

## Article

# A Comprehensive Tool for Scenario Generation of Solar Irradiance Profiles

Amedeo Buonanno <sup>1</sup>, Martina Caliano <sup>1</sup>, Marialaura Di Somma <sup>1,\*</sup>, Giorgio Graditi <sup>2</sup> and Maria Valenti <sup>1</sup><sup>1</sup> Department of Energy Technologies and Renewable Energy Sources, ENEA, 80055 Portici, NA, Italy<sup>2</sup> Department of Energy Technologies and Renewable Energy Sources, ENEA, 00123 Rome, Italy

\* Correspondence: marialaura.disomma@enea.it

**Abstract:** Despite their positive effects on the decarbonization of energy systems, renewable energy sources can dramatically influence the short-term scheduling of distributed energy resources (DER) in smart grids due to their intermittent and non-programmable nature. Renewables' uncertainties need to be properly considered in order to avoid DER operation strategies that may deviate from the optimal ones. This paper presents a comprehensive tool for the scenario generation of solar irradiance profiles by using historical data for a specific location. The tool is particularly useful for creating scenarios in the context of the stochastic operation optimization of DER systems. Making use of the Roulette Wheel mechanism for generating an initial set of scenarios, the tool applies a reduction process based on the Fast-Forward method, which allows the preservation of the most representative ones while reducing the computational efforts in the next potential stochastic optimization phase. From the application of the proposed tool to a numerical case study, it emerged that plausible scenarios are generated for solar irradiance profiles to be used as input for DER stochastic optimization purposes. Moreover, the high flexibility of the proposed tool allows the estimation of the behavior of the stochastic operation optimization of DER in the presence of more fluctuating but plausible solar irradiance patterns. A sensitivity analysis has also been carried out to evaluate the impact of key parameters, such as the number of regions, a metric, and a specific parameter used for the outlier removal process on the generated solar irradiance profiles, by showing their influence on their smoothness and variability. The results of this analysis are found to be particularly suitable to guide users in the definition of scenarios with specific characteristics.

**Keywords:** scenario generation; scenario reduction; solar irradiance profiles; smart grid

**Citation:** Buonanno, A.; Caliano, M.; Di Somma, M.; Graditi, G.; Valenti, M. A Comprehensive Tool for Scenario Generation of Solar Irradiance Profiles. *Energies* **2022**, *15*, 8830. <https://doi.org/10.3390/en15238830>

Academic Editor: Adalgisa Sinicropi

Received: 31 October 2022

Accepted: 20 November 2022

Published: 23 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The sustainability objectives set by the European Green Deal require the increasing use of generation systems based on renewable energy sources (RES), as well as the use of electricity as the main energy vector. The pathway to reducing carbon emissions by 2030 will require efforts across society and sectors. With the European Green Deal as the main plan to implement for promoting this change, the European Union (EU) finalized its master program to fight carbon emissions, namely, the "Fit for 55" package [1]. Released in two batches in July and December 2021, the package drafts of EU climate and energy legislation underpin the bloc's political pledge to cut emissions by at least 55% by 2030 compared with 1990 levels. This target is more ambitious than the previous one of a 40% reduction for 2030 and is the key to achieving carbon neutrality in the EU by 2050. In view of this scenario, solar energy will become one of the key players in the electricity generation sector, thanks to its viability in combating global warming and its effectiveness in reducing pollution caused by fossil-fuel-based generation and diversifying the energy mix to ensure energy security. In particular, the installed capacity of solar photovoltaics (PV) has grown rapidly over the past decade due to the great improvements in PV technology performance, reductions in cost, and the development of efficient business models that have fostered new

investments in this technology. This trend is expected to continue in the future, affecting not only large-scale centralized solar farms but, above all, small-/medium-scale PV at the distribution level, where the number of PV applications owned by residential and industrial prosumers (power producers and consumers) will also expand, driven by environmental policies and economic incentives.

