Renewable Energy 127 (2018) 398–411

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Solar resource assessment through long-term statistical analysis and typical data generation with different time resolutions using GHI measurements

Edgar F.M. Abreu [a,*], Paulo Canhoto [a,b], Victor Prior [c], R. Melicio [a,b,d]

[a] Instituto de Ciências da Terra, Universidade de Évora, R. Romão Ramalho 59, 7000-671 Évora, Portugal
[b] Departamento de Física, Escola de Ciências e Tecnologia, Universidade de Évora, R. Romão Ramalho 59, 7000-671 Évora, Portugal
[c] Observatório Meteorológico do Funchal, IPMA - Instituto Português do Mar e da Atmosfera, R. Lazareto 37-39, 9060-019 Funchal, Portugal
[d] IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

A R T I C L E I N F O

Article history:
Received 22 December 2017
Received in revised form
27 March 2018
Accepted 24 April 2018
Available online 25 April 2018

Keywords:
Solar resource assessment
Global horizontal irradiation
Typical meteorological year
Madeira island

A B S T R A C T

This work addresses the solar resource assessment through long-term statistical analysis and typical weather data generation with different time resolutions, using measurements of Global Horizontal Irradiation (GHI) and other relevant meteorological variables from eight ground-based weather stations covering the south and north coasts and the central mountains of Madeira Island, Portugal. Typical data are generated based on the selection and concatenation of hourly data considering three different time periods (month, five-day and typical days) through a modified Sandia method. This analysis was carried out by computing the Root Mean Square Difference (RMSD) and the Normalized RMSD (NRMSD) for each time slot of the typical years taking the long-term average as reference. It was found that the datasets generated with typical days present a lower value of overall NRMSD. A comparison between the hourly values of the generated typical data and the long-term averages was also carried out using various statistical indicators. To simplify this analysis, those statistical indicators were combined into a single Global Performance Index (GPI). It was found that datasets based on typical days have the highest value of GPI, followed by the datasets based on typical five-day periods and then those based on typical months.

1. Introduction

Solar resource information is crucial when designing or simulating solar energy systems. In the development of a solar energy project, various time scales of solar resource assessment are required, ranging from simple annual mean values in the first design approach to short-term characterization of solar radiation and of other relevant meteorological variables that affect the performance of the installation, such as air temperature and wind speed. Therefore, time resolution of the solar radiation data may have a significant impact on the viability analysis of a solar energy project [1]. The detailed design as well as its economic viability are directly related with both magnitude and variability of the local solar resource [2,3], thus making imperative its characterization in different time resolutions.

A common approach to quantify and evaluate the solar resource in a given region is the analysis of monthly and annual values of Global Horizontal Irradiation (GHI) [1,4]. Regarding the prediction of solar energy systems performance, it is recommended to use past measurements of solar radiation in the region of interest together with radiative transfer models [4,5], and thus radiometric measurements are essential to evaluate the solar resource (assuming proper calibration and maintenance of radiometers) [6]. When a sufficient number of ground-based stations is available in a given region, then the use of interpolation methods to evaluate the solar resource over that region is also appropriate [6]. Additionally, as solar energy power plants have a considerable cost, it is important to know in advance the energy that a given system or power plant will generate depending on its location. In that sense, modelling of solar energy systems for a long-term data series of radiometric measurements is a possible way to study the response of those systems under different meteorological conditions. However,

* Corresponding author.
E-mail address: eabreu@uevora.pt (E.F.M. Abreu).

https://doi.org/10.1016/j.renene.2018.04.068

Nomenclature

FS	Finkelstein-Schafer statistics
GPI	Global Performance Index
H	Hourly global horizontal irradiation (kWh/m^2-hour)
$\overline{H}$	Long-term average of hourly global horizontal irradiation (kWh/m^2-hour)
k	Ranked order number
MBE	Mean Bias Error (kWh/m^2-hour, $^\circ C$, % or m/s)
n	Total number of data records
N	Number of daily records in a month
NRMSD	Normalized root mean square difference
P	Number of hours in a day with $H > 0$
R	Correlation Coefficient
RMSD	Root mean square difference (kWh/m^2-hour, $^\circ C$, % or m/s)
S_n	Cumulative Distribution Function
SD	Standard Deviation (kWh/m^2-hour, $^\circ C$, % or m/s)
$t - stats$	t-statistics
U95	Uncertainty at 95% (kWh/m^2-hour, $^\circ C$, % or m/s)
WS	Weighted sum of the FS statistics
x	Meteorological parameter
$\tilde{y}_j$	Median of scaled values of the statistical indicator j
$\tilde{y}_{ij}$	Scaled value of the statistical indicator j and typical data set i

Greek symbols

δ	Absolute difference between CDFs
ω	Statistical weight

Acronyms

CDF	Cumulative Distribution Function
FS	Finkelstein-Schafer statistics
GHI	Global Horizontal Irradiation
TMM	Typical Meteorological Month
TMY	Typical Meteorological Year
TRY	Test Reference Year
TSRY	Typical Solar Radiation Year

performing such numerical simulations is a time consuming computational process. This problem can be overcome if typical weather data sets are used, such as Typical Meteorological Years (TMY), which also reproduce all the long-term data statistics. The Typical Meteorological Years are also useful for energy simulations in buildings, either for determining their thermal response under different environmental conditions or designing and simulating integrated renewable energy systems. Usually, dedicated software requires representative meteorological data files of hourly values for the specific location to assess the performance of buildings and solar energy systems. However, if hourly weather data series are not available then they can be generated through statistical methods from monthly averaged values of the required meteorological parameters measured in nearby locations [7].

The TMY and the Test Reference Year (TRY) datasets are widely used to study the performance of buildings and solar energy systems [8,9]. The TRY is generated through the concatenation of 12 actual months selected from the entire time series of meteorological measurements according to the standard EN ISO 15927-4 [10]. Although this standard describes a procedure for generating a reference year suitable for evaluating the annual heating and cooling requirements in buildings based on long-term data, it was found that the same reference year can also be used to predict the output of solar photovoltaic energy systems [11]. The resulting data series consists of 8760 hourly values of various selected meteorological parameters such as the air temperature, solar radiation, relative humidity, and wind speed [12]. Several authors [12–14] reported TRY data for different locations around the world. In a similar way, the TMY comprises hourly values of solar radiation and other meteorological parameters for a one-year period that is generated from long-term observed data [15]. The first version of the TMY was generated using the Sandia method [16] as proposed by Hall in 1978. The National Renewable Energy Laboratory (NREL) developed a second and a third versions of the TMY denominated TMY2 (1995) [17] and TMY3 (2008) [18], respectively, adding some changes in the selection procedure of the typical data. However, since the TMY represents typical rather than extreme conditions, it is not suitable for designing systems to meet the worst-case conditions occurring at a given location [16].

Several other methods have been proposed to generate typical weather data: the Danish method [19], the Festa-Ratto method [20], the Crow method [21], the Miguel-Bilbao method [22], the Gazela-Mathioulakis method [23] and the Stochastic Approach [7,24]. Some researchers, including Skeiker [25], Janjai and Deeyai [26] and Ebrahimpour and Maerefat [27], have compared various methods and concluded that the Sandia method is the one that generates the closest evaluation to long-term performance of thermal systems in buildings [8]. Kulesza [28] recently found that the TRY perform better than the TMY for central Poland. Additionally, several TMY data sets have been generated worldwide using the Sandia method [29–35] with the help of Finkelstein-Schafer (FS) statistics [36] in order to evaluate how close a given month is to the long-term data series of the corresponding calendar month, regarding the cumulative distribution of daily values and the monthly average. If only solar radiation data is considered, Typical Solar Radiation Years (TSRY) have been generated by some researchers [37–39]. The TSRY generation is carried out using daily global solar radiation data and the FS statistical method thus making it useful for the first approach in the design of solar energy conversion systems [38].

In this work, solar resource assessment is addressed through long-term statistical analysis and typical data generation with different time resolutions, using measurements of Global Horizontal Irradiation (GHI), air temperature, relative humidity and wind speed from eight ground-based weather stations in Madeira Island, Portugal. The Madeira Island is located in the North Atlantic Ocean, southwest of Portugal mainland, between parallels 32$^\circ N$ and 33$^\circ N$ of latitude, and has an area of 740.7 km^2. The climate of Madeira Island is classified as temperate with dry and warm summers according to the modified Köppen-Geiger classification system [40], except in some narrow strips along the coastal zones where a temperate climate with hot and dry summers is observed. The central mountains, with a maximum altitude of 1862 m above the mean sea level, together with the proximity to the ocean, strongly affect the local meteorological phenomena, including those closely related with the renewable energy resources, namely the wind speed, precipitation and solar radiation. Since there is no fossil fuel extraction and refining in the area, the generation of electric energy in the island is done mostly via imported fuel and thus the environmental resources such as solar energy should be considered for power generation to reduce the dependence from conventional sources of primary energy. In recent years the installed capacity of wind power increased and a growing investment in solar energy is expected soon, namely photovoltaic (PV) power, as a way of diversifying the sources of renewable energy while maintaining a background power generation capacity from conventional sources due to grid stability issues. Thus, the generation of Typical Meteorological Years for the eight locations analysed in this work also allows the characterization of the different climatic patterns in the island and provide datasets that can be used for studying energy efficiency in buildings and for dimensioning solar energy systems, thus contributing to a correct energy planning and, ultimately, for the sustainable development of the island. This also helps to identify the locations with higher levels of annual GHI and provides datasets of hourly values for the simulation of solar energy systems such as photovoltaic systems, including the effect of air temperature and wind speed, which are meteorological parameters that affect the energy generation of such systems via the temperature of the PV cells [41].

