Synthetic Generation of Plausible Solar Years For Long-Term Forecasting of Solar Radiation

Paola Jiménez-Valero
Miguel Larrañeta (mlarraneta@gter.es)
AICIA https://orcid.org/0000-0002-9090-4200
Elisa López-García
Sara Moreno-Tejera
Manuel Antonio Silva-Pérez
Isidoro Lillo-Bravo

Research Article

Keywords: Synthetic sequences, Solar variability, Plausible Meteorological Years (PMYs), Multiyear

Posted Date: March 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1350312/v1

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 of coupled global horizontal irradiance (GHI) and direct normal solar irradiance (DNI) in 1 min resolution. The algorithm uses 10–20 years of hourly coupled GHI + DNI datasets that can be retrieved for most of the locations of the world from satellite estimates. The algorithm is evaluated at five locations with 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.

1 Introduction

As recognized by different authors and institutions (Abreu et al., 2018; Peruchena et al., 2018; Renné, 2016) the main source of uncertainty in large solar energy projects is the available solar resource. Ideally, the long-term assessment of the solar resource should be based on high quality time series of measured data recorded during a long period, at least 10 consecutive years. Because such time series were very seldom available, different methodologies have been developed to characterize the solar resource with the lowest possible uncertainty, being the elaboration of time series obtained through combination of solar radiation estimation from satellite images and local measurements for site adaptation (adjustment) the most successful one (Sengupta et al., 2018).

Currently, the conventional approach to energy yield assessment of solar energy projects relies on the use of plant models one of whose main inputs is the solar resource. The solar resource is usually characterized by a typical meteorological year or TMY, a time series of hourly or sub-hourly values of solar radiation and other meteorological variables elaborated from different sources and following different statistical procedures (Hall et al., 1978; Nielsen, et al. 2017), intended to represent the average solar resource expected during the project lifetime, but the climatic characteristics of a given location are not completely represented by average values. Indeed, the characterization of the solar resource requires a detailed study of its frequency distribution and variability at different time scales (Engeland et al., 2017).

The yield estimates based on TMYs can be complemented using other time series (PoE95, PoE90, etc.) (Nielsen et al.,2016) representing extreme, unfavorable meteorological conditions, i.e., years with an unusually low solar resource, for financial risk assessment. However, this approach has inherent limitations (Pavón et al., 2016; Ramírez et al., 2017). The elaboration of these complementary series requires strong assumptions regarding the statistical distribution of yearly values of solar radiation, since they must be elaborated from relatively short time series of observed data where the probability of occurrence of extreme years is low or very low. A PoE90 time series represents a year whose annual solar

radiation is lower than 90% of the years, but in most cases, it must be constructed from observed (from ground measurements or satellite images) data covering not more than 10 to 20 years.

To overcome the limitations of the conventional approaches, some authors (Ho et al., 2011; Nielsen et al., 2017) have suggested a stochastic approach to the energy yield assessment of solar energy projects based on the use of plausible meteorological years (PMYs) or plausible solar years (PSYs). A PMY is defined as a synthetic annual time series of hourly or sub-hourly values of meteorological variables elaborated following statistical procedures, intended to preserve the observed highfrequency variability of the solar resource in a specific location (Fernández-Peruchena et al., 2015). A PSY is a simplified PMY containing only solar radiation data (usually GHI and DNI). The synthetic generation of PSYs requires a long-term series of lowresolution ($\geq$ 15 min) observed data, usually obtained from 10 to 25 years of satellite images, and at least one or two years of high-resolution solar radiation data ($\leq$ 1 min). The statistical distribution of daily, monthly and yearly values of solar radiation is inferred from the long-term series, while the highfrequency variability is characterized by the observed variability of the highresolution series, assuming that the highfrequency variability in the future is adequately characterized by the observed variability. As a result, the annual synthetic series describe can describe a wide range of solar resource scenarios, including extreme ones that have never been observed, all of them compatible with the observed highfrequency meteorological variability of the location. PMYs generation allows the stochastic simulation of a solar harnessing system considering the uncertainty and variability inherent of the solar resource (Larrañeta et al., 2021)