From a range of studies, solar PV is expected to contribute 36% to 69% of European electricity consumption by 2050 [2], and its role is predominant in Energy Communities [3]. Due to the variability and intermittent nature of the solar PV output, such a high share of solar PV will impact the overall system costs due to the increase in operating costs and the infrastructure needed. This problem is aggravated by the inaccuracy of the methodologies in modeling renewables' uncertainties, which are related to the uncertainty of weather conditions for RES [4], which represent a key factor to be properly handled in the smart grid environment. In fact, they may influence how distributed energy resources (DER) are scheduled in the short term to provide the available flexibility for system balancing at all times [5–8]. If such uncertainties are not identified and handled properly in the operation scheduling of DER, their operation strategies may deviate from the optimal ones by causing a number of issues, such as an increase in operational costs or system stability and security. Modeling RES uncertainties in the stochastic operation optimization of DER is thus extremely important [9–12].

Several works in the literature deal with different sources of uncertainties, such as renewables, electricity consumption, electric vehicles, etc. [13–17]. The objective of uncertainty-modeling methods is to evaluate the impact of uncertain input parameters on system output parameters. These methods can be subdivided into several groups, as suggested in [18,19]:

- Probabilistic: the probability density functions (PDFs) of the input parameters are used;
- Possibilistic (fuzzy): the uncertainty of the input parameters is modeled with a membership function (MF);
- Hybrid probabilistic and possibilistic: both probabilistic and possibilistic approaches are used;
- Based on Information Gap Decision Theory (IGDT): it measures the deviation of the estimation error;
- Robust optimization: the uncertainty of the input parameters is described using uncertainty sets;
- Interval analysis: the uncertain inputs can assume values in a known interval (similar to the probabilistic approach with uniform PDFs).

The approaches to estimating solar irradiance can be grouped into linear, nonlinear, Artificial Neural Network (ANN)-based, and Fuzzy Logic (FL) techniques [20]. For linear and nonlinear models, the authors have created associations between solar irradiance and other variables, such as meteorological ones [21–23]. In ANN-based approaches [24,25], the usual inputs used are geographical coordinates, meteorological data, and information related to the current date and time. In FL approaches, the input to the estimation model is the classified sky condition. Moreover, other authors have applied statistical methods to study the hourly variation in solar irradiance data considering different climatic locations [26] or empirical models to estimate solar irradiance on a monthly basis for different locations [27,28]. In [29], the authors state that FL can yield better estimation results when the data available for estimation are ambiguous and vague.

Among the probability distributions, beta is considered one of the most effective for modeling solar irradiance [30–33] and is often employed in planning studies related to PV systems [34–36]. Other works propose the Weibull distribution for modeling solar irradiance [37].

The contribution of this paper is the presentation of a comprehensive tool to generate solar irradiance profiles using a scenario generation approach and historical data for a specific location. The methodology is general and highly replicable and can thus be applied in several contexts for generating 24 h solar irradiance scenarios, which are useful for the

stochastic operation optimization of DER. The tool has been completely implemented in Python, and it is suitable for transformation into a Web Service to generate solar irradiance profiles related to a particular geographical region and a time period of interest.

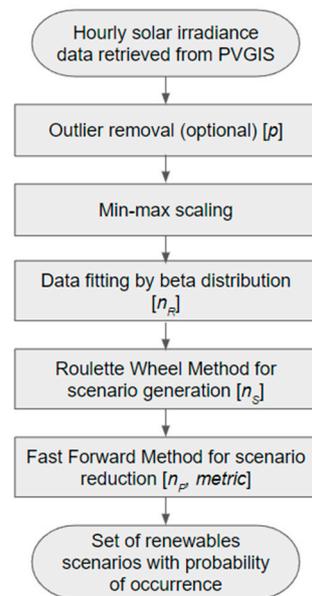
The historical hourly solar irradiance data were fitted using the beta distribution, and the Roulette Wheel mechanism [38] was used to generate an initial set of scenarios; then, a reduction process based on the Fast-Forward method [39,40] was applied in order to preserve the most representative ones while reducing the computational efforts in the next potential stochastic optimization phase. The generation and reduction phases are ruled by certain parameters, such as the number of regions, a metric, and a specific parameter used for an optional process for outlier removal that can be modified. Moreover, a sensitivity analysis was performed to evaluate the impacts of the variations in these parameters on the solar irradiance scenarios generated. The numerical results of the analysis show that these parameters have a visible effect on the smoothness and variability of the generated scenarios. Based on the current scientific literature, there are no previous works that examined these aspects, thereby highlighting the importance of this study, which could be useful as a guide for tuning the scenario generation process with the aim of obtaining scenarios with certain characteristics. In the case study, the proposed tool is found to be efficient in generating plausible scenarios for daily solar irradiance profiles with 1 h as the time-step to be used as input for DER stochastic operation optimization purposes. Moreover, the high flexibility of the proposed tool allows the estimation of the behavior of DER stochastic operation optimization in the presence of more fluctuating—but plausible—solar irradiance profiles. The current paper extends the results presented in [41] by including additional results in the case study, as well as introducing a new verification system of the plausibility of the reduced scenarios.

