A novel procedure for generating solar irradiance TSYs

Cite as: AIP Conference Proceedings **1850**, 140015 (2017); https://doi.org/10.1063/1.4984523
Published Online: 27 June 2017

Vicente Lara Fanego, Jesús Pulgar Rubio, Carlos M. Fernández Peruchena, Martín Gastón Romeo, Sara Moreno Tejera, Lourdes Ramírez Santigosa, Rita X. Valenzuela Balderrama, Luis F. Zarzalejo Tirado, Diego Bermejo Pantaleón, Manuel Silva Pérez, Manuel Pavón Contreras, Ana Bernardos García, and Sergio Macías Anarte

A Novel Procedure for Generating Solar Irradiance TSYs

Vicente Lara Fanego[1, a)], Jesús Pulgar Rubio[2], Carlos M. Fernández Peruchena[3], Martín Gastón Romeo[3], Sara Moreno Tejera[4], Lourdes Ramírez Santigosa[5], Rita X. Valenzuela Balderrama[5], Luis F. Zarzalejo Tirado[5], Diego Bermejo Pantaleón[6], Manuel Silva Pérez[4], Manuel Pavón Contreras[4], Ana Bernardos García[3], Sergio Macías Anarte[7]

[1]PM at SynerMet Weather Solutions S.L., c/ Los Almendros 12, 23002 Jaén (Spain), (+34) 953212474
[2]SynerMet Weather Solutions S.L., c/ Los Almendros 12, 23002 Jaén (Spain)
[3]CENER, c/ Isaac Newton 4 - Pabellón de Italia, 41092 Seville,(Spain)
[4]GTER, Avd. de los Descubrimientos s/n, 41092 Seville,(Spain)
[5]CIEMAT, Avd. Complutense 40, 28040, Madrid (Spain)
[6]IrSOLaV, Paseo de Almería 63, 6º, 04001, Almería (Spain)
[7]Abengoa Solar, c/ de la Energía Solar 1, 41014 Sevillle (Spain)

[a)] Corresponding author: vlara@synermet.com

Abstract. Typical Solar Years (TSYs) are key parameters for the solar energy industry. In particular, TSYs are mainly used for the design and bankability analysis of solar projects. In essence, a TSY intends to describe the expected long-term behavior of the solar resource (direct and/or global irradiance) into a condensed period of one year at the specific location of interest. A TSY differs from a conventional Typical Meteorological Year (TMY) by its absence of meteorological variables other than solar radiation. Concerning the probability of exceedance (Pe) needed for bankability, various scenarios are commonly used, with Pe90, Pe95 or even Pe99 being most usually required as unfavorable scenarios, along with the most widely used median scenario (Pe50). There is no consensus in the scientific community regarding the methodology for generating TSYs for any Pe scenario. Furthermore, the application of two different construction methods to the same original dataset could produce differing TSYs. Within this framework, a group of experts has been established by the Spanish Association for Standardization and Certification (AENOR) in order to propose a method that can be standardized. The method developed by this working group, referred to as the EVA method, is presented in this contribution. Its evaluation shows that it provides reasonable results for the two main irradiance components (direct and global), with low errors in the annual estimations for any given Pe. The EVA method also preserves the long-term statistics when the computed TSYs for a specific Pe are expanded from the monthly basis used in the generation of the TSY to higher time resolutions, such as 1 hour, which are necessary for the precise energy simulation of solar systems.

INTRODUCTION

A Typical Solar Year (TSY) is a commonly used tool for design and bankability analysis of solar energy projects, such as CSP, and in many other fields. A TSY is similar to the more ubiquitous Typical Meteorological Year (TMY), but only includes solar radiation data, such as direct normal irradiance (DNI) for CSP applications, whereas a TMY also contains information about many additional meteorological variables, such as temperature, humidity, or wind speed. Early developments of TSYs were made at National Renewable Energy Laboratory (NREL) for either DNI or global horizontal irradiance (GHI) [1]. They were referred to as TDY and TGY, respectively. Although the use of TMY or TSY is not recommended by experts for the precise design of solar-based renewable energy conversion systems [2, 3], such as CSP or PV, it has nevertheless become standard practice in the evaluation of the economic feasibility of such projects. Similarly to TMYs, TSYs are representative annual time-