The aim of this work is to study the impact of the time resolution used to generate typical data on the capacity of these datasets to represent the long-term statistics (long-term mean values) of GHI, air temperature, relative humidity and wind speed, thus contributing to improve solar resource assessment methodologies using typical weather data. With this purpose, a detailed analysis of three data sets generated through the selection and concatenation of different typical time periods (month, five-day and typical days) is presented and a comparison with the long-term mean values is carried out for each one. A modified Sandia method was implemented in order to select the different typical data periods. To the best knowledge of the authors, this is the first study that addresses the impact of time resolution on the generation of typical weather data sets through methods similar to the Sandia method. Since there are no Direct Normal Irradiation measurements in the Madeira Island, the original Sandia method is used as the starting base to generate the typical data sets, instead of the more recent TMY3 method. The paper is organized as follows: Section 2 presents the location of the measuring stations, measuring period and the quality control of solar radiation and weather data; Section 3 presents the solar resource assessment using a long-term statistical analysis and the generation of typical data with different time resolutions; Section 4 presents the results and discussion; and, finally, conclusions are drawn in Section 5.

2. Solar radiation and weather data

2.1. Location of the weather stations and measuring period

The global horizontal irradiation measurements and weather data used in this work are from the network of automatic ground-based stations of the national weather service, Instituto Português do Mar e da Atmosfera (IPMA), and the following locations in the island were considered: Areeiro, Caniçal, Lido, Observatório, Lombo da Terça, Lugar de Baixo, Ponta do Pargo and São Jorge [42]. The IPMA network comprises CM11 pyranometers from Kipp & Zonen with periodic and standard maintenance procedures following the World Meteorological Organization (WMO) guidelines. Depending on the location, at least five years in the period 2003–2014 of hourly averaged and extreme values were used. The following meteorological parameters were considered: hourly mean, maximum and minimum values of air temperature and relative humidity, hourly mean and daily maximum wind speed and hourly GHI. The location of the stations and the respective period of measurements are presented in Table 1. These sites were selected to cover the entire area of the island, thus including weather stations in the south coast, north coast and central mountains. The results for Observatório, São Jorge and Areeiro (representing the south coast, north coast and central mountains, respectively) are presented in detail in the next sections as an example of the work developed in this study, although the same outputs were obtained for the remaining locations.

Table 1
Location and period of measurement of selected stations in Madeira Island.

Station	Latitude [N]	Longitude [W]	Altitude (m)	Period
Areeiro	32° 43′ 20″	16° 54′ 55″	1610	2003–2014
Caniçal	32° 44′ 54″	16° 42′ 24″	68	2010–2014
Lido	32° 38′ 12″	16° 56′ 08″	13	2003–2014
Observatório	32° 38′ 51″	16° 53′ 33″	58	2004–2014
Lombo da Terça	32° 49′ 52″	17° 12′ 08″	660	2010–2014
Lugar de Baixo	32° 40′ 52″	17° 05′ 24″	15	2003–2014
Ponta do Pargo	32° 48′ 48″	17° 15′ 42″	312	2008–2014
São Jorge	32° 50′ 04″	16° 54′ 42″	82	2003–2014

2.2. Quality control and gap filling

Solar radiation and weather data series often contain incorrect time records, gaps or incorrect measurements. Therefore, it is necessary to do a quality control to identify these problems. Firstly, the GHI filters used in the Cabauw station were applied to the GHI data, according to the Baseline Surface Radiation Network guidelines [43]. These filters identify physically impossible values, extremely rare values and values that are not consistent regarding the ratio between diffuse and global solar irradiance values. Diffuse irradiation is available only in the Observatório station, thus it was not studied in detail here. Direct Normal Irradiation is not measured in the island. This is the main reason that the more recent TMY3 method [18] is not used here. Regarding gap filling, the missing records are less than 15% of the total number of records in all the cases, except for the hourly values of GHI from the Observatório station (23%). Gaps with one or two consecutive hours were filled using linear interpolation. Gaps with more than two consecutive hours were filled with an acceptable accuracy through correlations with other stations, with correlation coefficients higher than 0.8 in all the cases.

3. Solar resource assessment

3.1. Long-term statistical analysis

A common approach to quantify and evaluate the solar resource is the analysis of monthly and annual values of GHI [1,4]. This approach is a simple procedure to assess the solar energy availability at a given location and to perform comparisons between different locations. The monthly averages of the daily GHI and the respective annual mean and standard deviation are presented in Table 2 for all the stations included in this work.

The location with higher annual mean of daily GHI is Areeiro with 4.84 kWh/m^2-day. This is also the location where a higher variation of the daily solar radiation occurs along the year. The Lugar de Baixo and Ponta do Pargo stations also have high values of daily GHI, 4.73 kWh/m^2-day and 4.77 kWh/m^2-day, respectively, but present a lower variability if compared with Areeiro. The location with lower annual mean of daily GHI is Lombo da Terça followed by São Jorge, with 4.03 kWh/m^2-day and 4.06 kWh/m^2-day, respectively, both in the north coast of the island. The minimum value of monthly mean of daily GHI is 1.75 kWh/m^2-day and occurs in December at Lombo da Terça, while the maximum value is 7.71 kWh/m^2-day in July at Areeiro. To better understand the GHI variation along the year, the long-term average of the hourly GHI is shown in Fig. 1 for the Observatório, São Jorge and Areeiro stations.

Table 2
Monthly and annual averages of daily GHI (kWh/m^2-day).

Month	Station							
	Areeiro	Caniçal	Lido	Obs.	L. Terça	L. Baixo	P. Pargo	S. Jorge
Jan	2.84	2.57	2.69	2.81	2.17	2.86	2.65	2.22
Feb	3.27	3.17	3.34	3.44	2.48	3.52	3.27	2.83
Mar	4.64	4.64	4.48	4.68	3.79	4.74	4.48	4.02
Apr	5.38	5.58	5.09	5.38	4.66	5.66	5.80	5.07
May	6.49	5.88	5.88	5.68	5.13	6.17	6.29	5.56
Jun	7.25	6.31	5.95	5.81	6.35	6.38	7.04	5.70
Jul	7.71	6.07	6.09	5.91	6.28	6.43	7.05	5.63
Aug	7.06	6.12	5.81	5.83	6.55	6.25	6.75	5.39
Sep	4.87	5.00	4.87	5.05	4.25	5.12	5.01	4.56
Oct	3.75	3.91	3.88	3.98	3.13	4.06	3.85	3.52
Nov	2.49	2.67	2.86	3.06	1.81	3.00	2.77	2.23
Dec	2.32	2.28	2.45	2.47	1.75	2.53	2.36	2.02
Annual	4.84	4.52	4.45	4.51	4.03	4.73	4.77	4.06
std.dev.	1.94	1.53	1.37	1.29	1.79	1.48	1.79	1.45

3.2. *Typical data generation with different time resolutions*

According to the Sandia method [16], the TMY generation is carried out through the concatenation of twelve Typical Meteorological Months (TMMs). The selection of the TMM for each month of the calendar starts by evaluating the Finkelstein-Schafer statistics (FS) of the daily values of the nine meteorological parameters considered in the method (Appendix A). The method is based on the Cumulative Distribution Function (CDF) for each one of the meteorological variables, as shown in Fig. 2a for the GHI in June at the Observatório station, and using different statistical weights for the variables as presented in Table 3 [16]. Also, according to this method, GHI is the most relevant variable with a statistical weight of 50%. The TMMs are then selected based on its closeness to the long-term CDF, considering all the meteorological parameters.

A similar procedure was used to select the typical five-day periods and the typical days, as shown in Fig. 2 b and c, respectively. The TMYs based on five-day and typical days are generated using the same raw data except that hourly values are used to determine the long and short-term CDFs instead of using the daily values as before. In these cases, only the hourly mean wind speed is used, since there are no hourly maximum values of this parameter. The statistical weights are presented in Table 3, which maintain the relative weight between the four meteorological parameters. The five-day TMY is composed of 73 periods that are selected by

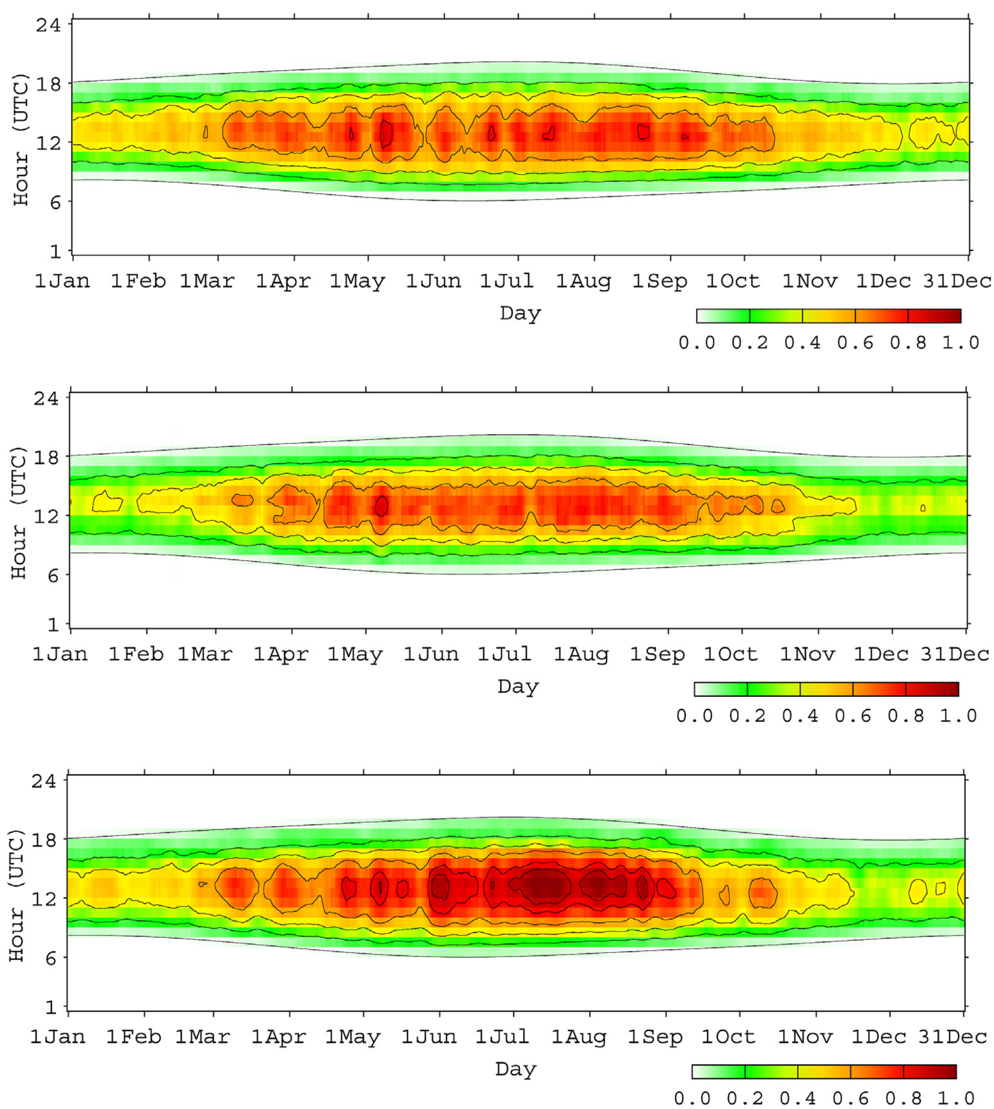

Fig. 1. Long-term average of hourly GHI (kWh/m^2-hour) at Observatório (top), São Jorge (center) and Areeiro (bottom).