In the following, the dataset, data preprocessing, data fitting, and methods for scenario generation and reduction are discussed in Section 2. The results of varying certain key parameters in the scenario generation and reduction processes are presented in Section 3. In Section 4, the sensitivity analysis is discussed, along with the obtained results.

## 2. Materials and Methods

The tool proposed in this paper is based on a statistical approach used to model solar irradiance based on historical data. By using the Roulette Wheel method [38], an initial set of scenarios is first generated, and then, through a reduction process, the most representative ones are preserved.

A scheme for describing the proposed tool is shown in Figure 1. In particular, the dataset retrieved from the Photovoltaic Geographical Information System (PVGIS) [42] (Section 2.1) is preprocessed by using an optional process for outlier removal and min–max scaling, as described in Section 2.2. For each hour, a data-fitting process (Section 2.3) is performed in order to obtain a probability distribution for each hour. From each hourly probability distribution, a sample is extracted to obtain 24 randomly sampled values, the so-called 24 h solar irradiance scenario (Section 2.4). This process is performed  $n_S$  times in order to obtain  $n_S$  scenarios. In order to reduce the computational complexity of the following optimization task, the generated scenarios have to be reduced by using an approach that preserves the “information content” present in the original set of scenarios. This is performed by using the procedure described in Section 2.5.



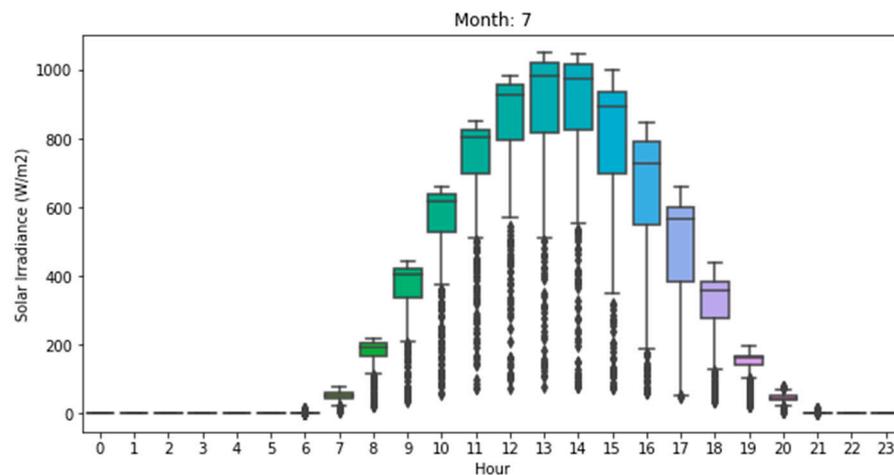
**Figure 1.** Overall description of the proposed tool. In square parentheses, the parameters that affect the considered step are reported.