SolarPACES 2016
AIP Conf. Proc. 1850, 140015-1–140015-8; doi: 10.1063/1.4984523
Published by AIP Publishing. 978-0-7354-1522-5/$30.00

series of typical solar radiation conditions expected at a specific location over a long time period, ideally the lifetime of the facility. A TSY does not correspond to any particular year of a specific period, but is artificially built as a composition of twelve "typical" months selected from different natural years, themselves extracted from a long-term time series of solar radiation data at the location of interest. The selection of those months to be used in the generation of the TSY involves a statistical characterization of the long-term time series of the dataset. The theoretical features of this synthetic year facilitate the objective quantification of the expected solar irradiance for different annual scenarios of available energy amount, usually measured by the probability of exceedance (Pe), and the corresponding estimated uncertainty. These valuable properties allow for an objective risk balance and safer analysis of the economic feasibility of the projects. Therefore, reliable TSYs for different solar resource scenarios—usually low-energy or "near worst case" (Pe90 to Pe99) and average or median (Pe50)—are demanded by the industry. Unfortunately, there is no scientific consensus on a standard method to generate such TSYs outside of the conventional Pe50 scenario. In fact, most likely only one method seems to have been proposed so far for the construction of TSYs corresponding to any probability scenario [4]. Nevertheless, the different strategies followed by the existing or future methods could undesirably generate different TSYs and thus different financing risk factors for the same initial meteorological dataset.

Within this framework, the Spanish Association for Standardization and Certification (AENOR) has established a working group of experts with the goal to design and standardize a method to generate TSYs for any solar energy scenario, specifically for applications in solar thermal power plants. In a first step, Fernandez-Peruchena et al. [5] recently presented a study that showed that any Pe (hereafter, Pexx) could be inferred by the estimation of the continuous cumulative distribution functions (CDF) evaluated from long-term annual series of data of GHI and DNI. The present contribution presents a novel procedure for the selection of the most appropriate individual months (among all those available in the long-term time series) to generate a TSY for any particular Pexx.

METHODOLOGY

For the generation of TSYs, the complete methodology must comprise two parts. In the first part, the annual values of each variable in the long-term time series are calculated. With this discrete number of annual values the estimated continuous CDF is derived using a Weibull distribution. The estimation of the parameters of the Weibull annual distribution is performed here using the *fitdistrplus* package [6] in R version 3.2.4 ("R: The R Project for Statistical Computing," 2003) with the maximum likelihood method. This procedure is well described and analyzed in [5]. This estimated CDF provides the values for any annual probability of exceedance (Pexx), hence for any desired scenario, corresponding to design, bankability, etc. It should be clarified here that there is a common acceptation of using the concepts of probability of exceedance and percentile (normally referred to as Pe and P, respectively) as if they were the same. Both are complementary but should not be confused: for a determined percentile value the probability of exceeding that value is the complementary to 100%; for instance, for a percentile 5, the probability of being exceeded is 95%. Therefore it would not be correct to use a percentile 95 to refer to a scenario of low energy. Conversely, a probability of exceedance of 95% (Pe95) more appropriately means that the value will be exceeded 95% of the time. In this work we used the concept of probability of exceedance instead of percentile, because it is more commonly used by the industry.

The second part of the method corresponds more specifically to the analysis presented in this work, and is referred to as the EVA method, which is an acronym constructed from the Spanish words for seasonality and variability. Once the annual target value for a specific Pexx of interest is obtained from the estimated continuous CDF in the first part, the next step is to concatenate a subset of twelve calendar months from the long-term time series of the original data, which might be measured (preferably) or modeled. Ultimately, this ensemble constitutes a TSY for the scenario determined by the target Pexx value. Therefore, the objective of the method is to determine which subset of twelve months shall be extracted from the complete long-term dataset. The EVA method is composed of two stages. In the first stage, the aim is to find those monthly values that respect two requisites: (i) in combination they must be statistically representative of the desired Pexx scenario; and (ii) their annual sum must correspond to the Pexx target value. Those monthly values are called "monthly expected values" (MEV), and they do not have to be necessarily equal to any of the available monthly values of the long-term time series. With this definition, the sum of all MEVs is exactly equal to the annual target value of Pexx. To determine these monthly values the method uses the following conditions:

1. For each month, the MEV should be the particular value that minimizes its distance to the corresponding monthly median (least-squares equation).
2. The MEV annual sum must be equal to the annual target value of Pexx (binding).
3. The intra-annual statistics that describe the natural behavior of the solar irradiance at the location of study is introduced by a composition of weights that conveniently modifies the least-squares equation. These weights, noted w_i, are determined by the product of two factors. The first one (${f^1}_i$) measures the variability of the irradiance during each month relative to the others. Prior to obtain these values the seasonality of the monthly time series must be removed by means of a clear-sky model. For this study, the Bird clear-sky model described by Iqbal (1983) as Model C [7], particularized to a clean and dry atmosphere version, is used. As a measure of the monthly variability the statistic named median absolute deviation (MAD) is computed for each month. The second factor (${f^2}_i$) accounts for each individual monthly energy contribution relative to the total annual energy. It is calculated as the mean of the monthly values of each month available in the original time series. The weights are combined by the product: $w_i = {f^1}_i \cdot {f^2}_i$, for $i = 1, \dots, 12$.

From a mathematical standpoint, the combination of the three conditions above can be described as a minimization problem with constraint. In other words, the problem consists in the minimization of the function:

$$f(x_1, \dots x_{12}) = \sum_{i=1}^{12} \left(\frac{\sum_{i=1}^{12} w_i}{w_i}\right) \left(x_i - Pe_i^{50}\right)^2 \qquad (1)$$

With the following constraint:

$$\sum_{i=1}^{12} x_i = Pe_y \qquad (2)$$

Where:

$i = 1, \dots, 12$; month of the year.

x_i: monthly expected value (MEV) for month i.

Pe_i^{50}: median of the available values of month *i* relative to the long-term time series.

Pe_y: annual probability of exceedance at the *y* level (Pexx).

w_i: weight.

This minimization problem is analytically resolved by the method of Lagrange multipliers. After application of the procedure an equation is finally obtained. The unknowns of this resulting equation are the 12 MEV(x_i) values.

The second stage simply consists in finding the available monthly values that are closest in distance to the corresponding MEV. These distances, called residuals, are obtained as the absolute value of the difference between the MEV and the available monthly solar radiation values. Finally, the 12 selected months constitute the desired TSY for the specific Pexx of interest.

In summary, this method can be said to be statistically based and analytically resolved. Because of its general definition it can be applied for both components GHI and DNI. Moreover, it should be pointed out that the method makes no assumptions about the monthly distributions. This is a key factor that justifies the selection of MAD as a measure of variability, since it is a robust statistic. Finally, it should be noted that the normalized weights are introduced inverted in function (1) by means of the factor ($\sum_{i=1}^{12} w_i / w_i$). Therefore, for the ith month, a low value of w_i (low variability and low energy) makes the ith MEV value closer to the median value of the corresponding monthly distribution, comparatively to MEVs whose w_i are higher.

RESULTS

The EVA method has been applied to a wide sample of different climatic locations around the world [5]. The selection of these stations is based on the availability of long-term time series (at least 20 years) of high-quality data of surface solar irradiance. The results presented below correspond to the evaluation that has been carried out for DNI and GHI at the Burns radiometric station (BRN, 43.52 ºN, 119.02 ºW) of the University of Oregon's Solar Radiation Monitoring Laboratory (UO-SLMR, http://solardat.uoregon.edu/). The high quality of the original data has been reinforced through additional quality checks [8]. Note that the irradiance measurements had differing time resolutions depending on period (starting at hourly in 1980 and finishing at 5-min from 1995 on). For consistency, the time series has been homogenized to the conventional hourly basis throughout. Hourly data were then integrated to obtain monthly and yearly values. For this study, the time series of both DNI and GHI covered the complete 33-year period (1980–2012). The procedure has been applied to derive TSYs for a wide range of Pexx values, including the most commonly ones required by the industry, namely: Pe99, Pe95, Pe90 and Pe50. For validation purposes, the generated TSYs have been analyzed against the initial long-term time series in order to consider the representativeness of each single artificial year under each possible Pexx scenarios.

To obtain an intuitive perspective of the problem, a possible way is to show the distributions of the monthly irradiation values, since the calendar month is the working unit of time adopted by the method. In Fig. 1, the distribution of the monthly DNI values is shown by means of boxplot diagrams. Different curves linking the twelve calendar months of a year would enclose the area of the total annual energy amount corresponding to different Pexx scenarios. Hence, for any specific value of this amount of energy determined by a certain Pexx, there is a set of curves whose integral approaches the annual target value of that specific Pexx. The aim of the method is to determine the shape of the particular curve that meets the former condition of the integral and provides the most representative behavior of the long-term irradiation among all possibilities.