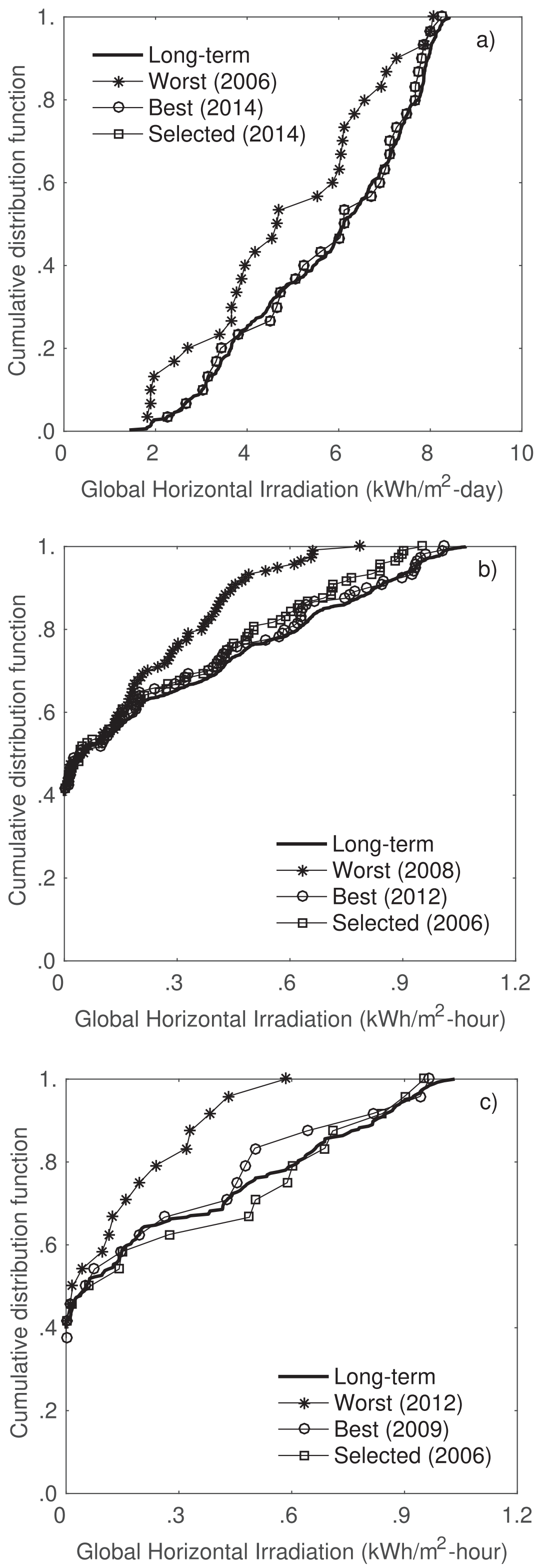

Fig. 2. Cumulative Distribution Functions of GHI values at Observatório: a) daily values of June; b) hourly values for the 31st five-day period; c) hourly values for June 1st.

comparing the sets of 5 days of each year of real measurements against the long-term CDF and average for the same time periods, while the daily TMY is composed of 365 typical meteorological days, i.e., each day of the TMY is selected by comparing all corresponding calendar days of the data series against the respective long-term CDF and average.

Table 3
Statistical weights of the meteorological parameters for the TMY generation [16].

TMY	Air temperature			Relative humidity			Wind speed		GHI
	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.	
Month	2/24	1/24	1/24	2/24	1/24	1/24	2/24	2/24	12/24
Five-day	2/24	1/24	1/24	2/24	1/24	1/24	4/24	–	12/24
Day	2/24	1/24	1/24	2/24	1/24	1/24	4/24	–	12/24

In Fig. 2, the long-term CDF, the CDF of the year with lowest *FS* value, the CDF of the year with highest *FS* value and the CDF of selected year are shown for the Observatório station. In Fig. 2a, the year with the lowest *FS* value, which corresponds to the year with the closest CDF to the long-term cumulative distribution, and the selected year for the TMY are the same. Since hourly measurements were used to generate the TMY based on five-day and typical days, the cumulative probability of GHI being zero in Fig. 2 b and c is around 40% for the respective time periods in June. This percentage represents the measurements taken during the night (GHI = 0 kWh/m^2-hour), which do not appear in Fig. 2a because daily values were used for the TMY generation based on typical months.

The smooth transition between the selected months, five-day periods and typical days was ensured by an interpolation in the period between 6 h before and after the mid-night where the selected typical data concatenate by using a centred moving average of three hourly values. It is worth noting that the transition between two consecutive typical days is not so abrupt because both are selected based on its closeness to the long-term CDF of hourly values, while the sequence of typical days retains the long-term climatological pattern. The difference between consecutive typical days will further decrease if the number of years in the data series increases. On the contrary, the transition between two consecutive typical months may be more abrupt because the selection procedure is based on the respective monthly CDFs of daily values and the last day of a typical month may be very different of the first day of the next typical month in terms of hourly values, although being only twelve transitions. This problem does not arise in the case of solar radiation since the transition between typical periods always occurs at night, and the interpolation that was implemented according to the Sandia method proved to be suitable to generate smooth typical data sets. On the other hand, the autocorrelation between typical periods may increase as the time slot for generating the CDF decreases, which means that the temporal correlation between two consecutive days in the daily TMY can be higher than that in the case of five-day TMY and monthly TMY. In any case, for a long data series, this autocorrelation is bounded by the persistence of weather conditions on consecutive days. That is, as typical days are selected based on the persistence of the meteorological conditions on each calendar day (closeness to the long-term CDF), if two or more consecutive typical days are highly correlated this means that similar conditions persist during those days in a long-term perspective, thus capturing the climatological pattern of that period. However, because the available number of years in the data series are limited and the changes in the weather conditions does not occur always in the same calendar day of every year, the resulting climatological pattern of the daily TMY can be slightly smoothed in the perspective of the daily mean values, thus contributing to increase the temporal correlation between typical days in the periods of persistent changes on the meteorological conditions, despite the hourly variability is that of selected days (measured hourly values). This issue needs to be further addressed in the future in order to establish a more precise relation between the time slot used to construct the CDF and the autocorrelation between the selected typical periods. The hourly GHI of the TMY based on typical months is shown in Fig. 3 for the Observatório, São Jorge and Areeiro stations.

A simple way of representing the selected years for each TMY is using the ratio between the number of times that a specific year was selected and the total number of periods, i.e., the percentage of occurrence of a specific year in the generated dataset. This frequency is shown in Tables 4–6, for the TMYs based on typical months, five-day periods and typical days, respectively, with the higher values highlighted in bold. The total period of measurements is very important because it has a strong impact on the determination of the long-term values, which are directly connected with the selection of the typical data. However, even for locations with the same period of measurements and not far from each other, as for example Lugar de Baixo and Lido, the selected years are different. This shows that close locations can have somewhat different climatological patterns due to the particular orographic and meteorological characteristics of the island. These differences are more evident between south coast, north coast and central mountains. A more detailed analysis is needed to better understand these aspects if required. Considering the TMYs based on typical five-day and day periods, Tables 5 and 6 respectively, and for a given location, the years with higher percentage of selection in the case of the TMY based on five-day period do not always correspond to the years with higher percentage of selection of the TMY based on typical days. This means that the time resolution used in the TMY generation also has a direct impact on the selected data, thus resulting in different final datasets.

4. Results analysis and discussion

The accuracy of the generated TMY in representing the long-term statistics of solar resource can be assessed through a comparison between the long-term average of daily GHI in a given time slot and the respective average of daily GHI data from the generated TMYs. This comparison is shown in Fig. 4 for the Observatório, São Jorge and Areeiro stations, using the monthly average of daily GHI. A very good agreement was found between the monthly TMY and the long-term monthly mean values, while the TMY based on typical days performs reasonably and the TMY based on five-day period does not perform so well in this monthly comparison.