### 2.1. Dataset Description

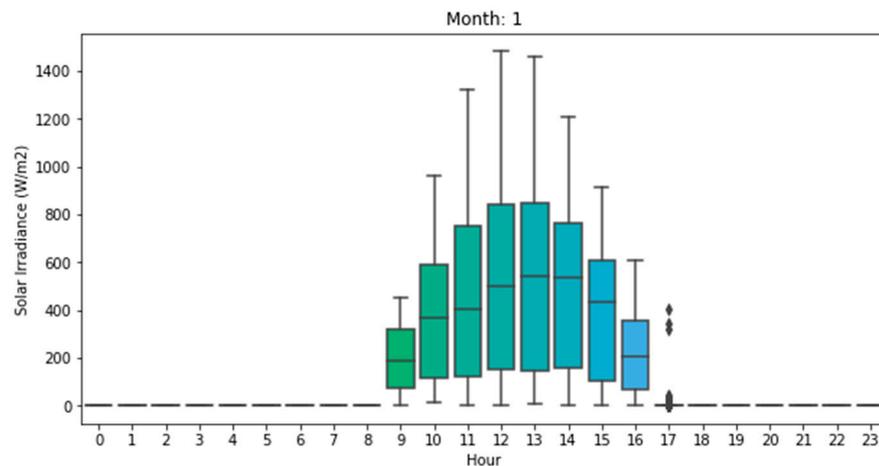
The hourly solar irradiance data from 2005 to 2016 for the city of Turin (Italy) were gathered using PVGIS [42]. In order to model both the winter and summer seasons, January and July were selected as months of interest.

The daily patterns of solar irradiance for the days in July and January from 2005 to 2016 are shown in Figures 2 and 3, respectively.

From the figures, it is possible to observe the great variability in each hour, which will be modeled by means of a probability distribution that fits the data.



**Figure 2.** Box-whisker plot of solar irradiance for the days in July from 2005 to 2016. The bold points represent values identified as outliers following Equation (1) with  $p = 1.5$ . To each hour is associated a different color of box.



**Figure 3.** Box-whisker plot of solar irradiance for the days in January from 2005 to 2016. The bold points represent values identified as outliers following Equation (1) with  $p = 1.5$ . To each hour is associated a different color of box.

### 2.2. Data Preprocessing

The hourly data for each considered month were normalized using min–max scaling in order to map the values observed in the range  $[0, 1]$ .

To reduce the variability in the observations, the outliers can be removed using the Interquartile Range (IQR) method [43], according to which the values are considered outliers—and hence removed—when they are outside of the following range:

$$[Q_1 - p \cdot IQR, Q_3 + p \cdot IQR] \tag{1}$$

where  $Q_1$  and  $Q_3$  are the 1st and 3rd quartiles (25th and 75th percentiles), respectively;  $IQR = Q_3 - Q_1$ ; and  $p$  is a value that permits the expansion or restriction of the range and hence the consideration of fewer or more values as outliers.

However, it is the user’s choice to enable (or not) the outlier removal process, as well as the value of  $p$ .

### 2.3. Data Fitting

The normalized hourly irradiance data were fitted using several probability distributions, namely, Weibull, beta, logistic, and arcsine. Among the tested distributions, beta was found to be the best fit for most hours in January and July, confirming what was observed in the relevant literature for solar irradiance data [32,33].

Beta is a continuous probability distribution with support in  $[0, 1]$ , and its PDF is defined as in [44]:

$$f_S(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)} \tag{2}$$

where  $B(a, b)$  is the beta function, formulated as:

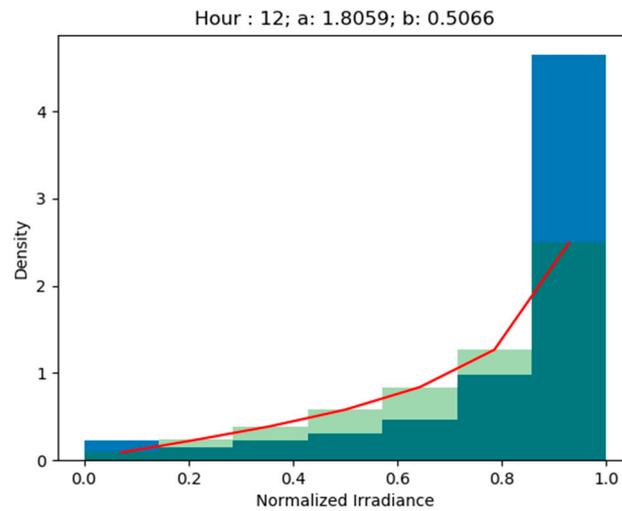
$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)} \tag{3}$$

where  $\Gamma(x)$  is the gamma function, formulated as:

$$\Gamma(x) = \int_0^\infty u^{x-1}e^{-u} du \tag{4}$$

Different values of  $a$  and  $b$  allow uniform ( $a = 1, b = 1$ ), bimodal ( $a < 1, b < 1$ ), or unimodal ( $a > 1, b > 1$ ) distributions to be obtained.  $n_R$  regions, or bins, are defined to divide the support of the distribution.