In Fig. 1, the boxplot diagrams show the set of twelve distributions of monthly DNI values at the Burns station. For each boxplot, the interquartile range (IQR) amounts to 50% of the data, between Q1 (25%) and Q3 (75%), and the whiskers represent the quantity 1.5·IQR. The mean and median values of each monthly distribution are highlighted. Interestingly, each monthly distribution is different from the others, and a normal distribution cannot always be assumed in all cases. For instance, in May and September at the Burns station, some values are outside of the whiskers interval and can be considered outliers. These circumstances justify the preference of using the robust MAD statistic rather than the usual standard deviation as a measure of the variability of the monthly distributions. Even though the standard deviation is also a measure of variability in data samples, it presents two important disadvantages when the distribution cannot be assumed Gaussian: outliers can strongly influence the standard deviation value, and the standard deviation can force a preference for lower vs. higher values, or vice versa. Figure 1 also shows how the natural seasonal tendency of the time-series exhibits a strong pattern with higher energy values during summer months, which is typical of temperate climates. In order to properly compare the variability of the 12 monthly distributions, the monthly dataset has been seasonally adjusted by means of the clear-sky model. Figure 1 presents the MEV values calculated with the EVA method for the "near worst-case" scenarios of Pe99 and Pe90 along with the MEV values obtained with the simplified method that would only use the variability factor ${f^1}_i$ to configure the weights. When only this variability factor ${f^1}_i$ is taken into account, the winter months (December, January and February) have higher weights due to their greater variability (note that, as defined in the Methodology section, the higher the variability the higher the weight, and thus the farther the MEV is to the median value of the monthly distribution, because of the special way the weights are defined in function (1)) Because winter months are naturally less energetic than summer months, MEVs for winter months should have very low values to compensate for the contribution of the less variable and more energetic summer months (June, July and August). In other words, winter months have to contribute with extremely low energy values relative to the total annual energy amount in order to achieve the unfavorable low-energy scenarios Pe99 or Pe90. This occurs in detriment of the higher energetic (but less variable) summer months, which cannot contribute to the low annual energy amount with low monthly values, but with energy values close to the median. As can be seen in Fig. 1, this is more pronounced for the extreme Pe99 case than for the milder Pe90 case. Counting only on variability factor ${f^1}_i$ by ignoring ${f^2}_i$ can produce a misrepresentation of the possible contribution of the high-energy months to the total amount of annual energy in unfavorable cases, such as those determined by scenarios Pe99 or Pe90. The high-energy months (as dependent on seasonality) should be considered a major potential contributing factor when extreme years of DNI and GHI must be constructed. This is simply because low values of high-energy months can notably reduce the available energy of the whole year, and should therefore be taken into account. With the EVA method, this is done by means of the energy factor ${f^2}_i$. As shown in Fig. 1, this produces a different distribution of the contribution of each month relative to the annual Pexx value, depending on the variability and energy for that month. Thus, winter months may reach low DNI values without being extreme, whereas summer months may reach lower values instead of being forced to be closer to the median. In particular, Fig. 1 shows that, for the extreme case represented by Pe99, the MEV values for all months are below their respective IQRs. In contrast, when the method is applied using only the first variability factor $f^1{}_i$, in July the MEV value is still almost within its IQR, whereas extreme values are reached in January and December.

