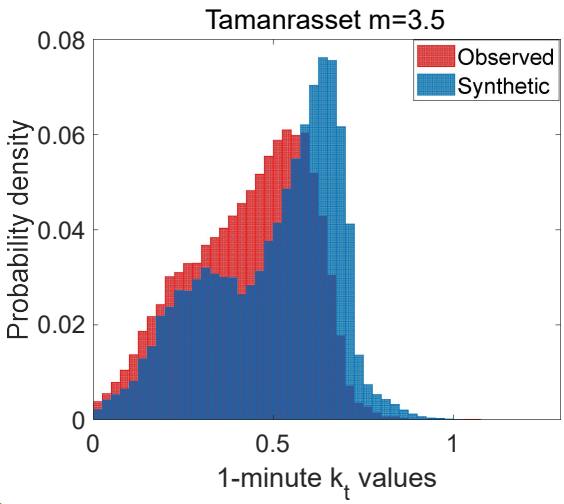

Tamanrasset m=1
Observed
Synthetic
Probability density
0.25
0.2
0.15
0.1
0.05
0
0
0.5
1
1-minute k_t values

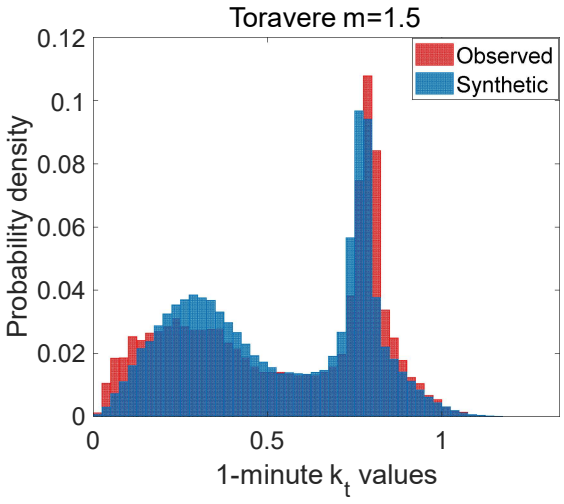

Tamanrasset m=1.5
Observed
Synthetic
Probability density
0.25
0.2
0.15
0.1
0.05
0
0
0.5
1
1-minute k_t values

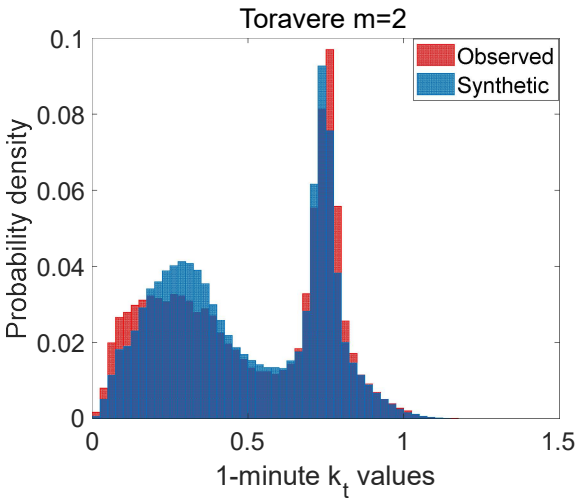

Tamanrasset m=2
Observed
Synthetic
Probability density
0.2
0.15
0.1
0.05
0
0
0.5
1
1-minute k_t values

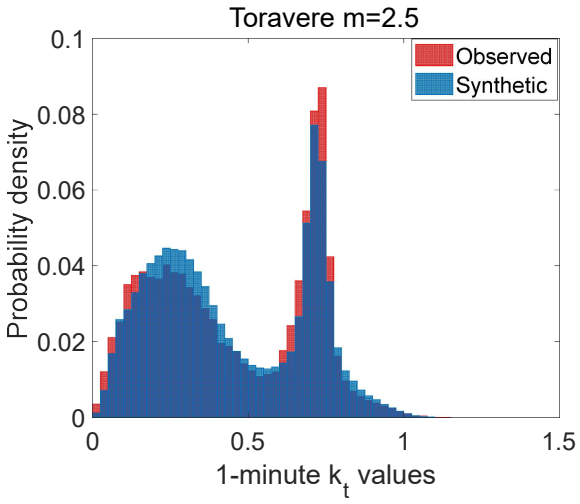

Tamanrasset m=2.5
Observed
Synthetic
Probability density
0.2
0.15
0.1
0.05
0
0
0.5
1
1-minute k_t values

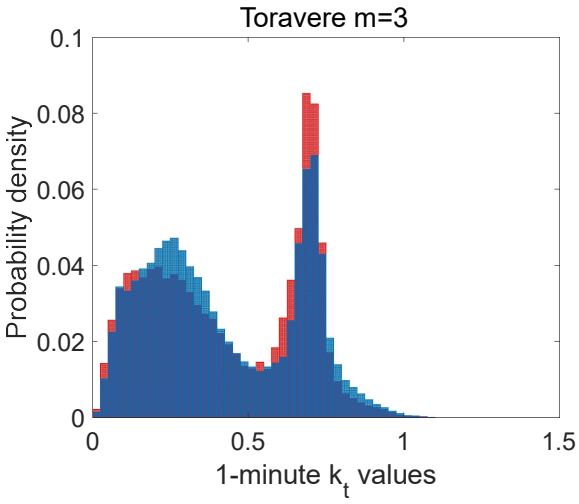

Tamanrasset m=3
Observed
Synthetic
Probability density
0.1
0.05
0
0
0.5
1
1-minute k_t values

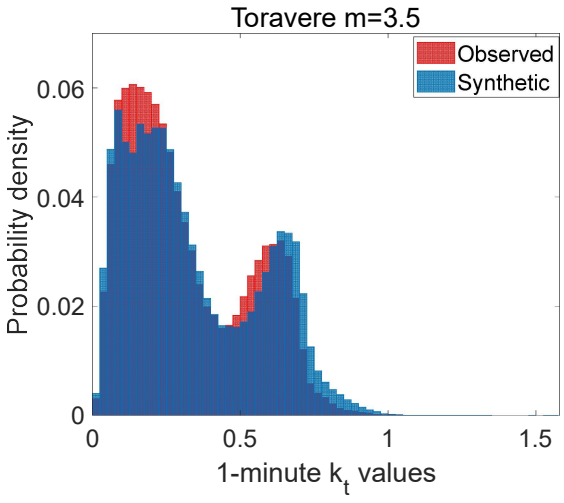

Tamanrasset m=3.5
Observed
Synthetic
Probability density
0.08
0.06
0.04
0.02
0
0
0.5
1
1-minute k_t values

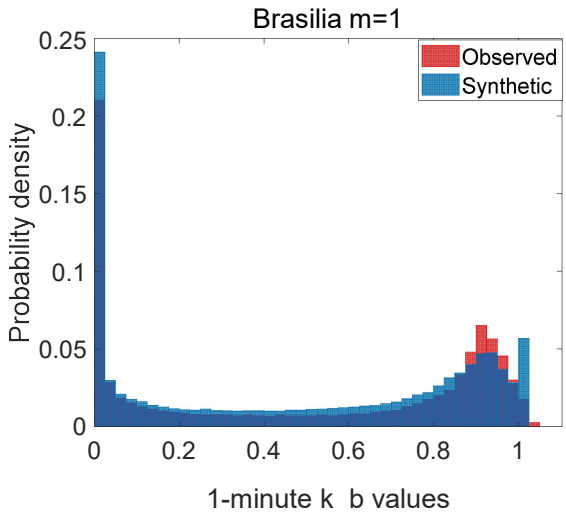

Toravere m=1.5
Observed
Synthetic
Probability density
0.12
0.1
0.08
0.06
0.04
0.02
0
0
0.5
1
1-minute k_t values

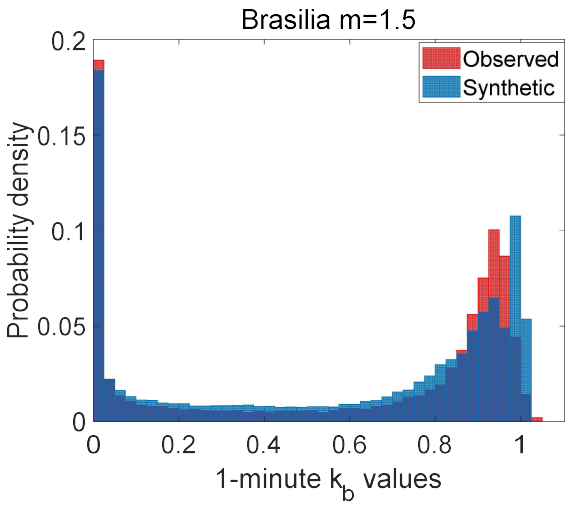

Toravere m=2
Observed
Synthetic
Probability density
0.1
0.08
0.06
0.04
0.02
0
0
0.5
1
1.5
1-minute k_t values