In this paper, we present and evaluate an algorithm based on the method used by Larrañeta et al., (2019) 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 (Larrañeta et al., 2021). The methodology does not need any local adaptation or calibration, although better results are obtained when using good quality high-frequency data measured on site. We present and update of the method proposed by Larrañeta et al. (2019) and test it on 5 locations with different climates. To this purpose, we generate 100 annual sets of coupled 1-min DNI and GHI solar data for each location using SolarGIS satellite derived data. SolarGIS professional data sets provide 20 to 25 annual sets of 15-min of coupled DNI and GHI and other relevant meteorological variables. This tool allows to obtain 100 PSYs (representing the percentiles 1 to 100 of annual DNI or GHI values) that maintain the characteristics of the original highfrequency series but provides a wide range of solar radiation scenarios suitable for the probabilistic description of the performance evaluation of solar systems. This method allows a better characterization of the solar resource variability, facilitating the financial risk analysis by reducing the uncertainty of the energy yield estimations.

The paper is organized as follows: Section 2 presents a description of the method used for the synthetic generation of the PSYs. Section 3 presents the data used for the evaluation of the method. Section 4 presents the results, evaluation, and discussion in annual, monthly, daily, and 1-min resolutions. Finally, conclusions are drawn in Section 5.

2 Methodology

In the methodology section we describe the method used to generate multiple synthetic annual solar radiation time series at different locations. This method requires a minimum of 10 annual sets of coupled DNI and GHI in hourly-resolution 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 irradiance reanalysis databases. The size of the sample of the observations should be sufficient to statistically characterize the annual and monthly distribution of the GHI and DNI. For this reason, it's advisable a minimum period of 10 years, either consecutive or non-consecutive, of observed data sets. The methodology applied for the synthetic generation consists of three steps (Larrañeta et al., 2019).

In the first step, we obtain 100 annual series at monthly scale. We use the probability integral transform method over the observed integrated monthly values, to generate the monthly synthetic solar data.. We constrain 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. In this case, we assume that the annual GHI and DNI follows a Normal 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 Fig. 1 we show the results of the first step calculations for the location of Seville. On the left we present the annual values of DNI versus their corresponding values of GHI of the observed data in a scatter plot, the 10000 synthetic data and the Normal distribution values. On the right we present the Normal distribution values and the 100 selected synthetic values closest to them

In the second step, we downscale the synthetic solar irradiance sets from monthly to daily time resolution. Based on (Grantham et al., 2018), the daily model consists of three components: a seasonal, an autoregressive (AR) and a random component. To determine if the set has an ARMA structure, we first perform a standardization of each annual set in the daily resolution. To that end, we estimate the seasonal component from the Fourier model, this component just depends on the day of the year, and then we divide by the standard deviation of the annual set. We analyze the correlations of this set and we observe that it can be represented by a first-order autoregressive progress (AR(1))

We then construct the AR (1) model for the standardized series. For the same locations shown in Fig. 1 the AR(1) coefficients are 0.34 and 0.42 in average for GHI and respectively. The random component or white noise is the difference between the standardized series and the AR(1) model.

We then generate synthetic daily values from the AR(1) model. The synthetic generation procedure for the daily coupled GHI + DNI data follows four main steps:

1. We first calculate the random component. We apply a bootstrapping technique to the white noise series. This technique uses random numbers assumed to be probabilities and calculate the estimated value of that probability from the daily 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.

2. Then we calculate the autoregressive component. The AR component depends on the previous solar radiation value. We calculate this component by multiplying the previous value of a given day by the estimated AR(1) coefficient for a given location.

3. In the last step, we take into to account 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.

Finally, for each synthetic day, we search the observed day with the closest value in terms of energy (kt and kb) and select their corresponding values of variability and distribution. The observed datasets have been also labeled with two more daily indexes. The variability index (VI) that provides information about the intra-daily variability of the solar radiation and the morning fraction index (Fm), that provides information about the intra-daily distribution of the solar radiation (Moreno-Tejera et al., 2017). 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 that more closely match the ones obtained in step 1.