Figure 4 shows the fitting results of data related to the hour 12:00 p.m. in July using the beta distribution and dividing its support into 7 regions.



**Figure 4.** Data fitting using beta distribution with 7 regions: (blue bars) histograms of the empirical distribution; (green bars overlaid with transparency) histograms of the fitted beta distribution; (red line) the line connecting the central points of all regions.

#### 2.4. Roulette Wheel Method

The Roulette Wheel method [38] was used to extract a set of samples from the fitted probability density functions with their supports divided into  $n_R$  regions [45].

The probability of the occurrence of a particular region  $r$  ( $r \in \{1, \dots, n_R\}$ ) at time  $t$  (the hour of the day,  $t \in \{0, \dots, 23\}$ ) can be computed as the product of width  $w_{t,r}$  (width of region  $r$  at time  $t$ ) and height  $h_{t,r}$  (height of region  $r$  at time  $t$ ):  $\alpha_{t,r} = w_{t,r} \cdot h_{t,r}$ , appropriately normalized as reported in Equation (5).

$$\hat{\alpha}_{t,r} = \frac{\alpha_{t,r}}{\sum_{\rho=1}^{n_R} \alpha_{t,\rho}} \quad (5)$$

The probabilities of the occurrence for all regions of a particular hour were sorted in descending order and cumulated.

In order to sample from the beta distribution with its support divided into  $n_R$  regions, it is possible to sample a value from a uniform distribution ( $v \sim U(0, 1)$ ) and, comparing it with the cumulative probability of occurrence, select one of  $n_R$  possible regions that  $v$  belongs to (this method is also known as the Inverse Transform [46]).

The central value of the selected region is the value sampled from the beta probability distribution with support divided into  $n_R$  regions. The procedure described above was performed for each hour, obtaining 24 sampled values that compose the scenario. To be precise, not all 24 values were sampled in this way because, in some hours, the observed solar irradiance was constantly zero, and this behavior was maintained in the scenario generation process.

The binary variable  $W_{k,t,r}$  is used to contain the information on whether region  $r$  is selected for a scenario  $s^k$  at hour  $t$  ( $W_{k,t,r} = 1$ ) or not ( $W_{k,t,r} = 0$ ).

The main assumption of the proposed model is that, at a specific hour, the solar irradiance is independent of the values observed in the previous hours. With this assumption, the probability of the occurrence of scenario  $s^k$ ,  $\pi_k$ , is the product of the probabilities of the regions that compose it [45]:

$$\pi_k = \frac{\prod_{t=0}^{23} \sum_{r=1}^{n_R} \{W_{k,t,r} \cdot \hat{\alpha}_{t,r}\}}{\sum_{k=1}^{n_S} \prod_{t=0}^{23} \sum_{r=1}^{n_R} \{W_{k,t,r} \cdot \hat{\alpha}_{t,r}\}}, \quad k = 1, \dots, n_S \quad (6)$$

The independence assumption can be relaxed, considering that the solar irradiance assumed at a particular hour depends on the solar irradiance values of previous hours. The methodology described here is, however, still valid, and this extension is reported elsewhere.

### 2.5. Scenarios' Reduction Process

From an initial set of  $n_S$  scenarios with their probabilities  $\pi_k$  (with  $k = 1, \dots, n_S$ ), it is necessary to obtain a reduced set of  $n_P$  scenarios to use in the successive potential stochastic optimization phase. The considered reduction method is the Fast-Forward method [39,40]. The algorithm creates a subset of scenarios with the minimum Kantorovich distance from the initial set. After computing the distance (with respect to a *metric* defined as the mean absolute distance, Euclidean distance, etc.) between all pairs of scenarios, the scenario with the minimum weighted distance (the weights are the probabilities of the occurrence of each scenario) from all of the other ones is selected. The probability of the occurrence of all removed scenarios is absorbed by the preserved scenario that is nearest to them.

The main steps of the Fast-Forward algorithm are described in the Algorithm 1 box [39].