In some way, this result is expected because we are comparing

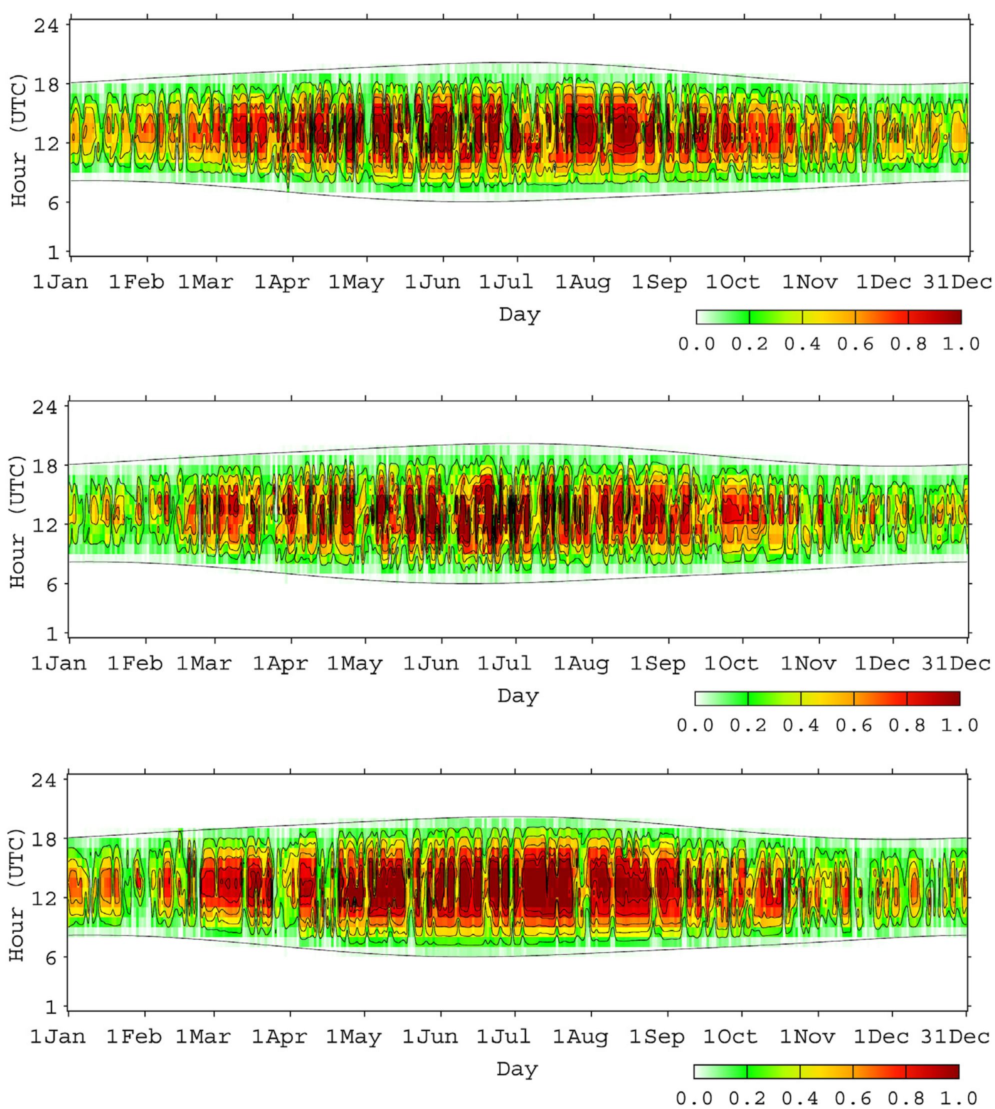

Fig. 3. Hourly GHI of the TMY based on typical months (kWh/m^2-hour) for Observatório (top), São Jorge (center) and Areeiro (bottom) stations.

Table 4
Percentage of selection of years in the case of the TMY based on typical months.

Station	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Areeiro	8.3	8.3	**16.7**	0	8.3	0	**16.7**	8.3	0	**16.7**	**16.7**	0
Caniçal	–	–	–	–	–	–	–	**25.0**	16.6	16.7	16.7	**25.0**
Lido	8.3	0	**16.7**	8.3	0	8.3	**16.7**	**16.7**	0	**16.7**	8.3	0
Observat.	–	8.3	0	16.7	8.3	16.7	**25.0**	0	8.3	0	0	16.7
L. Terça	–	–	–	–	–	–	–	**33.3**	0	8.4	25.0	**33.3**
L. Baixo	0	0	8.4	0	8.3	16.7	16.7	0	8.3	8.3	8.3	**25.0**
P. Pargo	–	–	–	–	–	0	8.3	16.7	**25.0**	16.7	**25.0**	8.3
S. Jorge	**16.7**	8.3	8.3	**16.7**	**16.7**	0	8.3	**16.7**	0	0	0	8.3

Table 5
Percentage of selection of years in the case of the TMY based on typical five-day periods.

Station	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Areeiro	15.1	2.8	8.2	6.8	9.6	4.1	8.2	4.1	12.3	5.5	5.5	**17.8**
Caniçal	–	–	–	–	–	–	–	17.8	**24.7**	17.8	16.4	23.3
Lido	11.0	8.2	**12.3**	5.5	5.5	**12.3**	8.2	9.6	8.2	5.5	5.5	8.2
Observat.	–	5.5	8.2	**13.7**	5.5	9.6	**13.7**	9.6	12.3	9.6	4.1	8.2
L. Terça	–	–	–	–	–	–	–	23.3	27.4	5.5	**28.8**	15.0
L. Baixo	12.3	6.8	**16.4**	6.8	4.2	2.7	6.8	6.8	11.0	12.3	9.7	4.2
P. Pargo	–	–	–	–	–	4.1	15.1	9.6	**19.2**	16.4	**19.2**	16.4
S. Jorge	**11.0**	5.6	8.2	6.8	6.8	8.2	**11.0**	6.8	9.6	8.2	9.6	8.2

Table 6
Percentage of selection of years in the case of the TMY based on typical days.

Station	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Areeiro	**12.9**	9.0	6.8	4.5	6.3	5.2	8.2	9.0	10.7	6.3	9.9	11.2
Caniçal	–	–	–	–	–	–	–	15.9	16.2	21.9	20.5	**25.5**
Lido	11.2	7.9	**12.6**	6.6	4.5	9.3	7.4	8.8	11.8	7.9	8.2	3.8
Observat.	–	**15.6**	11.0	10.1	13.2	7.4	5.8	9.6	6.6	6.2	6.6	7.9
L. Terça	–	–	–	–	–	–	–	**23.3**	15.6	21.1	23.0	17.0
L. Baixo	**17.8**	9.6	10.7	3.7	4.4	3.6	2.5	8.2	11.0	6.3	10.4	11.8
P. Pargo	–	–	–	–	–	4.7	7.9	17.3	**19.7**	13.4	18.6	18.4
S. Jorge	6.3	**12.3**	9.9	6.0	7.7	5.3	7.4	7.4	10.1	9.3	6.8	11.5

monthly averaged values and thus the TMY generated based on the monthly CDF will perform better. However, a question remains, that is, how each TMY performs for each variable (with distinct statistical weights) if different averaging time slots are used? To answer this question, the accuracy of the generated TMYs in representing the long-term statistics was also assessed through the Root Mean Square Difference (RMSD) and the Normalized Root Mean Square Difference (NRMSD) using the long-term mean values as reference, and for monthly, five-day and daily averaging periods, as shown in Table 7 for the Observatório station. In these calculations, the daily mean values were first determined for each variable, time slot and TMY, and then the RMSD values were determined using as reference the respective averaged values in the same time slots of the entire data series. Thus, the RMSD values are different for each averaging period because the daily mean values are calculated for each time slot prior to calculating the root mean square difference. This is also the reason why the values of RMSD increase as the averaging period decreases for a given TMY, that is, a larger averaging period reduces the scattering of TMY daily mean values with respect to the long-term daily mean. The NRMSD is defined as the ratio between the RMSD and the amplitude (difference between the maximum and minimum values) of each meteorological variable and station: $NRMSD = RMSD/(x_{max} - x_{min})$. In order to find which TMY presents the lower RMSD value considering all the meteorological parameters, a simple arithmetic average and a weighted average using the same statistical weights presented in Table 3 of the NRMSD values were also calculated. In this case, a total statistical weight of 4/24 was used for air temperature, relative humidity and wind speed.

In Table 7, the RMSD values of both TMY based on five-day and typical days are lower than the RMSD values for the TMY based on typical months for all the meteorological parameters and comparisons, except for wind speed and GHI in the monthly comparison. This can be possibly explained by the higher short-term variability of wind speed and GHI (due to the effect of scattered clouds and to the transition between clear and overcast skies), which benefits the comparison between mean values for longer averaging periods. It is worth noting that, if statistical weights other than those presented in Table 3 are used, then this analysis may change because different relative weights will be given to the meteorological variables, thus leading to the selection of other periods of typical data for the TMY generation. Nevertheless, regarding the data presented in Table 7 and the comparison done in Fig. 4 for the TMY generated from typical months, the RMSD values show that the TMYs generated from typical five-day and days are accurate, since similar results were obtained for all the stations and the weighted NRMSD average for both TMYs are lower than the values for the TMY based on typical months.

To identify the time resolution or typical period that better contributes to the representation of long-term statistics through the generation of Typical Meteorological Years, and considering all the meteorological parameters included in this analysis, an arithmetic average and a weighted average of the NRMSD of all meteorological parameters and locations were also calculated for the three generated TMYs. The results are presented in Table 8. The values in bold represent the TMY with the lower value of overall