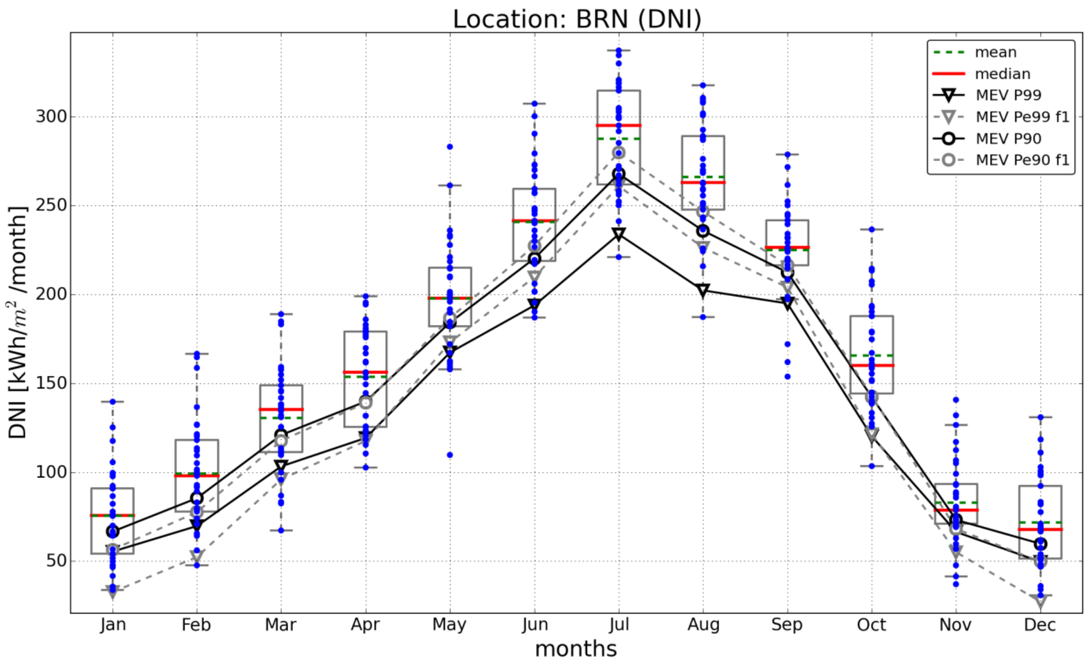

FIGURE 1. Boxplot diagrams of monthly DNI values for the complete dataset measured at the Burns (BRN) station. The IQR of the boxplots is the interquartile [Q1, Q3], and the whiskers equal to 1.5·IQR. The monthly mean and median values of the distribution are shown (green dashed and continuous red line, respectively). The MEV values for Pe99 and Pe90 annual values, along with those calculated from only the variability factor $f^1{}_i$ are also plotted.

By definition, the sum of the 12 MEVs is exactly equal to the target annual Pexx value. When minimizing the residuals, an error is inevitably introduced because the MEV values do not usually coincide with the available monthly values of the long-term time series, so that the sum of the residuals is not zero. These errors are presented in Figs. 2 and 3 along with the annual values of the constructed TSY over the estimated CDF curve, for DNI and GHI respectively. As shown in both figures, the value of the errors differs according to the specific target Pexx value. Errors are higher for DNI than for GHI, as could be expected since the former has more interannual variability than the latter [9]. Absolute values of the relative errors (in percent relative to the Pexx value) fall within the ranges 0.03–1.90 and 0.00–0.20 for DNI and GHI, respectively. The highest error is produced in the extreme case (Pe99) for the DNI variable (Fig. 2). This suggests that the more extreme Pexx is, the higher the error. However, this is not necessarily true in general. In Fig. 3, for instance, the errors for Pe60 and Pe10 are higher than those for Pe99. Furthermore, Figs. 1 and 2 show that the errors are not always of the same sign: they can indicate either overestimation (negative error) or underestimation (positive error). In general, it can be said that the magnitude of the error depends on the number of years available in the long-term time series to generate the TSY, because a longer time series implies more possibilities of finding actual monthly values that are closer to the MEV values. Finally, it should be highlighted that the errors are very low in all the cases presented in Figs. 2 and 3, and also for the pool of locations analyzed elsewhere [5] (not shown). The errors are below the usual standard limits that have been established to account for slight corrections –usually consisting in day substitutions- to obtain a better approach to the target Pexx.