In the last step, we use the synthetic daily quartets of kb, kt, VI and Fm obtained in step 2 to generate the 1-min synthetic data sets. In this step, we use the ND model tool (Larrañeta et al., 2019) 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 (M. Larrañeta et al., 2018). The non-dimensional database daily profiles are labeled with the indexes kt, kb, VI and Fm that provide information about energy, variability and distribution following Moreno-Tejera., 2017.). 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 database. 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 information of the locations of the dimensionless databases implemented in the algorithm. We have selected seven locations with different climates. The dimensionless databases are composed by 1-min observed GHI and DNI data obtained from the Australian Bureau of Meteorology website (Bureau of Meteorology, 2015) for six different locations in Australia. The 1-min observed GHI and DNI datasets for the location of Seville, Spain, are obtained from GTER meteorological database (S. Moreno-Tejera et al., 2016)..

Table 1
Dimensionless databases available sites and their Köppen-Geiger classification climate.

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 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 representative solar year (RSY) for each location. RSY is a simplification of TMY that uses only solar radiation data for the calculation of the most representative long term annual set. RSYs have been calculated following AENOR methodology (UNE 206011:2014., 2014)

3.1 Input databases

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 2 we present the main geographical and climatological information of the selected sites.

Table 2
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

3.2 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 RSY annual cumulative value as a green dot.

In Fig. 2 we observe that annual cumulative values of the RSY are very close to the median in the observed and synthetic datasets for both DNI and GHI at all locations, except in the case of the GHI component of Tamanrasset. In general, the synthetic values cover a wider range of scenarios than the observed values, meaning that the PSYs cover extreme scenarios that have not been observed preserving the natural variability of the resource in each selected location.

Figure 3 shows the annual GHI values versus the annual DNI values of the observed, PSYs and RSYs sets in a scatter plot at all the selected locations.

In Fig. 3 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 RSYs are located in the middle of the point cloud for all the locations, representing a long-term representative annual set.

3.3 Monthly evaluation

In this section, we compare the monthly values of the synthetic, observed and RSY sets. We calculate the clearness index (kt) and the direct fraction index (kb) in a monthly basis in order to exclude seasonality in the evaluation of GHI vs DNI. In Fig. 4, we represent both datasets in a scatter plot for the observed (red), synthetic (blue) and RSY (green).

We observe that we can reproduce the relation between 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 kt and kb 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 kt is close to 0.4 and kb 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 kt and kb 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 kt and kb of 0.82 and 1.05 respectively that, despite seeming an irregularity, the algorithm is capable to reproduce.

The RSY cloud points are located, as expected, inside the range of the observed sets.

3.4 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 Fig. 5, we present the CDFs of the daily cumulative GHI (left) and DNI (right) for the five locations under study.

The shape of the CDFs denotes significant variations depending on the location. The PMY CDFs reproduce the different shapes of the observed CDFs for all the locations and for both GHI and DNI. Synthetic CDFs cover a greater range of scenarios than the observed sets, although, as an exception, in Tamanrasset and Goodwin Creek we can observe an anomalous year separated from the rest. These CDFs correspond to the extreme annual values found in Fig. 3. The RSY presents an average CDF with respect to the CDFs of the observed data sets, since it is located in the middle of the observed CDFs.

In Fig. 6, 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.

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 shows a particular dispersion in the monthly (Fig. 4) and annual (Fig. 3) resolution evaluation.

We also observe greater kt-kb 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 (Polo et al., 2016).

RSYs point cloud stays in the middle of the point cloud of the observed sets reproducing an average performance.

3.5 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 (kt) and the direct fraction index (kb) to compare the distributions of the observed and synthetic series in the selected locations.

In Fig.7 we show the probability density distribution of 1-minute kt distribution (left) and 1-minute kb distribution (right) of the observed and synthetic sets at all the selected location

In this figure, we can observe some degree of bimodality in the 1-minute kt distributions, this is due to the presence of cloudy (peak at low kt) and cloud-less (peak at high kt) conditions (Tovar et al., 1998). In this case, we can see that the synthetic 1-minute kt values reproduce the distribution and trend of the observed series and maintain the bimodal character of the curve at all locations. The peaks corresponding to high kt are centered in similar positions except in the case of Tamanrasset. In the case of Boulder, Tamanrasset and Goodwin Creek, the synthetic series present a greater probability for the peak at high kt than the observed data. Although, in the case of Tamanrasset and Boulder, the area associated to high kt values is very similar. In the case of Goodwin Creek the difference of area at high kt values is balanced with a higher frequency of partially cloudy days. In general, the kt distributions maintain dispersion and amplitude between observed and synthetic sets. In the case of 1-minute observed and synthetic kb distributions, the shape of the probability density profile is maintained for low kb values. However, for high kb values we find greater differences, especially remarkable for Tamanrasset location. Although, as in the case of kt distributions, these differences seem to be balanced, this is, they present similar areas for kb values higher than 0.7.