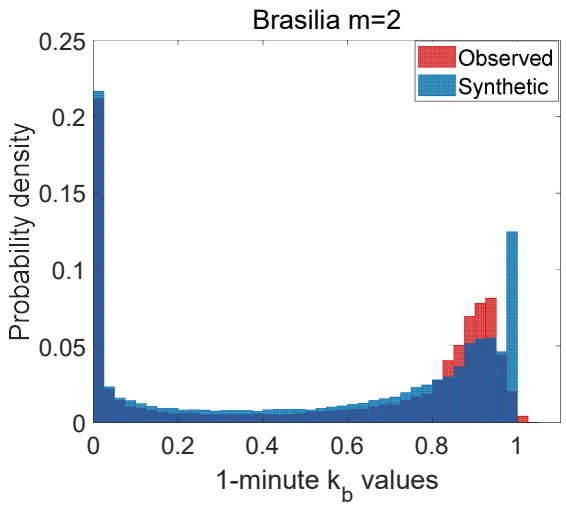

Toravere m=2.5
Observed
Synthetic
Probability density
0.1
0.08
0.06
0.04
0.02
0
0
0.5
1
1.5
1-minute k_t values

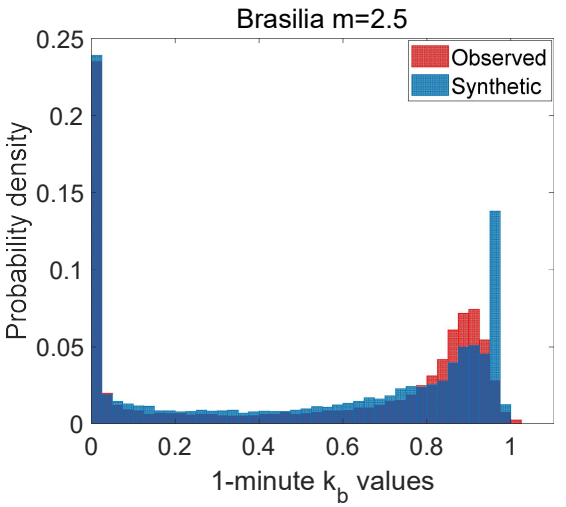

Toravere m=3
Observed
Synthetic
Probability density
0.1
0.08
0.06
0.04
0.02
0
0
0.5
1
1.5
1-minute k_t values

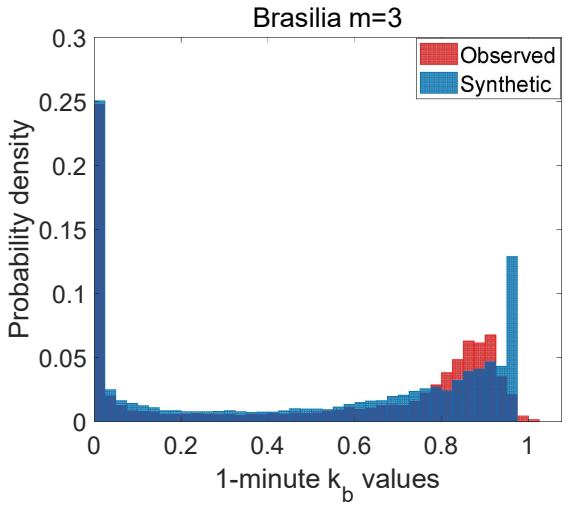

Toravere m=3.5
Observed
Synthetic
Probability density
0.06
0.04
0.02
0
0
0.5
1
1.5
1-minute k_t values

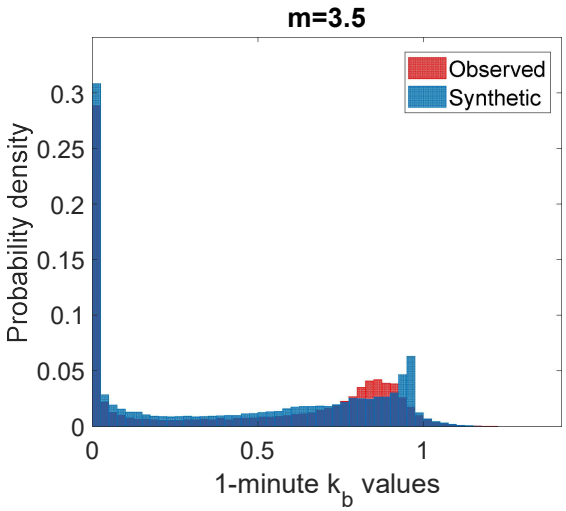

Brasilia m=1
Observed
Synthetic
Probability density
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
1-minute k b values

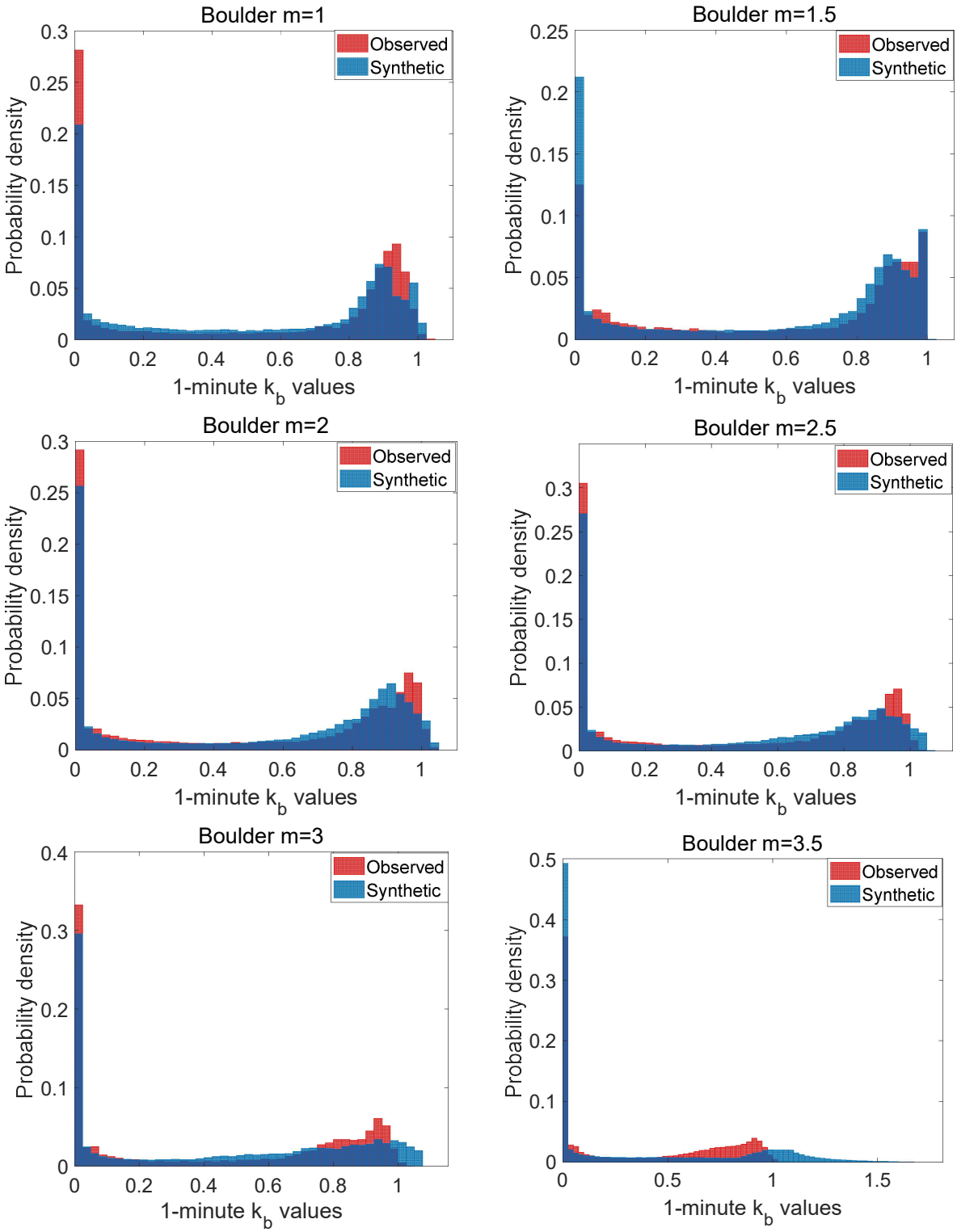

Brasilia m=1.5
Observed
Synthetic
Probability density
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
1-minute k b values