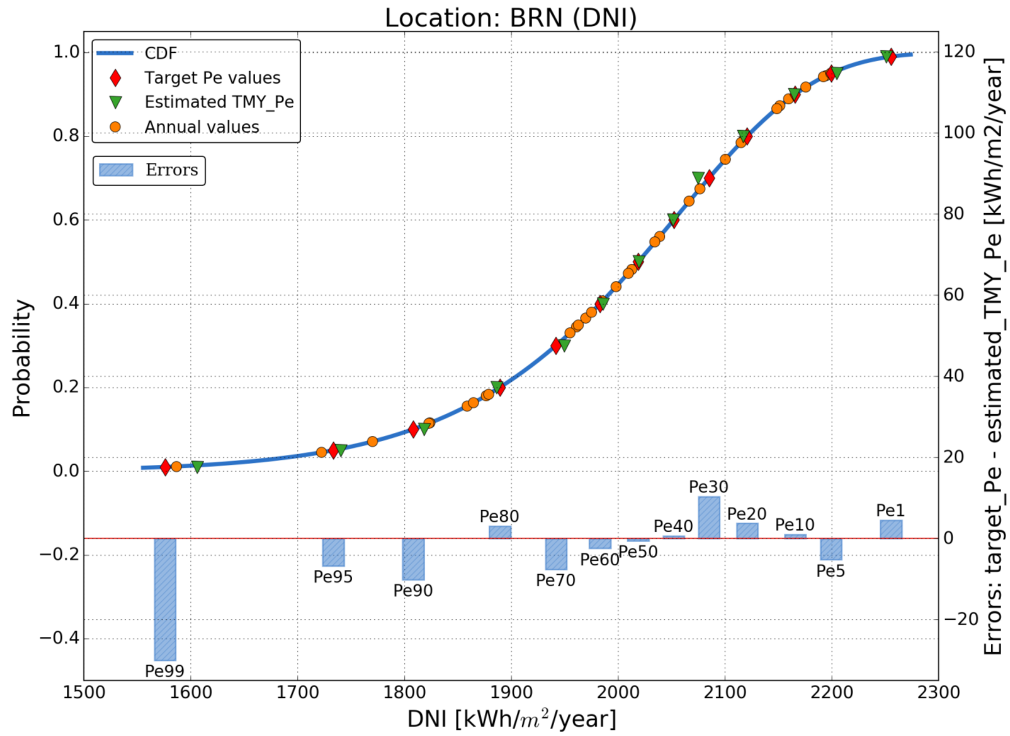

FIGURE 2. DNI annual values and estimated CDF at Burns (BRN) station, along with the annual target Pexx values and the estimated annual TSY-Pe values. Errors between the target and estimated TSY values for the different Pexx are also shown.

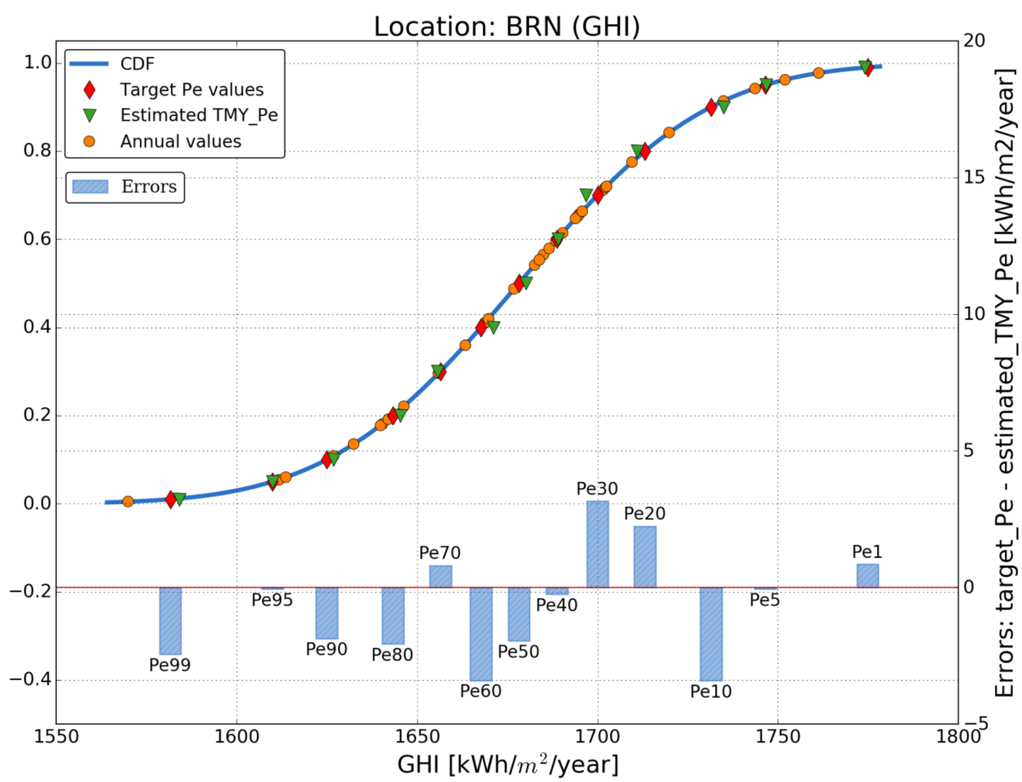

FIGURE 3. Same as Figure 2, but for GHI.