4 Conclusions

In this work, we test and improve an algorithm for the synthetic generation of plausible solar years. The algorithm uses an extended database (10–20 years) of coupled GHI + DNI at hourly resolution as input and provides 100 synthetic 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 describes potentially extreme scenarios maintaining the natural variability of the resource in all the locations. In the case of Tamanraset and Goodwin Creek, an extremely high solar

radiation dataset has been observed, but this anomaly seems not affect significantly to the obtained results. In the monthly resolution, it is remarkable how the algorithm reproduces significantly clear and significantly cloudy months, maintaining the kt-kb relation in locations with high and low dispersion. In the 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

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 will 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.

Declarations

Acknowledgments

The authors are grateful to Solargis and especially to Tomas Cebecauer for providing the satellite-derived databases

Statements and Declarations

Funding

This work has been carried out in the frame of the project “Exploitation of the synergies of solar thermal and photovoltaic technologies for the development of solar hybrid systems of electricity production (ASDELSOL - PY18-RE-0029)” funded by the Junta de Andalucía (Spain) in the frame of the Andalusian investigation development and innovation plan (PAIDI 2020).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose

Ethics approval

This publication has no implications for public health and does not include research involving humans and/or animals

Consent to participate

Not applicable

Consent for publication

Not applicable

Data Availability

The datasets generated during and/or analysed during the current study are not publicly available due to the terms of use signed with the satellite data provider but are available from the corresponding author on reasonable request.

Code Availability

The code generated during the current study is available from the corresponding author on reasonable request.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Paola Jiménez-Valero and Sara Moreno-Tejera. Algorithm improvement and test was performed by Paola Jimenez-Valero, Elisa López-García and Miguel Larrañeta. The first draft of the manuscript was written by Paola Jiménez-Valero and Miguel Larrañeta and reviewed by Sara Moreno-Tejera, Manuel Silva-Pérez and Isidoro Lillo-Bravo. All authors read and approved the final manuscript.

References

1. Abreu EFM, Canhoto P, Prior V, Melicio R (2018) Solar resource assessment through long-term statistical analysis and typical data generation with different time resolutions using GHI measurements. Renewable Energy 127:398–411. https://doi.org/10.1016/j.renene.2018.04.068
2. Bureau of Meteorology (2015) One minute Solar Data. *Bureau of Meteorology*. http://www.bom.gov.au/climate/data/oneminsolar/stations.shtml
3. CTN 206/SC 117-CENTRALES TERMOSOLARES UNE 206011:2014 (2014) *Centrales termosolares. Procedimiento de generación de Año Solar Representativo.*
4. Engeland K, Borga M, Creutin JD, François B, Ramos MH, Vidal JP (2017) Space-time variability of climate variables and intermittent renewable electricity production – A review. Renew Sustain Energy Rev 79(February):600–617. https://doi.org/10.1016/j.rser.2017.05.046