---

#### Algorithm 1. Fast-Forward

---

##### Step 1

1. For each pair of scenarios ( $s^k$  and  $s^u$ ), the distance is computed by using the metric  $c_T$ . The generic element  $C_{k,u}$  of matrix  $C$  in step 1 is:

2. 
$$C_{k,u}^{[1]} = c_T(s^k, s^u), k, u = 1, \dots, n_S \quad (7)$$

3. The metric usually used is the  $\ell_q$ -Norm of  $\mathbb{R}^T$ , which can be defined as:

4. 
$$c_T(s^k, s^u) = \left( \sum_{t=1}^T |s_t^k - s_t^u|^q \right)^{\frac{1}{q}} \quad (8)$$

5. Each scenario  $s^u$  is associated with the weighted distance to any other scenario  $s^k$ , where the weights are the probabilities of occurrence  $\pi_k$ :

6. 
$$z_u^{[1]} = \sum_{\substack{k=1 \\ k \neq u}}^{n_S} \pi_k C_{k,u}^{[1]}, u = 1, \dots, n_S \quad (9)$$

For example, the  $z$  values for scenarios  $s^1$  and  $s^2$  are the following:

7. 
$$z_1^{[1]} = \pi_2 C_{21}^{[1]} + \pi_3 C_{31}^{[1]} + \pi_4 C_{41}^{[1]} + \dots$$

$$z_2^{[1]} = \pi_1 C_{12}^{[1]} + \pi_3 C_{32}^{[1]} + \pi_4 C_{42}^{[1]} + \dots$$

8. Among the results, the index of the scenario with the minimum value of  $z$  is selected ( $u_1$ ):

9. 
$$u_1 \in \operatorname{argmin}_{u \in \{1, \dots, n_S\}} z_u^{[1]} \quad (10)$$

---

**Algorithm 1.** *cont.*

10. Then,  $s^{u_1}$  is preserved (operatively,  $u_1$  is removed from the indexes of scenarios to delete in step 1,  $J^{[1]}$ ):

$$11. \quad J^{[1]} = \{1, \dots, n_S\} \setminus \{u_1\} \quad (11)$$

**Step i**

Using the information from previous steps, the distance matrix is updated using Equation (12), new values of  $z$  are computed using Equation (13), and a new scenario is selected to be preserved ( $s^{u_i}$ ) using Equations (14) and (15):

$$13. \quad C_{ku}^{[i]} = \min \{C_{ku}^{[i-1]}, C_{ku_{i-1}}^{[i-1]}\}, k, u \in J^{[i-1]} \quad (12)$$

$$14. \quad z_u^{[i]} = \sum_{k \in J^{[i-1]} \setminus \{u\}} \pi_k C_{ku}^{[i]}, u \in J^{[i-1]} \quad (13)$$

$$15. \quad u_i \in \operatorname{argmin}_{u \in J^{[i-1]}} z_u^{[i]} \quad (14)$$

$$16. \quad J^{[i]} = J^{[i-1]} \setminus \{u_i\} \quad (15)$$

**Step  $n_P + 1$** 

In the final step, the list of scenarios to remove  $J = J^{[n_P]}$  is completed. Each scenario to be removed will be linked to a preserved scenario that will “substitute” it. In fact,  $j(i)$  is the index of the preserved scenario nearest to the removed scenario  $s^i$ :

$$18. \quad j(i) \in \operatorname{argmin}_{j \notin J} c_T(s^i, s^j), \forall i \in J \quad (16)$$

19. The set of indexes of the removed scenarios that have  $s^j$  as the nearest preserved scenario can be defined as follows:

$$20. \quad J(j) = \{i \in J : j = j(i)\} \quad (17)$$

21. Using the *optimal redistribution rule* [40], the probability of the occurrence  $\pi_j$  of the preserved scenario  $s^j$  is computed:

$$22. \quad \pi_j = \pi_j + \sum_{i \in J(j)} \pi_i \quad (18)$$

23. The probabilities of the occurrence of the removed scenarios nearest to  $s^j$  are added to the initial value of  $\pi_j$ .