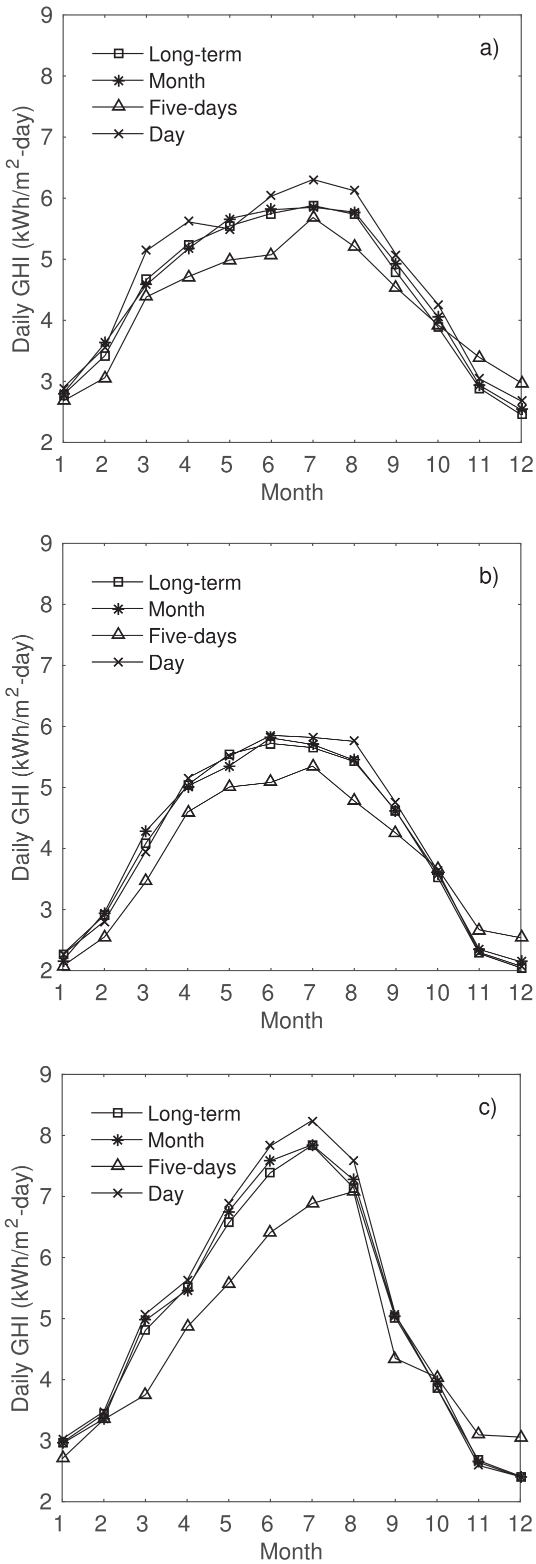

Fig. 4. Comparison between the monthly average of daily GHI of the TMY based on typical months, five-days and days, and the long-term mean: a) Observatório, b) São Jorge, c) Areeiro.

NRMSD for each averaging period and the underlined values correspond to the lower values for TMY based on typical months.

The values of Table 8 show that the conclusions drawn using the simple arithmetic average or the weighted average of all NRMSD values are not exactly the same. Taking the underlined values in Table 8, it is shown that the TMY based on typical months only presents lower NRMSD average value when a monthly comparison is carried out. On the contrary, the TMY based on typical days presents lower values in all the other cases, including for the weighted average of NRMSD. This means that, if a simulation of a solar energy application is to be performed on a daily or hourly basis including all the relevant meteorological variables, and the closest values to the long-term statistics are required, then the TMY based on typical days may be a more appropriate choice. However, if only GHI is considered and solar resource assessment is needed only in a monthly or annual basis, then the TMY generated with typical months represents better the long-term data (see Table 7).

This analysis was extended by comparing the generated TMYs against the long-term average year with the help of various statistical indicators, namely the Mean Bias Error (MBE), the Root Mean Square Difference (RMSD), the Uncertainty at 95% (U95), the t-statistics (t-stats) and the Correlation Coefficient (R), and without considering different averaging time slots, that is, using the hourly values directly. Lower values of these statistical indicators correspond to better performances, except in the case of the Mean Bias Error, in which values closer to zero correspond to a good performance, and for the case of the Correlation Coefficient, in which values closer to one indicate perfect linear relationship between two variables. These statistical indicators are defined as follows [44,45]:

Mean Bias Error:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\left(\bar{x}_{i,e} - \bar{x}_{i,m}\right) \tag{1}$$

Root Mean Square Difference:

$$RMSD = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\bar{x}_{i,e} - \bar{x}_{i,m}\right)^2\right]^{1/2} \tag{2}$$

Uncertainty at 95%:

$$U_{95} = 1.96\left(SD^2 + RMSD^2\right)^{1/2} \tag{3}$$

T-statistics:

$$t - stats = \left[\frac{(n-1)MBE^2}{RMSD^2 - MBE^2}\right]^{1/2} \tag{4}$$

Correlation coefficient:

$$R = \frac{\sum_{i=1}^{n}\left(\bar{x}_{i,e} - \bar{x}_{e,av}\right)\left(\bar{x}_{i,m} - \bar{x}_{m,av}\right)}{\left[\sum_{i=1}^{n}\left(\bar{x}_{i,e} - \bar{x}_{e,av}\right)^2\left(\bar{x}_{i,m} - \bar{x}_{m,av}\right)^2\right]^{1/2}} \tag{5}$$

in which, for each meteorological parameter x, $\bar{x}_{i,e}$ is the i^{th} hourly TMY value, $\bar{x}_{i,m}$ is the i^{th} hourly long-term mean, $\bar{x}_{e,av}$ is the average of the TMY hourly values, $\bar{x}_{m,av}$ is the mean value of the long-term averaged hourly values and SD is the standard deviation of $(\bar{x}_{i,e} - \bar{x}_{i,m})$.

The values of these statistical indicators are shown in Table 9 for the Observatório station, although this comparison was also performed for all the other stations. Based on the analysis it is observed that different statistical indicators give different results regarding the capability of the generated TMY in representing the long-term mean values. The RMSD, U95 and Correlation coefficient show that the TMY based on typical days is the one which performs better regardless of the meteorological variable being analysed. The other

Table 7
RMSD and NRMSD of TMY based on typical months, five-day and days with reference to the long-term mean values for the Observatório station.

Comparison	TMY	RMSD				NRMSD	
		Air temp. (°C)	Rel. hum. (%)	Wind (m/s)	GHI (kWh/m^2-day)	Average	Weighted average
Monthly	Month	0.5709	2.8197	0.1616	**0.1173**	**0.0428**	0.0092
	Five-day	0.2281	1.3261	0.1804	0.4276	0.1010	0.0046
	Day	**0.0963**	**0.8701**	**0.1596**	0.3026	0.0715	**0.0026**
Five-day	Month	0.8100	4.9926	0.3985	0.7642	0.1940	0.0171
	Five-day	0.4839	3.9894	0.3156	0.5044	0.1309	0.0119
	Day	**0.3591**	**1.8445**	**0.2255**	**0.3858**	**0.0962**	**0.0065**
Daily	Month	1.0603	7.5894	0.6342	1.3995	0.3439	0.0282
	Five-day	0.7373	6.4633	0.5055	1.2912	0.3111	0.0220
	Day	**0.7186**	**4.6416**	**0.3976**	**0.5820**	**0.1537**	**0.0153**

Table 8
Average and weighted average of NRMSD for all the eight stations.

TMY	Average			Weighted average		
	Monthly	Five-day	Daily	Monthly	Five-day	Daily
Month	<u>0.0421</u>	0.1885	0.3269	<u>0.0103</u>	0.0231	0.0352
Five-day	0.1336	0.1705	0.3235	0.0104	0.0182	0.0296
Day	0.0598	**0.0892**	**0.1609**	**0.0062**	**0.0109**	**0.0218**

Table 9
Statistical indicators and GPI values of the generated TMYs for the Observatório station.

TMY	MBE	RMSD	U95	t-stats	R	GPI
Air Temp (°C)						
Month	0.1085	1.2910	3.5722	7.8922	0.9098	−4.1578
Five-day	−0.0312	0.9763	2.7055	2.9930	0.9523	0.1614
Day	**−0.0087**	**0.9217**	**2.5547**	**0.8798**	**0.9562**	**0.6808**
Rel. hum. (%)						
Month	0.6012	9.4114	26.0613	5.9907	0.3482	−1.9069
Five-day	**0.2623**	8.1388	22.5545	**3.0173**	0.4714	0.7742
Day	−0.3614	**6.3682**	**17.6380**	5.3193	**0.5974**	**2.3190**
Wind (m/s)						
Month	**0.0174**	0.9301	2.5798	**1.7488**	0.4687	−1.7015
Five-day	−0.1183	0.7778	2.1435	14.3987	0.5313	0
Day	−0.1254	**0.6617**	**1.8175**	18.0714	**0.6311**	**1.2985**
GHI (kWh/m^2-hour)						
Month	**0.0026**	0.1226	0.3399	**1.9618**	0.8950	0.0546
Five-day	−0.0084	0.1150	0.3183	6.8495	0.8972	0.5648
Day	0.0110	**0.0678**	**0.1867**	15.4102	**0.9709**	**1.6194**

statistical indicators do not present conclusive results on which TMY performs better. To identify the TMY that performs better, a Global Performance Index (GPI) was used, which allows the combination of different statistical indicators [44,45]. To determine the GPI of the generated TMYs, the values of all statistical indicators must be scaled down, ranging between a minimum value of 0 and a maximum value of 1. Then, the GPI can be determined using:

$$GPI_i = \sum_{j=1}^{5} \alpha_j \left(\tilde{y}_j - \tilde{y}_{ij} \right) \tag{6}$$

where α_j equals −1 for the correlation coefficient R and 1 otherwise. $\tilde{y}_j$ is the median of scaled values of the statistical indicator j and $\tilde{y}_{ij}$ is the scaled value of the statistical indicator j for the typical data set i. Lower values of GPI imply less accuracy of the generated TMYs while higher values of GPI mean that the generated TMYs represent more accurately the long-term statistics.

In the case of the Observatório station, the values of GPI show that the TMY based on typical days performs better than the TMY based on typical months and five-day typical periods for all the meteorological variables. The GPI also shows that the TMY based on typical months presents lower accuracy compared to the other two TMYs if hourly values are considered. The GPI values of GHI for all the stations analysed in this work are presented in Table 10. The GPI values for the remaining meteorological variables and stations can be found in Appendix B. The information presented in Table 10 clearly shows that the TMY based on typical days presents higher values of GPI, meaning that this TMY represents better the long-term hourly mean values, followed by the TMY based on typical months.

The GPI values for the remaining meteorological variables (see Appendix B) also show that the TMY based on typical days represents more accurately the long-term hourly mean values, with an average GPI of 0.7459, 1.6800 and 1.6869 for air temperature, relative humidity and wind speed, respectively. Averaging all the GPI values of all the meteorological variables and meteorological stations, mean GPI values of −1.1154, −0.1230 and 1.5207 are obtained for the TMY based on typical months, typical five-day periods and typical days, respectively. This clearly shows that the TMY based on typical days performs well and should be considered if the long-term hourly mean values are to be better represented by typical data. In Appendix C is presented the spatial distribution of the annual GHI in the Madeira Island.