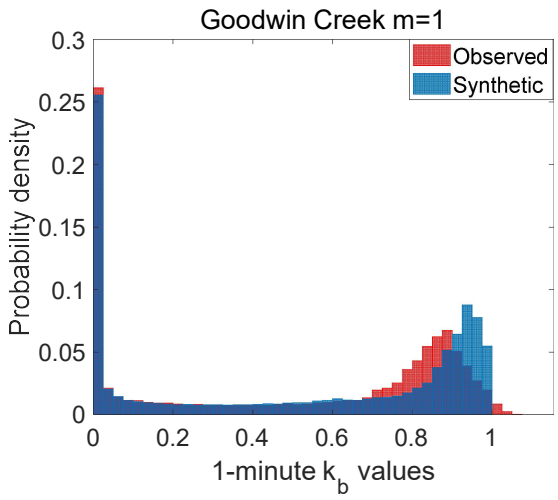

Brasilia m=2
Observed
Synthetic
Probability density
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
1-minute k b values

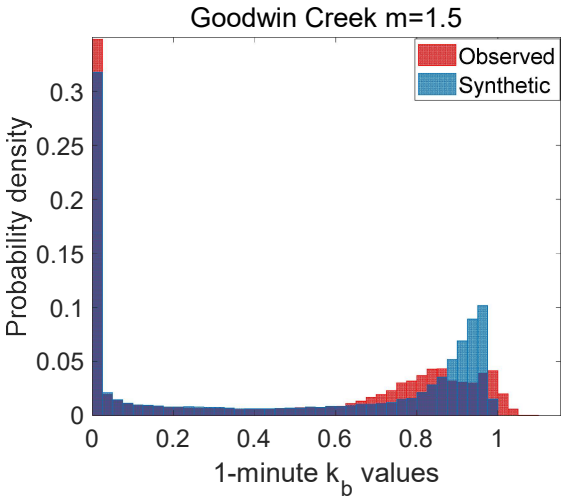

Brasilia m=2.5
Observed
Synthetic
Probability density
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
1-minute k b values

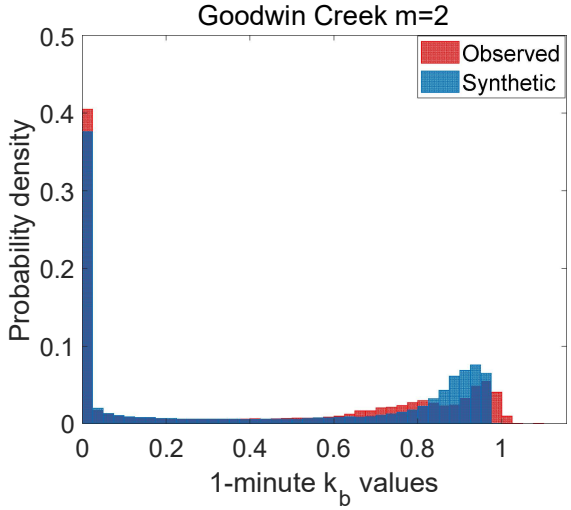

Brasilia m=3
Observed
Synthetic
Probability density
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
1-minute k b values

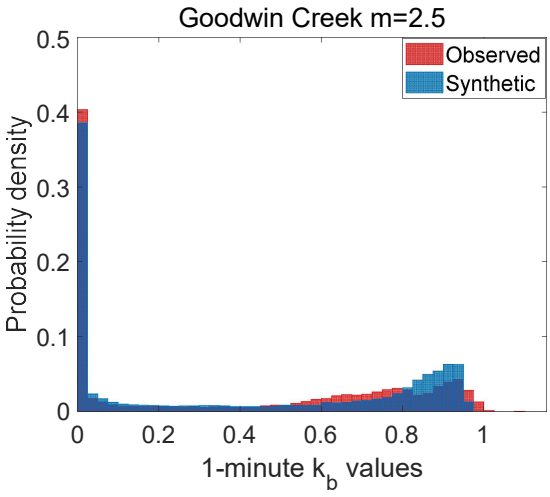

m=3.5
Observed
Synthetic
Probability density
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.5
1
1-minute k b values

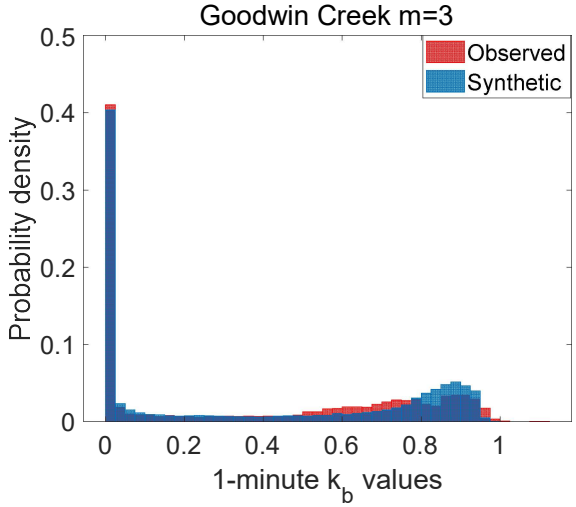

Boulder m=1
Boulder m=1.5
Boulder m=2
Boulder m=2.5
Boulder m=3
Boulder m=3.5
Observed
Synthetic
Probability density
1-minute k_b values