5. Fernández-Peruchena CM, Gastón M, Sánchez M, García-Barberena J, Blanco M, Bernardos A (2015) MUS: A multiscale stochastic model for generating plausible meteorological years designed for multiyear solar energy yield simulations. Sol Energy 120:244–256. https://doi.org/10.1016/j.solener.2015.07.037
6. Grantham AP, Pudney PJ, Boland JW (2018) Generating synthetic sequences of global horizontal irradiation. *Solar Energy*, *162*(November 2017), 500–509. https://doi.org/10.1016/j.solener.2018.01.044
7. Hall L, Prairie R, Anderson H, Boes E (1978) *Generation of a Typical Meteorological Year for 26 SOLMET stations.*
8. Larrañeta M, Cantón-Marín C, Silva-Pérez MA, Lillo-bravo I (2019) Use of the ND Tool: An Open Tool for the Synthetic Generation of 1-min Solar Data from Hourly Means with Geographic Flecibility. *SolarPaces Conference*
9. Larrañeta M, Fernandez-Peruchena C, Silva-Pérez MA, Lillo-Bravo I (2018) Methodology to synthetically downscale DNI time series from 1-h to 1-min temporal resolution with geographic flexibility. Sol Energy 162(January):573–584. https://doi.org/10.1016/j.solener.2018.01.064
10. Larrañeta M, Fernández-Peruchena C, Silva-Pérez MA (2021) "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), Chap. 5, pp. 5-1-5–34*
11. Moreno-Tejera S, Silva-Pérez MA, Lillo-Bravo I, Ramírez-Santigosa L (2016) Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions. Sol Energy 132:430–441. https://doi.org/10.1016/j.solener.2016.03.032
12. Moreno-Tejera S, Silva-Pérez MA, Ramírez-Santigosa L, Lillo-Bravo I (2017) Classification of days according to DNI profiles using clustering techniques. Sol Energy 146:319–333. https://doi.org/10.1016/j.solener.2017.02.031
13. Nielsen K, Blanc P, Vignola F, Ramírez L, Blanco M, Meyer R (2016) *Meteorological Data Sets for CSP/STE Performance Simulations – Discussion of current practices; SolarPACES Report*
14. Nielsen KP, Vignola F, Ramírez L, Blanc P, Meyer R, Blanco (2017) M. Excerpts from the report: "BeyondTMY - Meteorological data sets for CSP/STE performance simulations." *AIP Conf. Proc.*, *1850*, *1400*. https://doi.org/10.1063/1.4984525
15. Pavón M, Fernandez C, Silva M, Moreno S, Guisado M, Bernardos A (2016) *Statistical Analysis of CSP Plants by Simulating Extensive Meteorological Series. Proceedings of the SolarPACES Conference, Abu Dhabi, UAE*
16. Peruchena CF, Larrañeta M, Blanco M, Bernardos A (2018) High frequency generation of coupled GHI and DNI based on clustered Dynamic Paths. *Solar Energy*, *159*(August 2017), 453–457. https://doi.org/10.1016/j.solener.2017.11.024
17. Polo J, Wilbert S, Ruiz-Arias JA, Meyer R, Gueymard C, Súri M, Martín L, Mieslinger T, Blanc P, Grant I, Boland J, Ineichen P, Remund J, Escobar R, Troccoli A, Sengupta M, Nielsen KP, Renne D, Geuder N, Cebecauer T (2016) Preliminary survey on site-adaptation techniques for satellite-derived and

reanalysis solar radiation datasets. Sol Energy 132:25–37. https://doi.org/10.1016/j.solener.2016.03.001

18. Ramírez L, Pagh K, Vignola F, Blanco M, Blanc P, Meyer R, Wilbert S (2017) *Road Map for Creation of Advanced Meteorological Data Sets for CSP Performance Simulations. Proceedings of the SolarPACES Conference.* http://www.solarpaces.org/wp content/uploads/SolarPACES T5_BeyondTMY_Road_Map.pdf
19. Renné DS (2016) Resource assessment and site selection for solar heating and cooling systems. Advances in Solar Heating and Cooling. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100301-5.00002-3
20. Sengupta M, Xie Y, Lopez A, Habte A, Maclaurin G, Shelby J (2018) The National Solar Radiation Data Base (NSRDB). *Renewable and Sustainable Energy Reviews*, *89*(January 2018), 51–60. https://doi.org/10.1016/j.rser.2018.03.003
21. Tovar J, Olmo FJ, Alados-Arboledas L (1998) One-minute global irradiance probability density distributions conditioned to the optical air mass. Sol Energy 62(6):387–393. https://doi.org/10.1016/S0038-092X(98)00035-8

Figures

Figure 1

Result of the first step for location of Seville. In left figure, blue dots correspond to the synthetically generated 10000 values, red dots corresponds to the observed annual values and yellow dots correspond to the percentiles from 1 to 99 of the normal distribution fitted to the observed values. On the right, yellow dots represent the same percentiles of the normal distribution and blue dots represent the 100 synthetic GHI and DNI annual values closest to the percentiles of the normal distribution.