As consequence of the lack of measurements of diffuse horizontal (DHI) and direct normal irradiance (DNI) across Madeira Island, a preliminary analysis using the typical periods of each typical year was performed to understand the impact of the used time resolutions on the corresponding typical DHI and DNI data. In this case, the DHI and DNI data was generated through the Boland-Ridley-Lauret model (BRL model) [46], and the present preliminary analysis was performed for the Observatório station, since some DHI measurements were available at that location in a shorter period than that of the entire GHI series. The available measured DHI data need to be corrected due to the shadow band effect, with the associated approximations, and thus the entire series of DHI

Table 10
GPI values of GHI for all the eight stations.

Stations	TMY		
	Month	Five-day	Day
Areeiro	0.8939	0.0911	**2.1972**
Caniçal	0.3728	−0.2581	**2.3691**
Lido	0.2949	0.0772	**1.3721**
Observatório	0.0546	0.5648	**1.6194**
L. Terça	0.0134	−0.7121	**2.2745**
L. Baixo	0.6606	0.2679	**1.9285**
P. Pargo	0.6050	−0.1822	**1.4229**
S. Jorge	0.0371	0.2196	**2.5781**
Average	0.3665	0.0085	**1.9702**

was obtained through the BRL model for consistency. The BRL model [46] uses the experimental clearness index based on the GHI data, the solar zenith angle and the apparent solar time to generate the diffuse fraction. By comparing the modelled and the corrected experimental DHI values, a good agreement was found. Then, using both GHI and DHI data and the cosine of the solar zenith angle, the hourly DNI data was obtained and the same analysis as that presented in Table 9 was also performed, which is presented in Table 11.

The TMY based on typical days performs better for both DHI and DNI data. Despite the lower MBE and t-statistics of the monthly TMY regarding the DHI data, the GPI of the daily TMY is higher. In the case of the DNI data, the monthly TMY only performs better according to the t-statistics. All the other statistical indicators as well as the GPI associate the daily TMY as the best performing typical year taking the entire hourly data series as reference. These results suggest that despite the DHI and DNI data were not used for the selection of the typical periods, the same conclusions can be drawn regarding the best performing typical year, i.e., the daily TMY represents in a more accurate way the long-term series of GHI, DHI and DNI.

5. Conclusions

In this paper, solar resource assessment is addressed through long-term statistical analysis and typical data generation with different time resolutions using data from eight meteorological stations in Madeira Island, Portugal. A modified Sandia method was used to generate Typical Meteorological Years (TMY) with three different typical time periods: months, five-days and days. Various statistical indicators were used to assess the generated typical data sets and a Global Performance Index (GPI) was defined. Although all the generated TMYs are in good agreement with the long-term averages, it was found that the TMY based on typical days is the one that presents the closest results to the long-term hourly data, either if all the variables included in the TMY generation are considered or not, while the TMY based on typical months performs better if only the monthly values of Global Horizontal Irradiation (GHI) are considered. Therefore, the use of the TMY based on typical days is suitable for energy systems modelling and it improves the representation of a large data series of hourly values on a single typical year, with a mean GPI of 1.5207, in opposition to GPI values of −1.1154 and −0.1230 for the TMYs based on typical months and typical five-day periods, respectively. This work provides an alternative method to generate typical data sets that can be used to study energy efficiency of buildings and renewable energy systems, including not only GHI data but also the main meteorological parameters that affect the performance of such systems. It was also found that Areeiro presents the highest annual and monthly averages of daily GHI (4.84 kWh/m^2-day and 7.71 kWh/m^2-day, respectively) while Lombo da Terça presents the lowest annual and monthly averages of daily GHI (4.03 kWh/m^2-day and 1.75 kWh/m^2-day, respectively). Solar resource assessment was performed using typical data generated through different time resolutions using GHI measurements only. It is of interest extending the statistical analysis performed in this work to locations where DHI and DNI measurements are available and to study the impact of using different typical years on the simulation of solar energy systems.

Table 11
Statistical indicators and GPI values for DHI and DNI for the Observatório station.

TMY	MBE	RMSD	U95	t-stats	R	GPI
DHI (kWh/m^2)						
Month	**−0.0010**	0.0458	0.1270	**0.0006**	0.9369	−0.7719
Five-day	0.0017	0.0438	0.1213	0.0010	0.9465	−0.1299
Day	0.0140	**0.0398**	**0.1068**	0.0097	**0.9755**	**0.2281**
DNI (kWh/m^2)						
Month	0.0049	0.1733	0.4802	**0.0007**	0.6847	−1.4999
Five-day	−0.0113	0.1610	0.4458	0.0018	0.6870	−3.0299
Day	**−0.0043**	**0.1077**	**0.2985**	0.0010	**0.8453**	**1.3002**

Acknowledgments

The authors acknowledge to the Instituto Português do Mar e da Atmosfera (IPMA) for providing the meteorological data for this study. The authors acknowledge the funding provided by the European Union through the European Regional Development Fund, included in the COMPETE 2020 (Operational Program Competitiveness and Internationalization) through the ICT project (UID/GEO/04683/2013) with the reference POCI-01-0145-FEDER-007690 and the project DNI-Alentejo with the reference ALT20-03-0145-FEDER-000011. This work has been funded by Portuguese Funds through the Foundation for Science and Technology under the project LAETA 2015–2020, reference UID/EMS/50022/2013.

Appendix A. Sandia method for the TMY generation based on typical months

The Sandia method requires the determination of the cumulative distribution function (CDF) for each candidate month in each individual year and for each of the meteorological parameters. In a general form, the CDF is given by:

$$S_n(x) = \begin{cases} 0 & if \quad x < x_{(1)} \\ (k-0.5)/n & if \quad x_{(k)} \le x \le x_{(k+1)} \\ 1 & if \quad x \ge x_{(n)} \end{cases} \tag{A.1}$$

where $S_n(x)$ is the value of the CDF for the daily value of the meteorological parameter x under consideration, n is the total number of elements (number of days in the month) and k is the ranked order number ($k = 1, 2, \ldots, n$ - 1). A long-term CDF is also determined using (A.1) for each month of the calendar and meteorological parameter by considering all the corresponding months in the data series, thus covering the entire period of measurements. In this case, the total number of elements, n, is the number of years of the data series times the number of days of the given calendar month. Then, the Finkelstein-Schafer (FS) statistics, which compares individual months and long-term data for the same calendar month, is given by:

$$FS = \frac{1}{N}\sum_{i=1}^{N}\delta_i \tag{A.2}$$

where δ_i is the absolute difference between the long-term and candidate month CDFs and N is the number of daily records for each month. After this, a weighted sum, WS, of the FS statistics for each month/year is calculated by considering different statistical weights for the meteorological parameters, given by:

$$WS = \sum_{j=1}^{9}\omega_j * FS_j \tag{A.3}$$

where ω_j and FS_j are, respectively, the statistical weight and the FS statistics of the meteorological parameter j. This allows setting the relative importance of each meteorological parameter in the selection of the TMMs, depending on the intended use for the TMY data. In the case of solar energy applications and energy studies in buildings the statistical weights are presented in Table 3 [16].

Then, the years of the data series are sorted in ascending order of *WS* and the five years with the lowest *WS* values are selected as candidates for each one of the calendar months. The *FS* values for the daily GHI and the *WS* values at Observatório are presented in Tables A.1 and A.2, respectively. The underlined values indicate the best five years for each calendar month, selected according to the criterion of lowest weighted sum of the Finkelstein-Schafer statistics. It can be seen that the five selected years does not correspond exactly to the five years with lower values of *FS* for the GHI, although this parameter is the one with highest statistical weight. This shows the combined effect of the other meteorological parameters on the selection of the candidate years, mainly the mean values of temperature, relative humidity and wind speed. The values in bold indicate the years that were selected for the TMY based on typical months according to the criterion described in the following.

Table A.1
FS statistics of the daily GHI for the Observatório station

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Jan	**0.666**	1.082	5.845	1.645	1.982	1.364	3.748	0.865	1.375	4.243	2.449
Feb	2.857	4.670	2.786	1.814	**1.969**	2.980	3.049	3.674	1.497	2.956	1.953
Mar	2.188	1.886	1.062	**1.358**	1.217	3.346	4.707	0.686	4.630	1.378	1.123
Apr	4.703	2.561	**1.267**	1.033	1.430	1.136	1.333	1.527	2.318	1.648	2.812
May	3.311	2.487	1.158	1.114	2.293	**0.921**	2.434	2.569	1.900	4.821	5.595
Jun	2.767	2.691	4.185	1.370	1.191	1.939	2.494	2.176	0.876	1.797	**0.752**
Jul	2.308	3.226	2.100	3.551	3.117	**1.422**	4.100	2.378	1.548	2.085	4.707
Aug	4.853	2.147	2.232	4.457	2.393	2.604	2.079	**0.783**	1.762	3.733	3.959
Sep	2.342	1.806	4.052	5.345	**0.876**	2.033	3.024	1.179	1.812	1.170	1.297
Oct	5.073	3.220	**1.745**	2.566	0.921	1.572	2.355	3.100	2.718	1.504	1.862
Nov	4.061	1.985	1.361	3.000	1.542	**0.873**	1.473	1.415	2.012	4.897	1.767
Dec	1.006	2.449	1.677	2.378	1.211	3.255	2.214	2.258	2.584	1.411	**1.053**