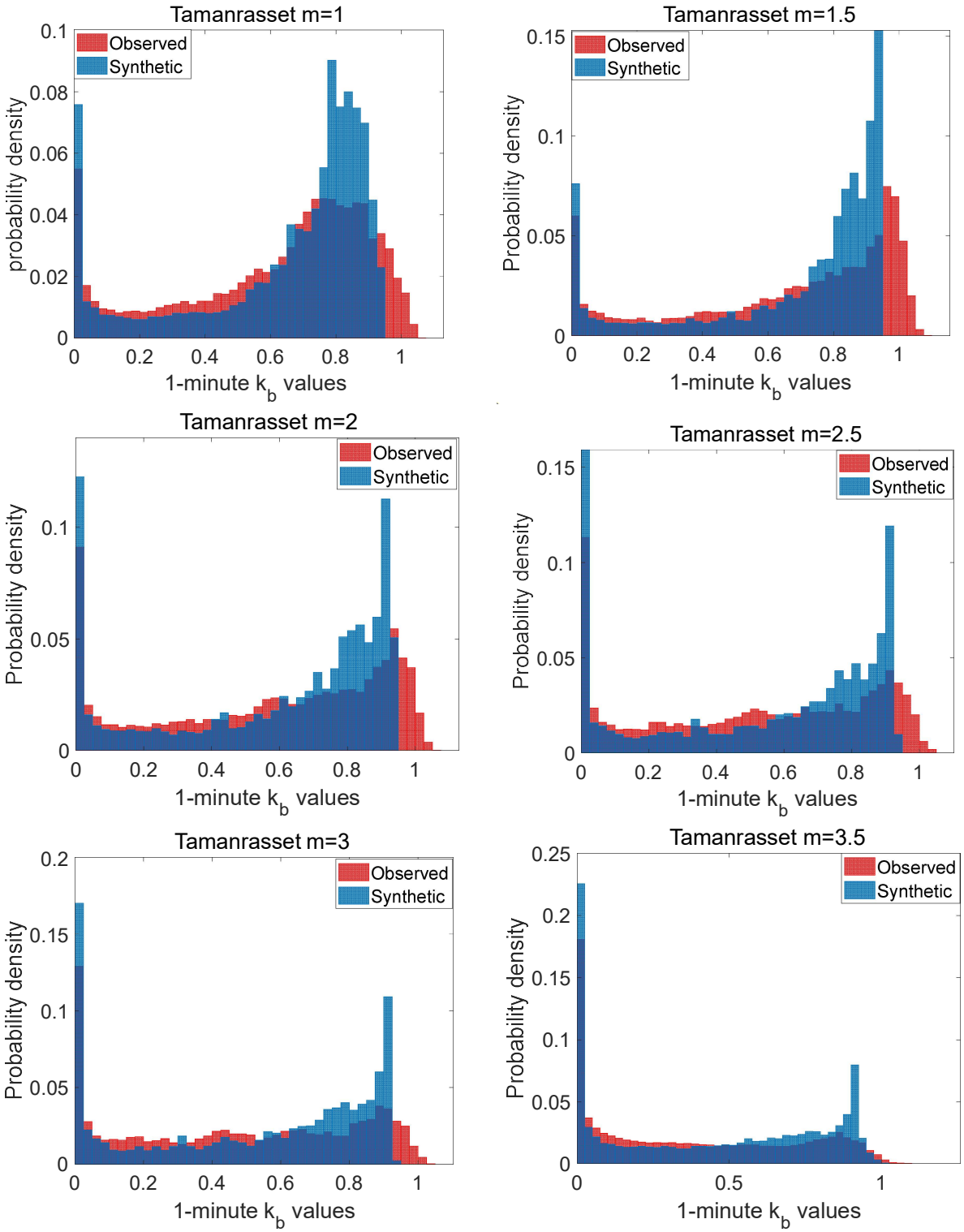

Goodwin Creek m=1
Observed
Synthetic
Probability density
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
1-minute k_b values

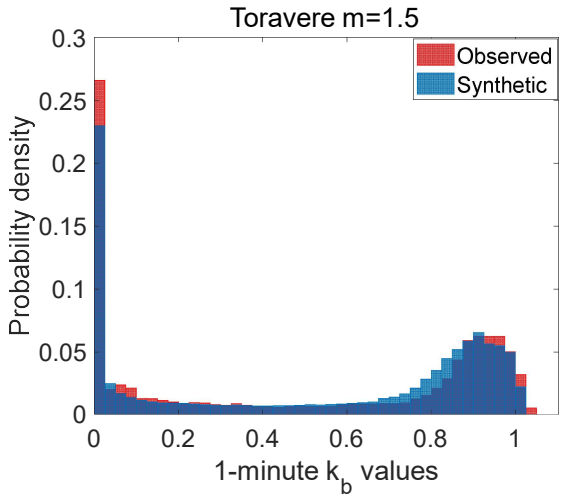

Goodwin Creek m=1.5
Observed
Synthetic
Probability density
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
1-minute k_b values

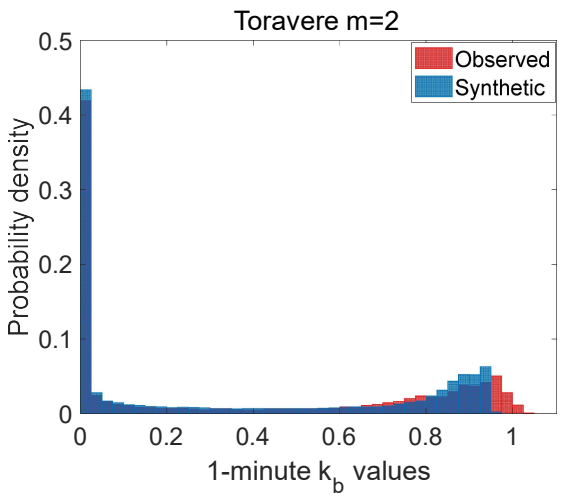

Goodwin Creek m=2
Observed
Synthetic
Probability density
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1-minute k_b values

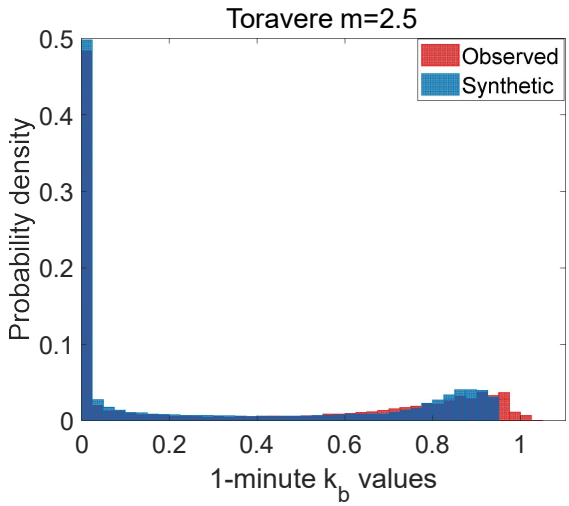

Goodwin Creek m=2.5
Observed
Synthetic
Probability density
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1-minute k_b values

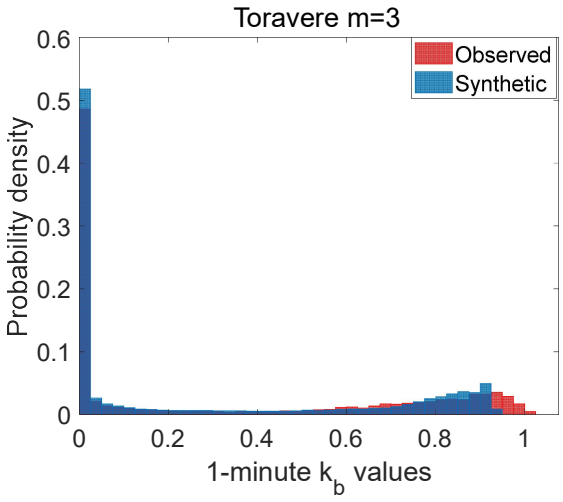

Goodwin Creek m=3
Observed
Synthetic
Probability density
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1-minute k_b values

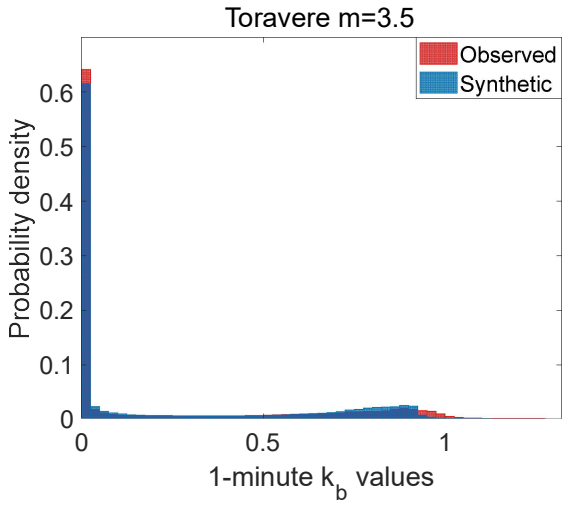

Tamanrasset m=1
Observed
Synthetic
probability density
1-minute k_b values
Tamanrasset m=1.5
Probability density
Tamanrasset m=2
Tamanrasset m=2.5
Tamanrasset m=3
Tamanrasset m=3.5


Toravere m=1.5
Observed
Synthetic
Probability density
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
1-minute k_b values


Toravere m=2
Observed
Synthetic
Probability density
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1-minute k_b values


Toravere m=2.5
Observed
Synthetic
Probability density
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1-minute k_b values


Toravere m=3
Observed
Synthetic
Probability density
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1-minute k_b values


Toravere m=3.5
Observed
Synthetic
Probability density
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.5
1
1-minute k_b values

RESUMEN DEL PROYECTO

Synthetic generation of Plausible Solar Years for long-term forecasting of solar radiation.

Generación sintética de Años Solares Plausibles para la predicción a largo plazo de la radiación solar.

Keywords:

Synthetic sequences; Solar variability; Plausible Meteorological Years (PMYs); Multiyear.

Palabras clave:

Secuencias sintéticas; Variabilidad solar; Años solares plausibles; Múltiples series anuales.