Finally, it is important to consider that, although the methodology to generate TSYs is based on a monthly time unit, the representative annual time series is most usually needed at higher time resolutions. In principle, since the TSY is constructed from blocks of monthly periods extracted from the original long-term time series, the original time resolution (e.g., hourly) should be conserved. In fact, this desirable feature of the TSY both restricts and determines the way it is defined. To evaluate the capability of the EVA method to generate representative years for higher time resolutions, the constructed TSYs are set to the original hourly resolution of their constituent months in order to compare the frequency distribution of the hourly TSYs with that of the long-term. Figure 4 shows the frequency distributions of TSY for Pe95 and Pe50 in the case of DNI at hourly time resolution, compared to that of the original long-term time series. The figure shows that the shapes of the frequency distributions of the TSYs are quite similar to that of the long-term. The reduction in annual energy from the Pe50 to the Pe95 scenarios is mainly produced by the higher DNI values, which appears logical.

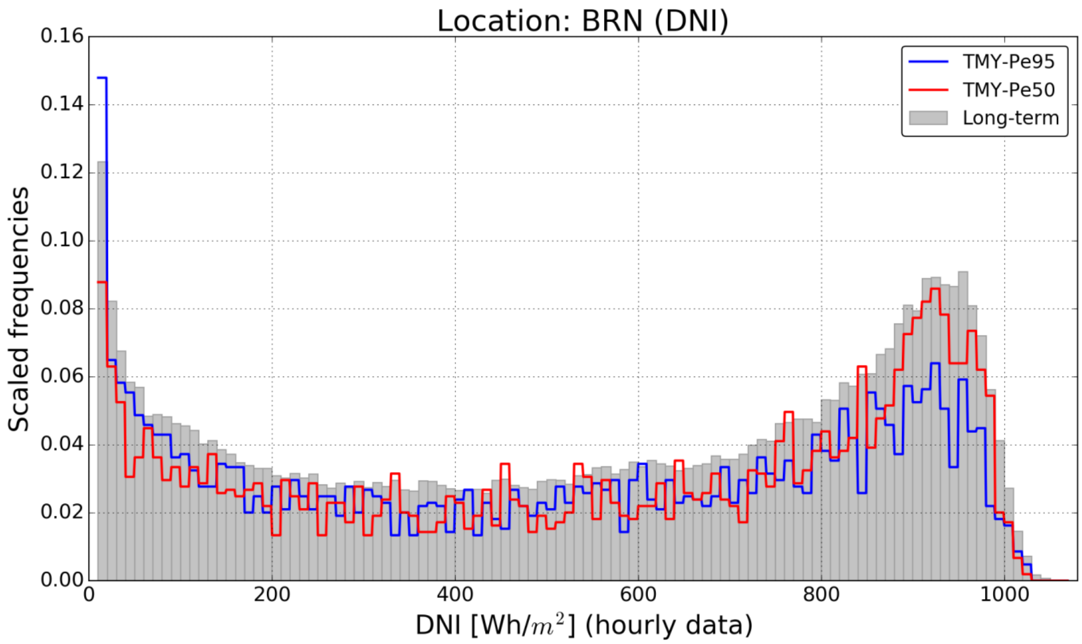

FIGURE 4. Frequency histograms of sun-up hourly values of DNI for generated TSYs at Pe95 and Pe50 and for the original long-term time series.

CONCLUSION

TMYS and TSYs are demanded by the solar energy industry because of their usefulness, mainly for the design and bankability analysis of solar projects. An issue, however, is that some temporal variability information is lost during the construction of such artificial years, with respect to the long-term time series of solar data they are based upon. Thus, the use of TMYs or TSYs is not recommended by experts for the precise design of solar energy conversion systems. Nevertheless, TMYs and TSYs have become a standard source of data for typifying the expected energy production at CSP or PV plants. Due to the lack of scientific consensus on how to define a method for generating TSYs, several initiatives have surfaced to propose methods to generate "standard" TSYs or TMYs. The design characteristic of representativeness and the constraint of being conformed by data of the historical long-term time series are the commonly elements used by the proposed methods. Whereas well-established methods exist to develop synthetic years for average or median conditions, the current challenge is to represent more extreme solar resource situations, such as what a financial institution would consider a worst-case scenario for interest repayment, which is indicated by the widely-used concept of probability of exceedance.