Figure 2

Boxplot of the observed and synthetic annual values for GHI (left) and DNI (right). The RSY is presented in a green dot.

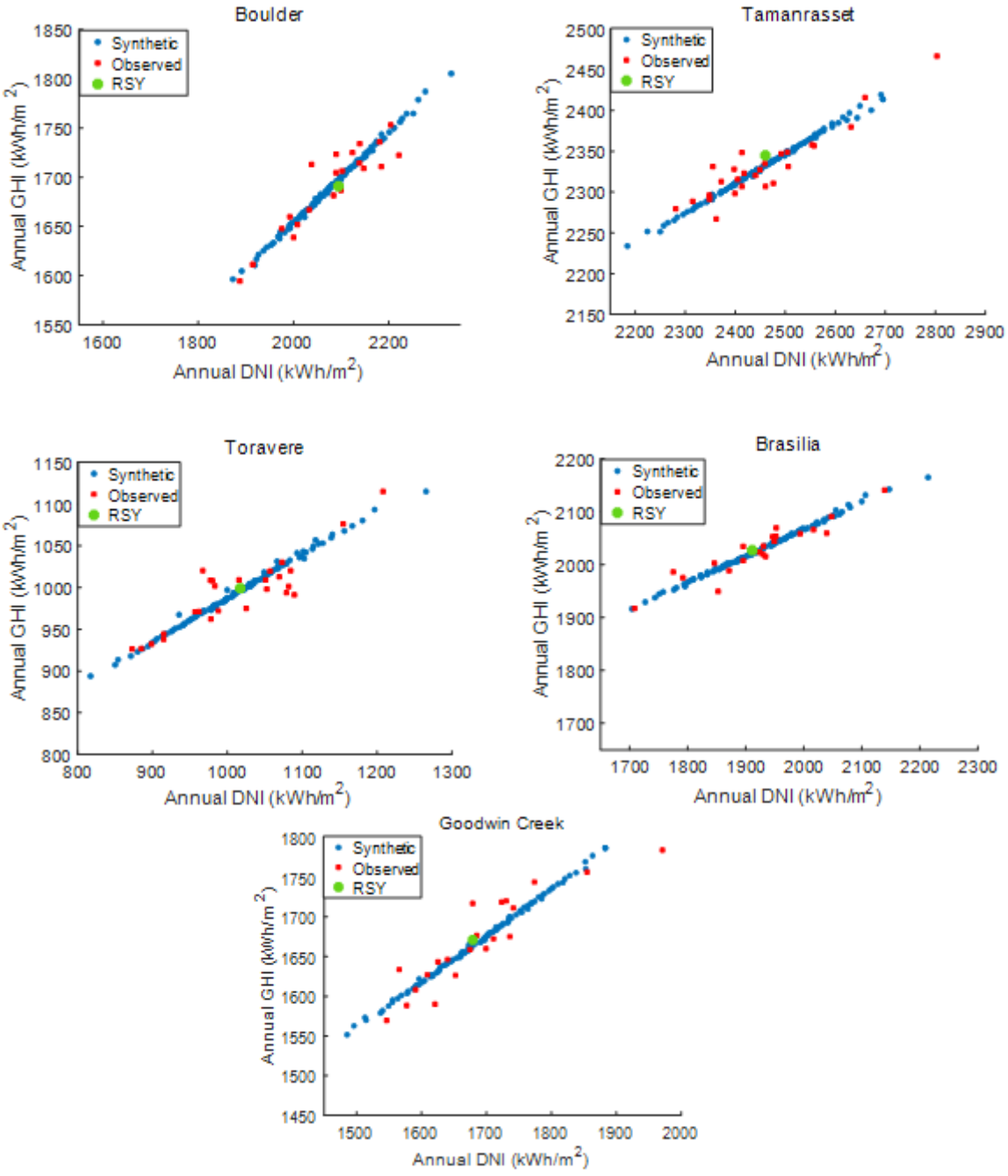

Figure 3

Scatter plot of the annual GHI versus the annual DNI. Observed sets are presented in red dots, synthetics PSYs in blue and RSY in green.