Resumen

El diseño y desarrollo de sistemas de energía solar requiere una aproximación del rendimiento de la planta y de su viabilidad económica en condiciones normales de funcionamiento. Para ello, se utilizan series temporales de irradiación solar y otras variables meteorológicas que permiten representar las condiciones normales de funcionamiento. En general, estas series de datos se basan en la definición de series anuales representativas que son adecuadas para la evaluación detallada del rendimiento de los sistemas solares. Actualmente, la mayoría de los métodos utilizados para representar las condiciones normales de funcionamiento en los sistemas de simulación se basan en la selección del año más representativo. Se trata de un conjunto meteorológico de un año representativo de las condiciones "típicas", definido como Año Meteorológico Típico (TMY) que representa un año de condiciones climáticas típicas para un lugar determinado. Desde su concepto inicial, se han propuesto variamos modificaciones del TMY para responder a las necesidades de otras tecnologías, pero el procedimiento inicial no ha cambiado significativamente. El principal punto débil del método TMY es que un único año representativo no permite representar las condiciones extremas ni caracterizar la variabilidad del recurso solar en detalle. Esto supone un problema, ya que la vida útil de estas centrales es de 25 a 30 años y un año típico no representa todos los posibles escenarios futuros.

Debido al coste considerable que tienen las centrales solares, es importante conocer la energía que generará un determinado sistema en función de su ubicación. Por lo tanto, la caracterización del recurso solar en una ubicación específica es clave para las decisiones de política energética, el diseño de ingeniería y el desarrollo de sistemas de energía solar. El uso de series representativas de irradiación solar y variables meteorológicas no es suficiente ya que son series de valores promediados que proporcionan una caracterización incompleta del clima. Para caracterizar el recurso solar se requiere un estudio detallado de la distribución de frecuencias y de la variabilidad en diferentes escalas de tiempo.

La incertidumbre relacionada con la variabilidad del recurso solar juega un papel importante en la estimación de la probabilidad del rendimiento energético futuro de una planta solar y sus condiciones financieras. Los promotores y prestamistas, además de la serie anual del TMY, necesitan información relacionada con la amortización anual del proyecto en años extremos [8]. Por esta razón, el TMY puede no es lo suficientemente adecuado para diseñar sistemas que satisfagan las peores condiciones climáticas que puedan darse en un lugar. Se ha observado que la magnitud de los cambios de variabilidad depende claramente de la zona geográfica. La evaluación de la variabilidad espacio-temporal a largo plazo puede tener repercusiones importantes en el diseño, la capacidad financiera y el funcionamiento de los sistemas de conversión de energía solar que utilizan tecnologías como la fotovoltaica o la CSP. Además, el estudio de la variabilidad del recurso solar en diferentes tipos de

clima proporciona información valiosa sobre la vida útil de los materiales solares en diferentes lugares y sobre la selección del emplazamiento durante la fase de prefactibilidad de los proyectos solares que utilizan estas tecnologías.

La evaluación del recurso solar es importante para comprender la variabilidad solar a largo plazo en una región o ubicación determinada. Las series de datos de radiación solar deben ser precisas y fiables para minimizar el riesgo relacionado con la incertidumbre de las series de datos disponibles. Estas series de datos permiten planificar los recursos, mejorar la modelización de los costes de producción, así como apoyar a la industria solar reduciendo los costes de desarrollo de los proyectos solares, por lo que lo ideal es disponer de una larga serie de datos de radiación medidos con precisión en una localización, aunque esto no es lo habitual. La existencia y disponibilidad de series temporales de radiación solar derivadas de satélites permite entender la dinámica de la radiación solar en diferentes escalas de tiempo para modelar la capacidad de los proyectos y reducir el riesgo financiero.

En la actualidad, es cada vez más importante el enfoque probabilístico para modelar la variabilidad y las incertidumbres del sistema de energía solar. Estos métodos probabilísticos permiten modelar los cambios y evaluar las probabilidades de que la métrica simulada se sitúe por encima o por debajo de un valor determinado o dentro de un rango específico. En este escenario, cobra importancia la generación de series de datos solares sintéticos para ampliar la información relacionada con la variabilidad de la radiación solar. La ventaja de las series de datos sintéticos es que permiten representar el comportamiento de la radiación solar en un gran rango de escenarios y facilita la evaluación del riesgo de los proyectos solares.

Por esta razón, la generación del Año Meteorológico Plausible adquiere mayor importancia. Los años meteorológicos plausibles (PMYs) se definen como series anuales de alta frecuencia de la Irradiancia Horizontal Global (GHI), la Irradiancia Normal Directa (DNI) y otras variables meteorológicas relevantes (temperatura, humedad relativa y velocidad del viento, entre otras), que preservan la variabilidad natural del recurso en un lugar específico. Los años solares plausibles (PSYs) son una simplificación de los PMYs donde sólo se generan datos solares. Los métodos para la generación sintética de los PMYs utilizan grandes series de datos que podrían proporcionar los proveedores de datos satelitales para la mayoría de los lugares del mundo. Estas series tienen en cuenta la variabilidad observada del recurso solar en el lugar, asumiendo que las tendencias futuras pueden caracterizarse por las tendencias pasadas. La generación de PMYs permite la simulación estocástica de un sistema de aprovechamiento solar teniendo en cuenta la incertidumbre y la variabilidad inherentes al recurso solar.

Los PMYs tienen grandes ventajas sobre los TMY, ya que estos conjuntos de datos consideran un rango más amplio de escenarios de radiación solar preservando su variabilidad inherente anual, mensual, diaria e intradiaria. El uso de estos PMYs como entrada en la evaluación del rendimiento de los sistemas solares permite obtener una descripción probabilística del rendimiento del sistema y una información valiosa para la fase de financiación del proyecto.

Larrañeta et al., (2019) presentan un modelo para la generación sintética de años solares plausibles de alta resolución. El algoritmo es adecuado para cualquier lugar en el que se disponga de conjuntos de datos derivados de satélites. El proceso de generación sintética se ha automatizado y se ha desarrollado recientemente una herramienta abierta. La metodología no necesita ninguna adaptación o calibración local, pero una menor incertidumbre en el conjunto de datos permitirá obtener mejores resultados.

En este trabajo mejoramos el método presentado por Larrañeta et al. (2019) y lo probamos en 5 localizaciones con climas diferentes Generamos 100 conjuntos anuales de datos solares acoplados de 1 minuto de DNI y GHI para estas localizaciones utilizando datos derivados del satélite SolarGIS y los comparamos con el comportamiento de las series de datos observados de cada localización. Los conjuntos de datos profesionales de SolarGIS proporcionan entre 20 y 25 conjuntos anuales de 15 minutos de DNI y GHI acoplados y otras variables meteorológicas relevantes. Este método permite

obtener 100 PSYs que mantienen las características de las series originales, pero proporcionan un amplio rango de escenarios de radiación solar adecuados para la descripción probabilística de la evaluación del rendimiento de los sistemas solares. Este método proporciona una caracterización más profunda asociada a la variabilidad del recurso solar, permitiendo obtener mejoras en las condiciones financieras y fomentando la inversión en la planta solar.

El método utilizado para la generación sintética de múltiples series temporales anuales de radiación solar en diferentes lugares requiere un mínimo de 10 series temporales anuales no consecutivas de valores de DNI y GHI acoplados cada hora como entrada y proporciona 100 Años Solares Plausibles (PSYs) anuales del conjunto de datos de DNI y GHI acoplados de 1 minuto cuyos valores acumulados anuales corresponden a la Probabilidad de Excedencia anual de 1 a 100. Los datos de entrada pueden obtenerse a partir de observaciones, estimaciones por satélite o bases de datos de irradiación de reanálisis. El tamaño de la muestra de las observaciones debe ser suficiente para caracterizar estadísticamente la distribución anual y mensual de la radiación solar. Por esta razón, es aconsejable un periodo mínimo de 10 años, consecutivos o no, de conjuntos de datos observados. El método para la generación sintética consiste en la conjunción de tres pasos (Larrañeta et al., 2019).

En el primer paso, obtenemos 100 series anuales a escala mensual. En primer lugar, integramos los datos horarios observados en la resolución mensual y calculamos las funciones de distribución acumulativa (CDF) de cada mes. A continuación, aplicamos el método de la transformada integral de probabilidad para generar los datos solares sintéticos mensuales. Limitamos la generación sintética para mantener la relación entre el GHI y el DNI en un lugar determinado (para más detalles ver Larrañeta et al., 2019).

El método se aplica 10000 veces para cada mes del año. Los valores mensuales se pueden concatenar para obtener 10000 conjuntos acumulativos anuales de datos sintéticos en base mensual.

Los datos horarios del GHI y del DNI observados también se integran en sumas acumulativas anuales y luego se ajustan a una distribución Normal para estimar las probabilidades teóricas de superación. En este paso, realizamos una modificación del algoritmo original que suponía que el GHI anual sigue una distribución Normal pero el DNI anual sigue una distribución Weibull. Se supone que el DNI horario y subhorario sigue una distribución de Weibull pero aún no está claro si el DNI anual sigue una distribución de Weibull.