In this work, a novel procedure for generating TSYs of solar irradiance -both DNI and GHI components- for any probability scenario is presented. The method, referred to as the EVA method, is based on statistical criteria and it has an analytical definition. The objective of the method is the determination of the energy monthly values of each calendar month whose annual sum is equal to the annual target value of probability of exceedance Pexx. These 12 months are called monthly expected values (MEV). The TSY is comprised by the available 12 calendar months whose absolute difference respect to the corresponding MEV is minimal. The EVA method makes use of a composition of weights that accounts for: i) the variability of the seasonal adjusted monthly distributions, ii) the individual monthly energy contribution of each calendar month respect to the total annual energy.

The results show that the method provides reasonable results, with low differences between the annual target value for a certain probability of exceedance and the annual value of the irradiance generated by the TSY. It also preserves the long-term statistics when the constructed TSY of a determined probability of exceedance is set to the original higher time resolution of its constituents months (1 hour). It has also other valuable properties such as its statistical base and analytical definition, flexibility and facility to be implemented in a software code.

Deeper research should be carried out to extend the pool of sites where the evaluation presented here is possible. Further work should also analyze other essential aspects of bankability, like uncertainty. In particular, it would be important to establish how to take the uncertainty in the EVA method into account with respect to the total uncertainty in the generation of a TSY. Such uncertainties include those related to the generation of the data in the original long-term time series, along with those due to the representativeness of the available dataset (usually of limited duration) with respect to actual long-term conditions. It would also be useful to compare results obtained from long-term time series of measured data to those derived from other sources, such as satellite-based modeled data. Finally, it would be important to carry out a study in which the solar plant's energy production would be analyzed in direct connection with the solar irradiance data. Such study would examine the relations between the solar resource and the expected vs. actual energy production, as affected by the use or not of a TSY at the design stage.

ACKNOWLEDGMENTS

This work has been possible thanks to the availability of a long period of historical time series of the variables analyzed. The authors would like to give special thanks to the past and current staff of the UO SRML network that has been making the great and important effort of maintaining and operating the stations along all these years. Authors would also like to thank to Dr Christian Gueymard for the valuable contributions that he has kindly provided with the revision of this work.

REFERENCES

1. A. Habte, A. Lopez, M. Sengupta, and S. Wilcox, "Temporal and Spatial Comparison of Gridded TMY, TDY, and TGY Data Sets", National Renewable Energy Laboratory, Golden, CO. (No. NREL/TP-5D00-60886) (2014).
2. T. Stoffel, D. Renné, D. Myers, S. Wilcox, M. Sengupta, R. George, C. Turchi, "Best practices handbook for the collection and use of solar resource data", National Renewable Energy Laboratory, Golden, CO. (No. NREL/TP-550-47465) (2010).
3. M. Sengupta, A. Habte, S. Kurtz, A.P. Dobos, S. Wilbert, E. Lorenz, T. Stoffel, D. Renné, C. Gueymard, D. Myers, S. Wilcox, P. Blanc, R. Pérez, "Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications", National Renewable Energy Laboratory, Golden, CO. (No. NREL/TP-5D00-63112) (2015).
4. T. Cebecauer and M. Suri. Energy Procedia **69**, 1958 – 1969 (2015).
5. C.M.F. Peruchena, L. Ramírez, M.A. Silva-Pérez, V. Lara, D. Bermejo, M. Gastón, S. Moreno-Tejera, J. Pulgar, J. Liria, S. Macías, R. Gonzalez, A. Bernardos, N. Castillo, B. Bolinaga, R.X.. Valenzuela, L.F. Zarzalejo, Sol. Energy **123**, 29–39 (2016).
6. M.L. Delignette-Muller, C. Dutang, R. Pouillot, J. B. Denis, Journal of Statistical Software, **64**(4), 1-34 (2015).
7. M. Iqbal. *An Introduction to Solar Radiation*. (Academic Press, New York, 1983), pp. 188-195.
8. C.N. Long, Y. Shi, The Open Atmospheric Science Journal, **2**, 23-37 (2008).
9. C. A. Gueymard, Solar Energy, **86**, 3544-3553 (2012).