Table A.2
Weighted sum of the FS statistics for the Observatório station

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Jan	**1.623**	2.223	4.544	2.132	2.668	1.829	3.084	1.279	2.406	3.301	2.459
Feb	3.323	4.527	2.872	3.061	**2.451**	2.371	3.535	2.855	3.296	3.127	1.828
Mar	2.680	2.125	1.709	**1.760**	1.518	2.870	3.038	2.252	3.342	3.351	1.952
Apr	4.014	1.952	**1.491**	1.310	2.040	1.448	1.324	1.568	2.688	1.867	2.830
May	3.413	2.394	1.568	1.159	2.282	**1.427**	2.281	2.431	2.204	4.046	4.356
Jun	3.688	2.132	2.878	1.783	1.947	2.119	3.461	2.770	2.288	2.616	**2.045**
Jul	3.874	2.571	2.292	3.438	3.136	**2.064**	3.177	2.849	1.863	2.949	3.857
Aug	4.788	2.482	2.093	4.025	2.762	2.830	2.361	**1.416**	2.101	3.299	3.865
Sep	3.246	1.970	3.159	4.306	**1.712**	1.982	2.763	1.618	2.355	2.749	2.495
Oct	3.253	2.865	**2.162**	2.697	2.279	2.143	2.067	2.881	1.970	2.268	1.849
Nov	3.166	2.160	2.034	2.351	2.598	**1.653**	2.144	1.898	2.108	3.586	1.996
Dec	2.056	2.158	2.115	2.022	1.878	3.651	2.504	2.587	2.842	1.891	**1.950**

The next step of the Sandia method consists in evaluating the persistence of the global solar radiation and mean temperature of the air. This is done by determining the frequency and length of runs of consecutive days with values above and below fixed percentiles. However, a most simple procedure for selecting the TMM as proposed by Pissimanis [29] was adopted in the present work. This procedure starts by determining the deviations of the short-term mean values of global horizontal irradiation (GHI) with respect to the long-term values for each of the five candidate years of each calendar month. Since hourly values were available, the deviations were estimated through the Root Mean Square Difference of the mean hourly GHI distribution for each candidate year with respect to the long-term hourly distribution [29], given by:

$$RMSD = \sqrt{\sum_{l=1}^{P}\left(H_{y,l} - \overline{H}_l\right)^2 \Big/ P} \tag{A.4}$$

where P is the number of hours of the day with global solar radiation different from zero, $H_{y,l}$ is the hourly average of GHI for the candidate year y and hour l and $\overline{H}_l$ is the long-term mean for the hour l. The values of the *RMSD* for the five candidate years for each month are given in Table A.3 for the Observatório station. From these results, the year with the lowest *RMSD* and the years within the range $\min\{RMSD\} + 0.02\ kWh/m^2 - hour$ were kept. Then, the *FS* values for the daily global solar irradiation of those years were analysed and the year with lowest *FS* and the years within the range $\min\{FS\} + 0.03$ were selected. If that was not enough to determine the TMM, the *FS* values for the mean dry bulb temperature are also analysed and the year with the lowest *FS* value is selected. In this work, the analysis of the daily GHI was enough to select all the TMMs for all the locations in the island. The selected years for the Observatório station are represented in bold in Table A.3.

Table A.3
Root mean square difference of the mean hourly GHI ($kWh/m^2 - hour$) for the Observatório station

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Jan	**0.038**	0.070	–	0.045	–	0.040	–	0.034	–	–	–
Feb	–	–	0.143	–	**0.068**	0.117	–	0.194	–	–	0.109
Mar	–	0.074	0.133	**0.057**	0.108	–	–	–	–	–	0.131
Apr	–	–	**0.053**	0.077	–	0.132	0.045	0.070	–	–	–
May	–	–	0.066	0.062	–	**0.062**	0.069	–	0.151	–	–
Jun	–	0.117	–	0.060	0.082	0.138	–	–	–	–	**0.046**
Jul	–	0.196	0.138	–	–	**0.053**	–	0.146	0.111	–	–
Aug	–	0.116	0.107	–	–	–	0.135	**0.053**	0.082	–	–
Sep	–	0.079	–	–	**0.064**	0.090	–	0.063	0.090	–	–
Oct	–	–	**0.078**	–	–	0.101	0.121	–	0.118	–	0.087
Nov	–	–	0.043	–	–	**0.048**	–	0.091	0.073	–	0.074
Dec	0.068	–	–	0.055	0.039	–	–	–	–	0.032	**0.042**

Consecutive months are concatenated to generate the Typical Meteorological Year composed of 8760 hourly values for each meteorological parameter. An interpolation in the period between the 6 h before and 6 h after the end of each month was also carried out to obtain a smooth transition of the meteorological parameters, by using a centred moving average.

Appendix B. GPI values for air temperature, relative humidity and wind speed

Table B.1
GPI values of air temperature for the eight stations in Madeira Island

Station	TMY		
	Month	Five-day	Day
Areeiro	−1.8212	−0.0383	**1.2170**
Caniçal	−1.3302	0.3266	**1.2475**
Lido	−1.3908	**0.3139**	−0.0768
Observatório	−4.1578	0.1614	**0.6808**
L. Terça	−1.8573	−0.3544	**0.6644**
L. Baixo	−1.3999	**0.1031**	−0.2969
P. Pargo	−2.3631	−0.0284	**1.7911**
S. Jorge	−1.7842	0.4094	**0.7398**
Average	−2.0131	0.1117	**0.7459**

Table B.2
GPI values of relative humidity for the eight stations in Madeira Island

Station	TMY		
	Month	Five-day	Day
Areeiro	0.5304	−0.1287	**1.6591**
Caniçal	−0.3191	−2.1223	**1.6591**
Lido	−2.4854	0.9768	**1.5378**
Observatório	−1.9069	0.7742	**2.3190**
L. Terça	−1.2099	−2.0161	**0.1938**
L. Baixo	−1.0176	0.1575	**1.4432**
P. Pargo	−2.0821	−0.0568	**2.8611**
S. Jorge	−1.4430	−0.2102	**1.7672**
Average	−1.2417	−0.3282	**1.6800**

Table B.3
GPI values of wind speed for the eight stations in Madeira Island

Station	TMY		
	Month	Five-day	Day
Areeiro	−1.5566	0.0911	**2.1972**
Caniçal	−0.8456	−1.9086	**1.9370**
Lido	−1.0964	0.0000	**1.9036**
Observatório	−1.7015	0.0000	**1.2985**
L. Terça	−0.9291	−0.1891	**3.8818**
L. Baixo	−2.6157	0.0000	**0.3843**
P. Pargo	−1.6449	−0.2647	**1.0904**
S. Jorge	−2.1979	0.0000	**0.8021**
Average	−1.5735	−0.2839	**1.6869**

Appendix C. Spatial distribution of the solar resource in Madeira Island

Aiming a better understanding of the geographic distribution of the solar resource in the Madeira Island, a spatial interpolation was carried out to estimate the total annual GHI in the entire island, using the data from TMY based on typical months for all the eight sites analysed, as shown in Figure C.1.

The annual GHI is higher in the center of the island (Areeiro) and in the south coast (Lugar de Baixo and Ponta do Pargo). The north coast presents lower GHI values (Lombo da Terça and São Jorge). One of the reasons for the higher values of GHI in Areeiro is the lower absorption and scattering of radiation in the atmosphere, because Areeiro is located at higher altitude. Another possible explanation is that clouds may form below that altitude, which needs to be studied in detail to be confirmed. The area north of Lugar de Baixo, in the central part of the island, is the area with higher uncertainty due to the distribution of the available stations for analysis. In addition, when more GHI data for the stations with shorter time series become available, the generated TMYs can be updated, thus improving its representativeness, while some other stations that were not considered in this work can be included.

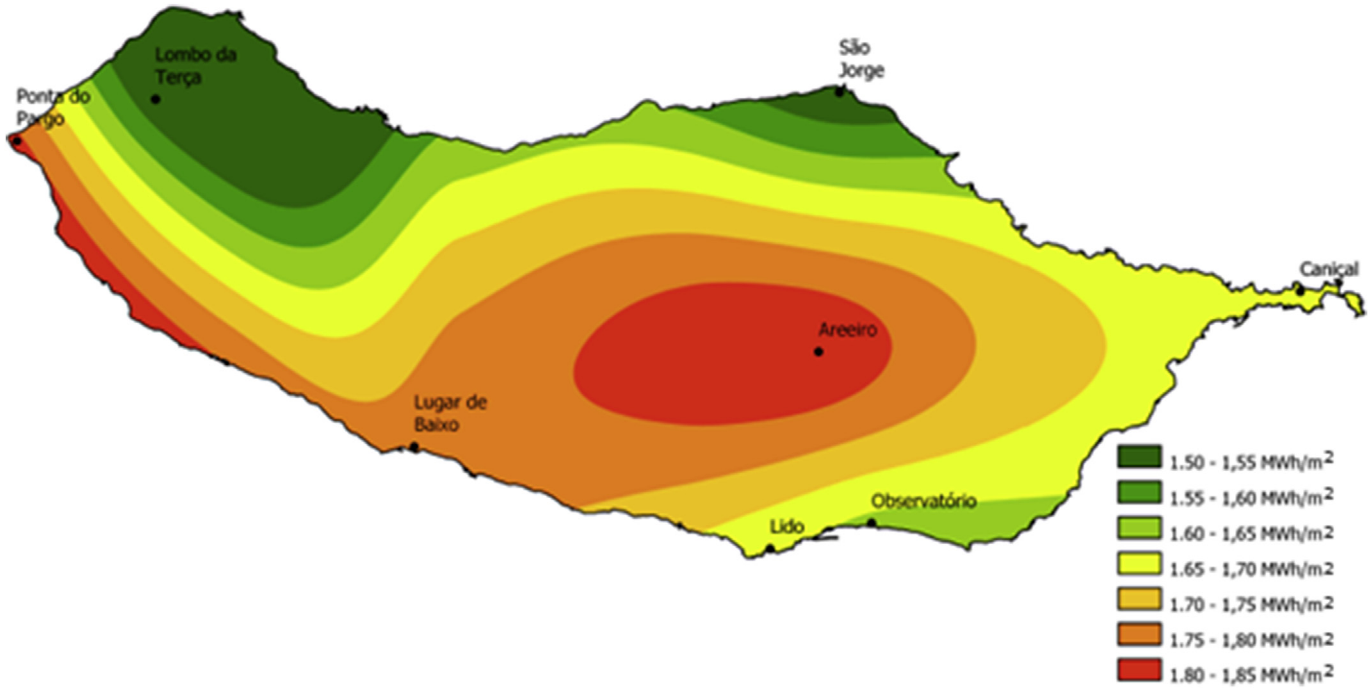

Fig. C.1. Madeira Island: Spatial distribution of annual global solar irradiation.