Al final del paso 1, buscamos y seleccionamos los 100 conjuntos más cercanos (de P1 a P100) dentro de los 10000 generados sintéticamente a los estimados ajustando una distribución normal a los conjuntos observados. En la Figura 1 presentamos los resultados de los cálculos del primer paso para la localidad de Sevilla.

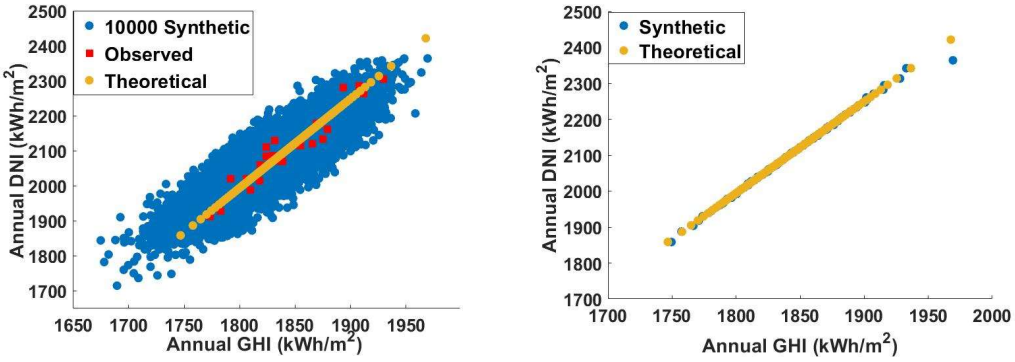


Figure 29 Resultados del primer paso en Sevilla.

En el segundo paso, reducimos la escala de los conjuntos de irradiación solar sintética de una

resolución temporal mensual a una diaria. Basándonos en (Grantham et al., 2018), el modelo diario consta de tres componentes: un componente estacional, uno autorregresivo (AR) y uno aleatorio. Para construir un modelo AR(1), primero realizamos una normalización de cada conjunto anual en la resolución diaria. Para ello, estimamos el componente estacional a partir del modelo de Fourier, este componente sólo depende del día del año, y luego lo dividimos por la desviación estándar del conjunto anual.

A continuación, construimos el modelo AR (1) para la serie estandarizada. El coeficiente AR (1) es de 0,34 y 0,42 en promedio para GHI y DNI para la localidad de Sevilla. El componente aleatorio o ruido blanco es la diferencia entre la serie estandarizada y el modelo AR (1).

A continuación, generamos valores diarios sintéticos a partir del modelo AR(1). El procedimiento de generación sintética para los datos diarios acoplados GHI+DNI sigue cuatro pasos principales:

1.Primero calculamos el componente aleatorio. Aplicamos una técnica de bootstrapping a las series de ruido blanco, esta técnica utiliza números aleatorios que se suponen probabilidades y calcula el valor estimado de esa probabilidad a partir de las FDC diarias de ruido blanco de cada mes. Mantenemos la relación entre el GHI y el DNI utilizando el mismo número aleatorio para cada par de valores diarios de radiación solar.

2.A continuación, calculamos el componente autorregresivo. El componente AR depende del valor de radiación solar anterior. Calculamos este componente multiplicando el valor anterior de un día determinado por el coeficiente AR(1) estimado para un lugar determinado.

3.En el último paso, procuramos la estacionalidad al conjunto sintético. Añadimos tanto los componentes aleatorios como los autorregresivos para cada día y luego deshacemos la normalización. Para ello, multiplicamos cada día por su correspondiente valor de desviación estándar y finalmente añadimos la contribución del modelo de Fourier.

Finalmente, para cada día sintético, buscamos el día observado con el valor más cercano en términos de energía (kt y kb) y le asignamos la misma información en términos de variabilidad y distribución mediante el índice de variabilidad (VI) y el índice de fracción matinal (Fm). Ejecutamos este procedimiento para la generación de 1000 conjuntos anuales de valores diarios. Agregamos los conjuntos sintéticos diarios en acumulados mensuales para seleccionar aquellos meses que se ajustan más a los obtenidos en el paso 1.

En el paso 3 generamos valores sintéticos de 1 minuto a partir de los cuartetos sintéticos diarios de kb, kt, VI y Fm obtenidos en el paso 2. En este paso, utilizamos la herramienta del modelo ND) para la generación sintética de datos GHI y DNI acoplados de 1 minuto. El modelo ND utiliza bases de datos adimensionales de una localización con clima similar a los datos de entrada; por tanto, puede ser utilizado en en localización sin adaptación local (Larrañeta et al., 2018). Los perfiles diarios de la base de datos adimensional están etiquetados en términos de energía (kt, kb), variabilidad (VI) y distribución (Fm). Generamos cuartetos diarios sintéticos de kb, kt, VI y Fm en el paso 2 que se utilizan para buscar y encontrar el día más similar entre la base de datos adimensional. Los perfiles adimensionales seleccionados se "abren" a valores reales multiplicando los valores de kt y kb de 1 minuto por la irradiancia extraterrestre y el DNI de cielo despejado de cada día correspondiente.

Se ha utilizado una base de datos adimensional para 1 implementación del algoritmo. Hemos seleccionado siete lugares con climas diferentes. Utilizamos los datos de GHI y DNI observados a 1 minuto obtenidos del sitio web de la Oficina Australiana de Meteorología (Bureau of Meteorology, 2015) para seis localizaciones y los conjuntos de datos de GHI y DNI observados a 1 minuto de la base de datos meteorológicos GTER) para la localización de Sevilla, España.

Se evalua el rendimiento del algoritmo en cinco lugares con climas diferentes. Los resultados se evalúan en cada paso del algoritmo: resolución anual, mensual, diaria y de 1 minuto. Comparamos los datos observados con los PSY sintéticos y con el año solar representativo (SRY) para cada ubicación. El SRY es una simplificación del TMY que utiliza únicamente datos solares para el cálculo del conjunto anual más representativo a largo plazo.

Para ello se utiliza una extensa base de datos derivadas de satélites (Tabla 1) para cinco ubicaciones como entrada del algoritmo multianual para la evaluación de los resultados. Las ubicaciones se han seleccionado con el objetivo de cubrir la mayoría de los climas significativos para los sistemas de aprovechamiento solar. Hemos utilizado las series temporales profesionales de Solargis climData, que consisten en series temporales de 15 minutos de datos solares y meteorológicos acoplados para un periodo de 20-25 años consecutivos.

Tabla 1 Emplazamientos seleccionados, period de años disponibles y clima Köppen-Geiger

Location (ID)	Country	Latitude (ºN)	Longitude (ºE)	Altitude (m)	Years	Köppen classification climate
Brasilia (BRB)	Brazil	-15.60	-47.71	1023	1999-2020	Aw
Boulder (BOU)	United States	40.13	-105.24	1689	1999-2020	Bsk
Tamanrasset (TAM)	Algeria	22.79	5.53	1385	1994-2020	Bwh
Goodwin Creek (GCR)	United States	34.25	-89.87	98	1999-2020	Cfa
Toravere (TOR)	Estonia	58.26	26.46	70	1994-2020	Dfb

Los resultados obtenidos para la resolución anual, mensual diaria y minutal se muestran a continuación:

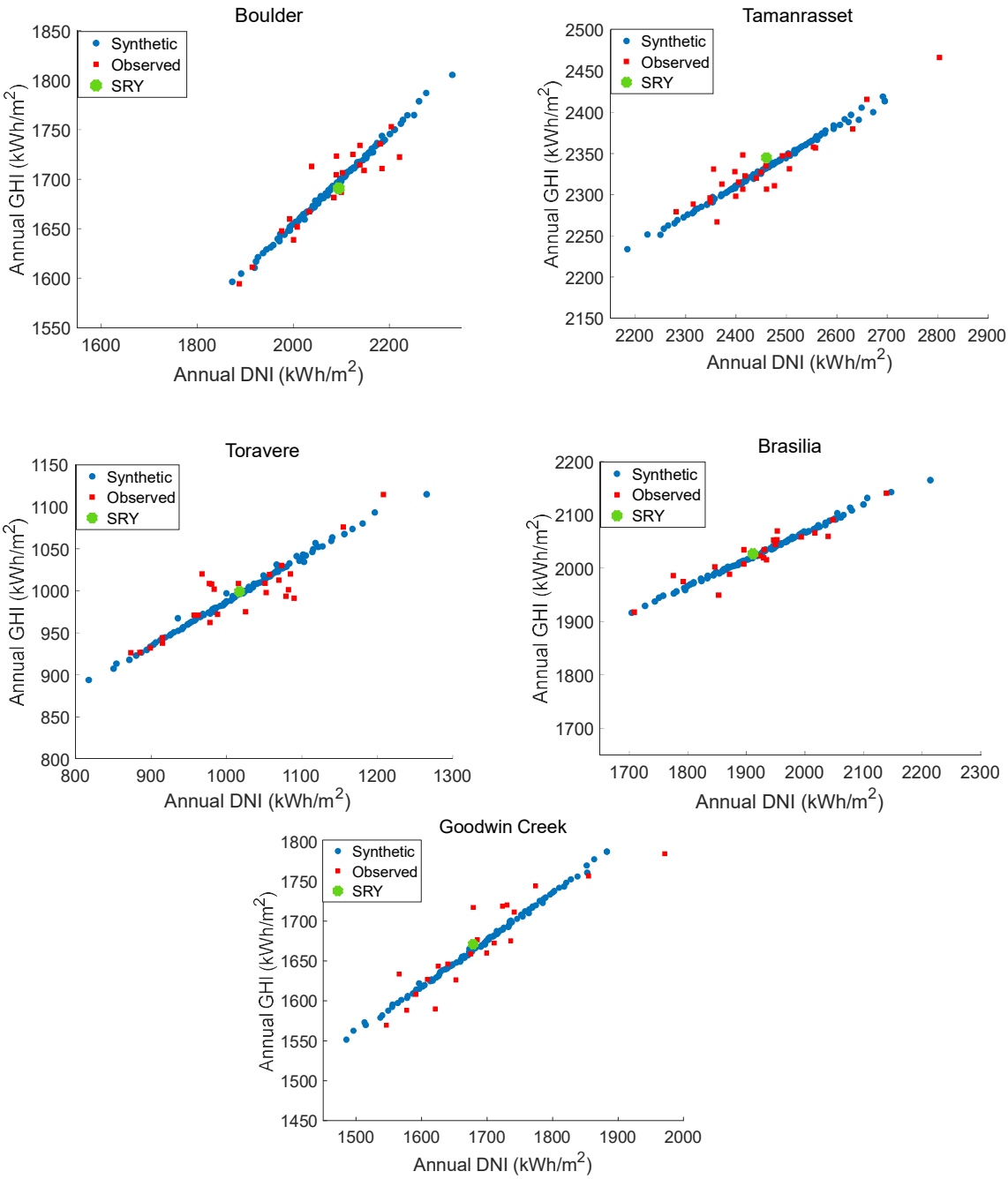


Figure 2 Gráfico de dispersión del GHI anual frente al DNI anual. Los conjuntos observados se presentan en puntos rojos, los sintéticos PSY en azul y los SRY

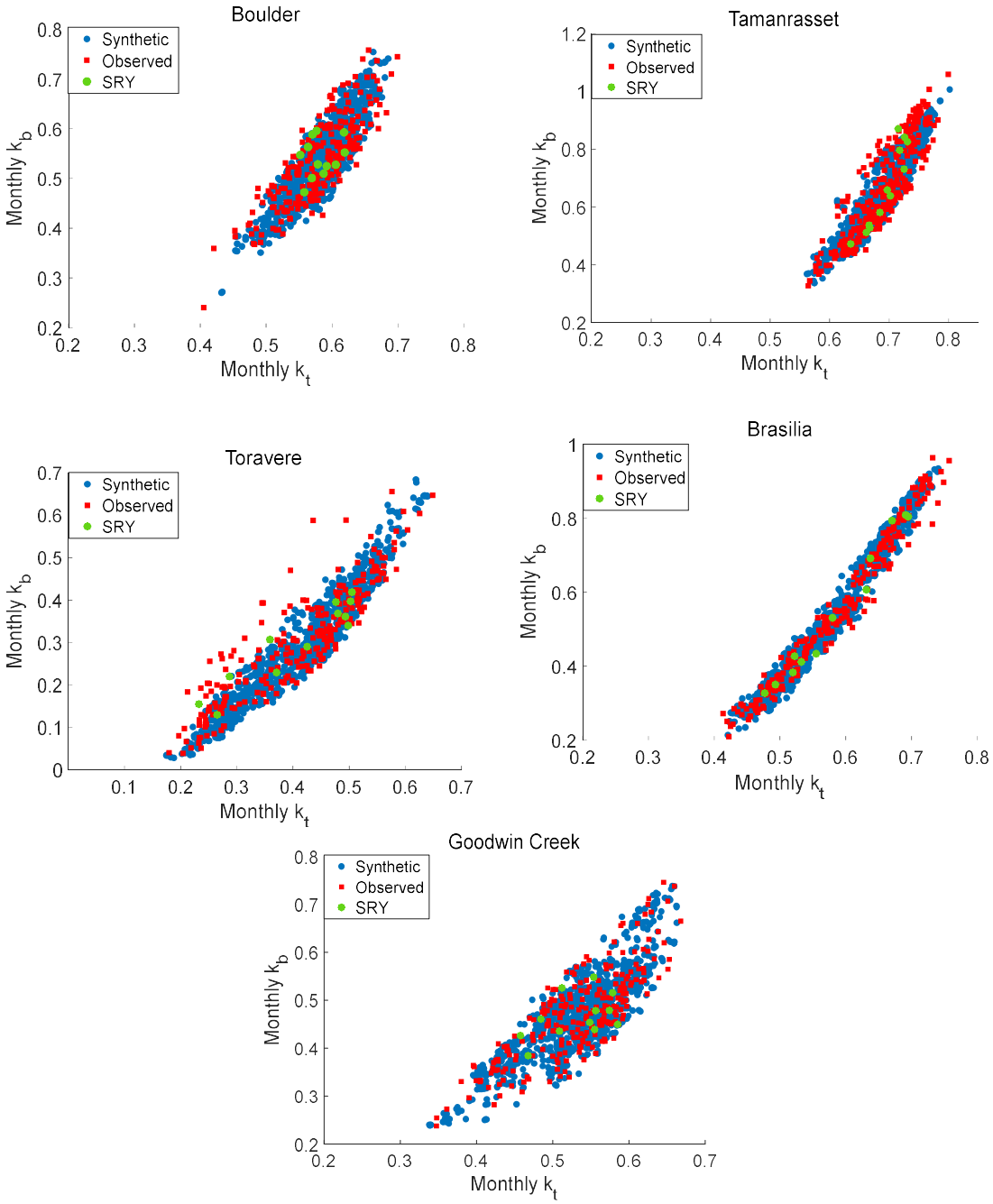


Figure 3 Gráfico de dispersión del índice de claridad mensual frente al índice de fracción directa. Los conjuntos observados se presentan en puntos rojos, los sintéticos PSY en azul y los SRY en verde.