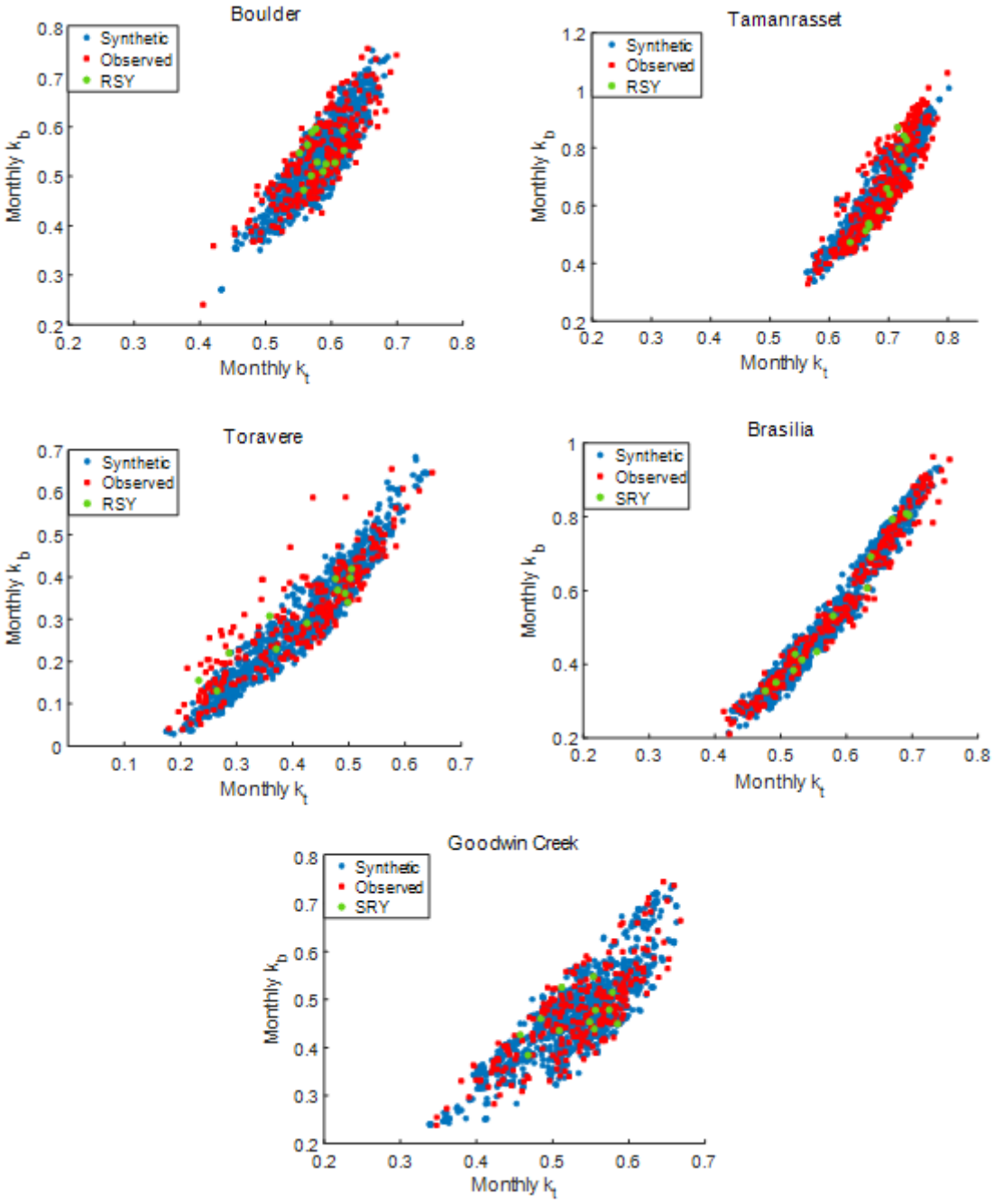

Figure 4

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

Figure 5

CDFs of the daily cumulative GHI (left) and DNI (right). Observed sets are presented in red, synthetic PSYs blue and RSY in green.