References

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

[2] C.A. Gueymard, S.M. Wilcox, Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data, Sol. Energy 85 (2011) 1068–1084, https://doi.org/10.1016/j.solener.2011.02.030.

[3] M. Journée, R. Muller, C. Bertrand, Solar resource assessment in the Benelux by merging Meteosat-derived climate data and ground measurements, Sol. Energy 86 (2012) 3561–3574, https://doi.org/10.1016/j.solener.2012.06.023.

[4] E. Zawilska, M.J. Brooks, An assessment of the solar resource for Durban, South Africa, Renew. Energy 36 (2011) 3433–3438, https://doi.org/10.1016/j.renene.2011.05.023.

[5] J.a. Duffie, W. a. Beckman, W.M. Worek, Solar Engineering of Thermal Processes, 4nd ed., John Wiley & Sons, 2013.

[6] J.A. Ruiz-Arias, S. Quesada-Ruiz, E.F. Fernández, C.A. Gueymard, Optimal combination of gridded and ground-observed solar radiation data for regional solar resource assessment, Sol. Energy 112 (2015) 411–424, https://doi.org/10.1016/j.solener.2014.12.011.

[7] M. David, L. Adelard, P. Lauret, F. Garde, A method to generate Typical Meteorological Years from raw hourly climatic databases, Build. Environ. 45 (2010) 1722–1732, https://doi.org/10.1016/j.buildenv.2010.01.025.

[8] A.L.S. Chan, Generation of typical meteorological years using genetic algorithm for different energy systems, Renew. Energy 90 (2016) 1–13, https://doi.org/10.1016/j.renene.2015.12.052.

[9] E.F.M. Abreu, P. Canhoto, V. Prior, R. Melício, Assessment of PV systems performance in the Madeira island using typical meteorological year data, in: Conf. Electron. Telecommun. Comput., Lisbon, Portugal, 2016. hdl.handle.net/10174/19283.

[10] ISO, EN ISO 15927-4:2005, Hygrothermal Performance of Buildings — Calculations and Presentation of Climatic Data — Part 4: Hourly Data for Assessing the Annual Energy Use for Heating and Cooling, European Committee for Standardization, 2005.

[11] I. García, J.L. Torres, Assessment of the adequacy of EN ISO 15927-4 reference years for photovoltaic systems, Prog. Photovoltaics Res. Appl. 23 (2015) 1956–1969, https://doi.org/10.1002/pip.2617.

[12] K. Lee, H. Yoo, G.J. Levermore, Generation of typical weather data using the ISO Test Reference Year (TRY) method for major cities of South Korea, Build. Environ. 45 (2010) 956–963, https://doi.org/10.1016/j.buildenv.2009.10.002.

[13] R. Layi Fagbenle, Generation of a test reference year for Ibadan, Nigeria, Energy Convers. Manag. 36 (1995) 61–63, https://doi.org/10.1016/0196-8904(94)00039-3.

[14] T. Kalamees, J. Kurnitski, Estonian test reference year for energy calculations, Proc. Est. Acad. Sci. Eng. 12 (2006) 40–58.

[15] T. Lhendup, S. Lhundup, Comparison of methodologies for generating a typical meteorological year (TMY), Energy Sustain. Dev. 11 (2007) 5–10, https://doi.org/10.1016/S0973-0826(08)60571-2.

[16] I. Hall, R. Praire, H. Anderson, E. Boes, Generation of Typical Meteorological Year for 26 SOLMET Stations, 1978.

[17] W. Marion, K. Urban, User's Manual for TMY2s—typical Meteorological Years, 1995.

[18] S. Wilcox, W. Marion, Users Manual for TMY3 Data Sets, 2008 doi:NREL/TP-581-43156.

[19] H. Lund, The Design Reference Year Users Manual, 1995. Lyngby.

[20] R. Festa, C.F. Ratto, Proposal of a numerical procedure to select Reference Years, Sol. Energy 50 (1993) 9–17, https://doi.org/10.1016/0038-092X(93)90003-7.

[21] L.W. Crow, Weather year for energy calculations, ASHRAE 26 (1984) 42–47.

[22] A. De Miguel, J. Bilbao, Test reference year generation from meteorological and simulated solar radiation data, Sol. Energy 78 (2005) 695–703, https://doi.org/10.1016/j.solener.2004.09.015.

[23] M. Gazela, E. Mathioulakis, A new method for typical weather data selection to evaluate long-term performance of solar energy systems, Sol. Energy 70 (2001) 339–348, https://doi.org/10.1016/S0038-092X(00)00151-1.

[24] H. Yang, Y. Li, L. Lu, R. Qi, First order multivariate Markov chain model for generating annual weather data for Hong Kong, Energy Build. 43 (2011) 2371–2377, https://doi.org/10.1016/j.enbuild.2011.05.035.

[25] K. Skeiker, Comparison of methodologies for TMY generation using 10 years data for Damascus, Syria, Energy Convers. Manag. 48 (2007) 2090–2102, https://doi.org/10.1016/j.enconman.2006.12.014.

[26] S. Janjai, P. Deeyai, Comparison of methods for generating typical meteorological year using meteorological data from a tropical environment, Appl. Energy 86 (2009) 528–537, https://doi.org/10.1016/j.apenergy.2008.08.008.

[27] A. Ebrahimpour, M. Maerefat, A method for generation of typical meteorological year, Energy Convers. Manag. 51 (2010) 410–417, https://doi.org/10.1016/j.enconman.2009.10.002.

[28] K. Kulesza, Comparison of typical meteorological year and multi-year time series of solar conditions for Belsk, central Poland, Renew. Energy 113 (2017) 1135–1140, https://doi.org/10.1016/j.renene.2017.06.087.

[29] D. Pissimanis, G. Karras, V. Notaridou, K. Gavra, The generation of a "typical meteorological year" for the city of Athens, Sol. Energy 40 (1988) 405–411, https://doi.org/10.1016/0038-092X(88)90095-3.

[30] S.A. Kalogirou, Generation of typical meteorological year (TMY-2) for Nicosia, Cyprus, Renew. Energy 28 (2003) 2317–2334, https://doi.org/10.1016/S0960-1481(03)00131-9.

[31] K. Skeiker, Generation of a typical meteorological year for Damascus zone using the Filkenstein-Schafer statistical method, Energy Convers. Manag. 45 (2004) 99–112, https://doi.org/10.1016/S0196-8904(03)00106-7.

[32] A.L.S. Chan, T.T. Chow, S.K.F. Fong, J.Z. Lin, Generation of a typical meteorological year for Hong Kong, Energy Convers. Manag. 47 (2006) 87–96, https://doi.org/10.1016/j.enconman.2005.02.010.

[33] S. Pusat, I. Ekmekçi, M.T. Akkoyunlu, Generation of typical meteorological year for different climates of Turkey, Renew. Energy 75 (2015) 144–151, https://doi.org/10.1016/j.renene.2014.09.039.

[34] F. Bre, V.D. Fachinotti, Generation of typical meteorological years for the argentine littoral region, Energy Build. 129 (2016) 432–444, https://doi.org/10.1016/j.enbuild.2016.08.006.
[35] S. Murphy, The construction of a modified Typical Meteorological Year for photovoltaic modeling in India, Renew. Energy 111 (2017) 447–454, https://doi.org/10.1016/j.renene.2017.04.033.
[36] J.M. Finkelstein, R.E. Schafer, Improved goodness-of-fit tests, Biometrika 58 (1971) 641–645, https://doi.org/10.1093/biomet/58.3.641.
[37] H. Bulut, Typical solar radiation year for southeastern Anatolia, Renew. Energy 29 (2004) 1477–1488, https://doi.org/10.1016/j.renene.2004.01.004.
[38] J. Zhou, Y. Wu, G. Yan, Generation of typical solar radiation year for China, Renew. Energy 31 (2006) 1972–1985, https://doi.org/10.1016/j.renene.2005.09.013.
[39] H. Zang, Q. Xu, H. Bian, Generation of typical solar radiation data for different climates of China, Energy 38 (2012) 236–248, https://doi.org/10.1016/j.energy.2011.12.008.
[40] A. Chazarra, A. Mestre, V.C. Pires, S. Cunha, Á. Silva, J. Marques, F. Carvalho, M.T. Mendes, J. Neto, L. Mendes, L.F. Nunes, Atlas Climático dos arquipélagos das Canárias, da Madeira e dos Açores, 2011 nipo: 281-12-006-X.
[41] E.F.M. Abreu, Avaliação do recurso solar e modelação de um sistema de energia fotovoltaica ligado à rede elétrica da ilha da Madeira, Universidade de Évora, Portugal, 2015 hdl.handle.net/10174/18417.
[42] E.F.M. Abreu, P. Canhoto, R. Melicio, Tratamento de dados associados com a radiação solar, Évora, 2016 hdl.handle.net/10174/19214.
[43] C.N. Long, E.G. Dutton, BSRN Global Network Recommended QC Tests, V2.0, 2002.
[44] B. Jamil, N. Akhtar, Comparative analysis of diffuse solar radiation models based on sky-clearness index and sunshine period for humid-subtropical climatic region of India: a case study, Renew. Sustain. Energy Rev. 78 (2017) 329–355, https://doi.org/10.1016/j.rser.2017.04.073.
[45] M. Despotovic, V. Nedic, D. Despotovic, S. Cvetanovic, Review and statistical analysis of different global solar radiation sunshine models, Renew. Sustain. Energy Rev. 52 (2015) 1869–1880, https://doi.org/10.1016/j.rser.2015.08.035.
[46] B. Ridley, J. Boland, P. Lauret, Modelling of diffuse solar fraction with multiple predictors, Renew. Energy 35 (2010) 478–483, https://doi.org/10.1016/j.renene.2009.07.018.