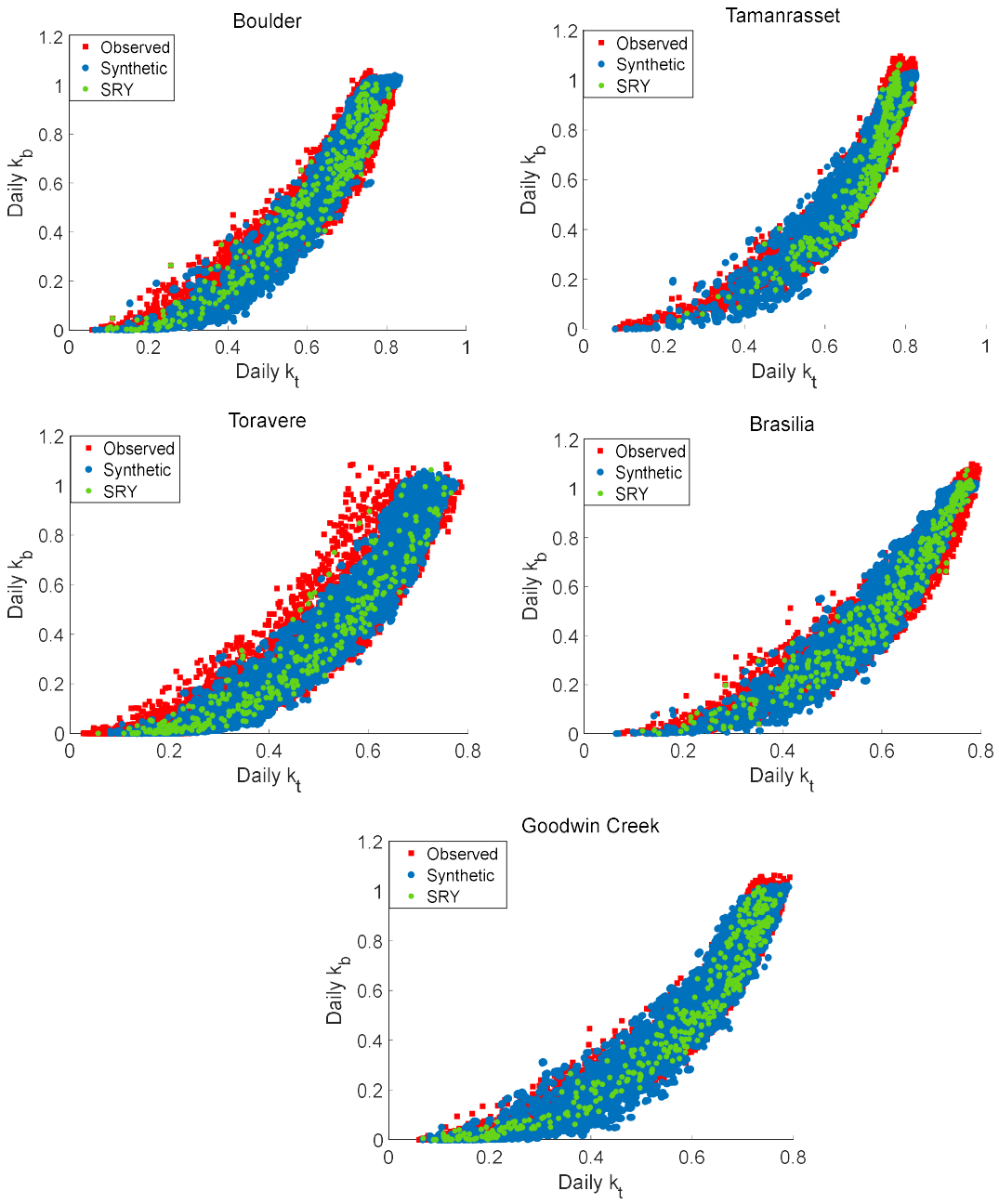


Figure 4. Gráfico de dispersión del índice de claridad diario frente al índice de fracción directa. Los conjuntos observados se presentan en puntos rojos, los PSY sintéticos en azul y los SRY en verde.

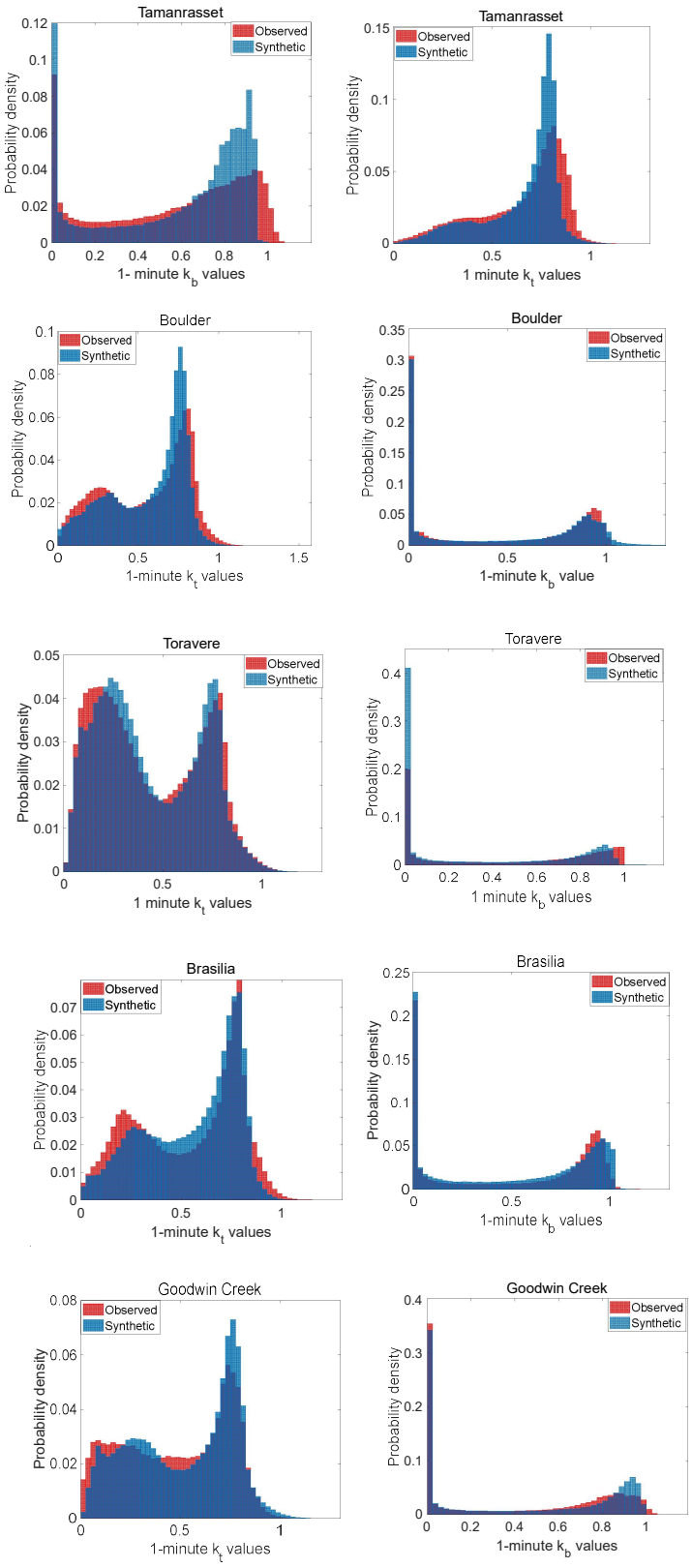


Figure 5 Distribución de la densidad de probabilidad de los valores kt y kb de 1 minuto. Los conjuntos observados se representan en rojo y los PSY sintéticos en azul.

Conclusiones

A partir de los resultados mostrados podemos concluir que en resolución anual los PSYs suscriben escenarios extremos manteniendo la variabilidad natural del recurso en todas las localizaciones. En el caso de Tamanrasset y Goodwin Creek, se ha observado un conjunto de datos de radiación solar extremadamente alto, pero esta anomalía no parece afectar significativamente a los resultados obtenidos. En la resolución mensual, es destacable cómo el algoritmo reproduce meses significativamente claros y nublados, manteniendo la relación kt-kb en localizaciones con alta y baja dispersión. En la resolución diaria, las localizaciones seleccionadas describen varios escenarios de radiación solar que el algoritmo es capaz de reproducir y ampliar, proporcionando un mayor rango de escenarios. Finalmente, las distribuciones sintéticas de frecuencia de kt y kb de 1 minuto reproducen la tendencia de las series observadas manteniendo la dispersión y amplitud de los perfiles de distribución en todas las localizaciones evaluadas

Una vez evaluado el rendimiento del método, podemos concluir que los PSYs sintéticos obtenidos con esta metodología reproducen el comportamiento de los conjuntos de datos observados manteniendo la relación entre GHI y DNI y permitiendo obtener escenarios extremos que aún no han sido observados. Los resultados muestran que este enfoque puede complementar el enfoque común del TMY aportando la oportunidad de realizar una evaluación estocástica de los sistemas de aprovechamiento solar.

Los trabajos futuros podrían centrarse en la generación de Años Meteorológicos Plausibles en lugar de Años Solares Plausibles, considerando otras variables meteorológicas relevantes como la velocidad del viento, la temperatura y la humedad relativa, además de los datos de radiación solar. Además, la inclusión de nuevos climas en la base de datos utilizada para la generación sintética de 1 minuto podría mejorar los resultados obtenidos.

El enfoque considera la hipótesis de que las tendencias futuras pueden caracterizarse por las tendencias pasadas, pero se aplicará el impacto del cambio climático en las tendencias a largo plazo, ya sea en la radiación solar o en las variables meteorológicas.