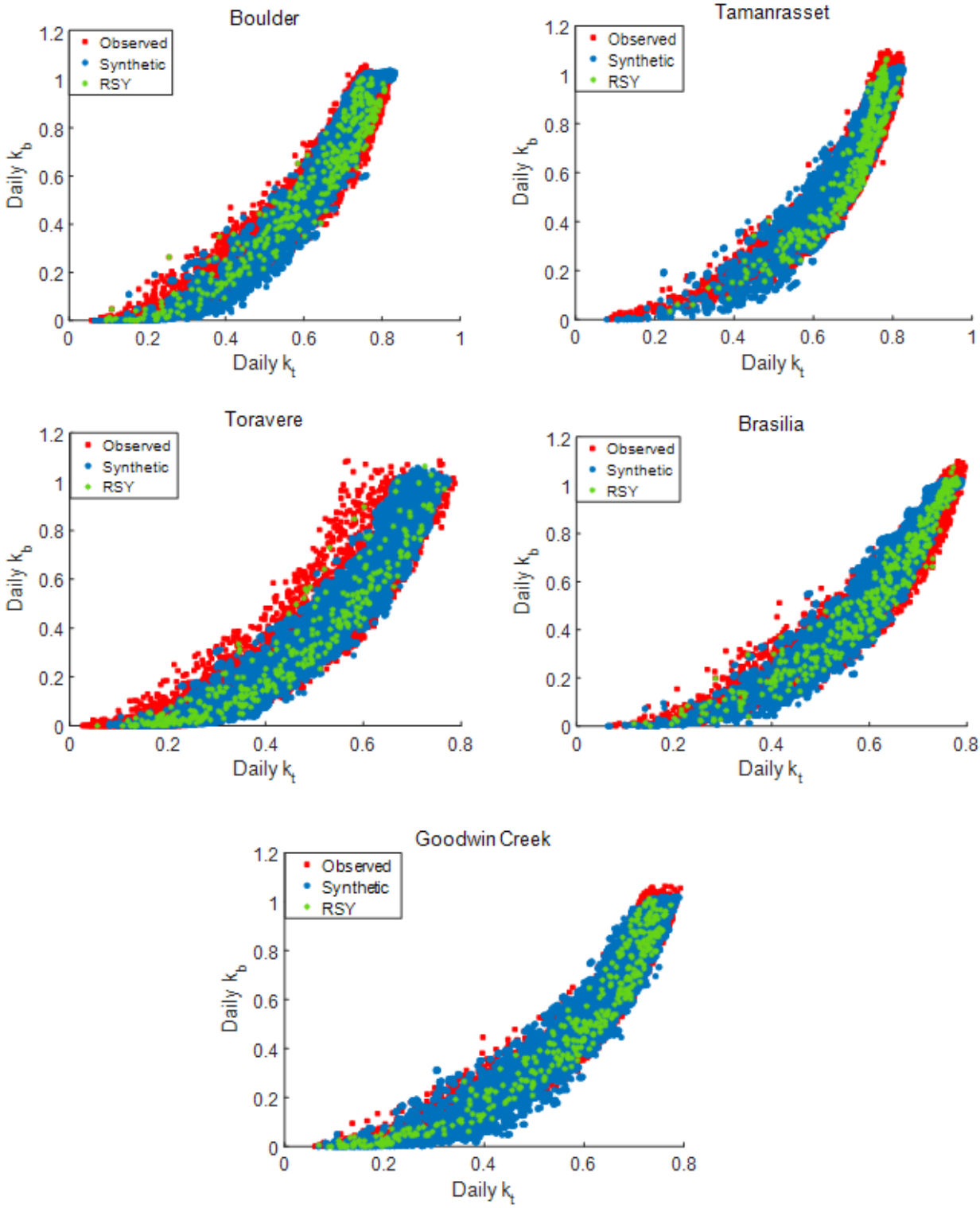

Figure 6

Scatter plot of the daily clearness index versus the direct fraction index. Observed sets are presented in red dots, synthetic PSYs in blue overlapped dots and RSY in green dots.

Figure 7

Probability density distribution of 1-minute kt and kb values. Observed sets are represented in red and synthetic PSYs in blue.