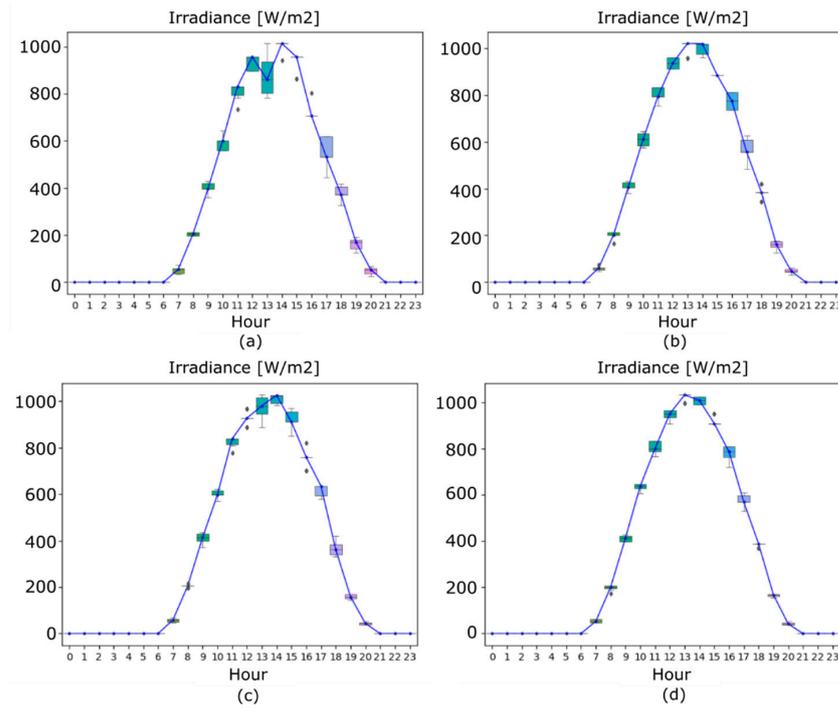
**3. Numerical Results**

In this section, the numerical results of the scenario generation process carried out considering different values of  $n_R$ , *metric*, and  $p$  are presented.

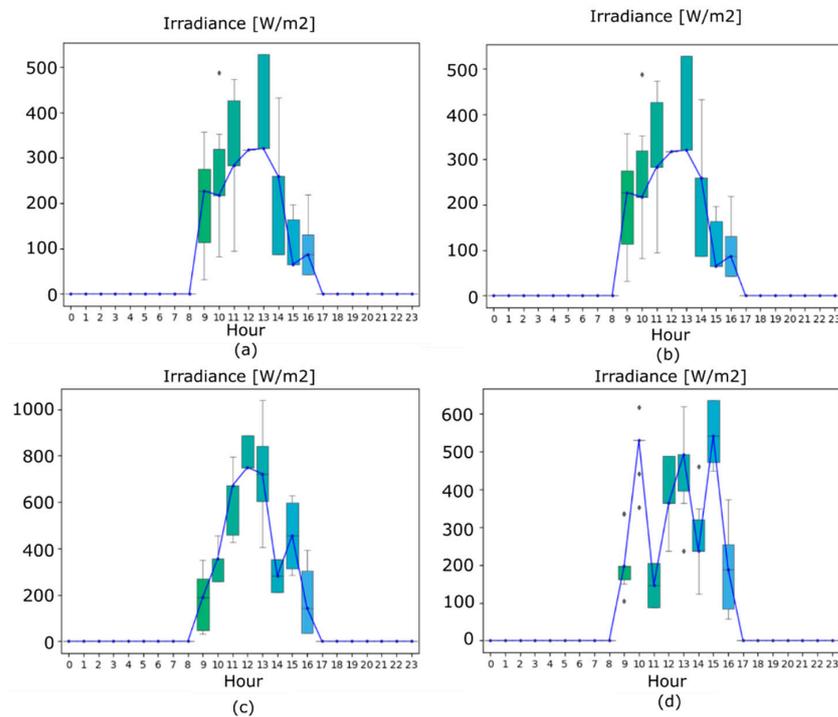
Due to the randomness of sampling from uniform random variables in the Roulette Wheel method, two successive executions of the scenario generation process with the same parameters could return different scenarios unless fixing the *seed* for sampling from the uniform distribution. In order to compare the several generated scenarios with different parameters, the same *seed* was used for all trials related to the same month.

**3.1. Outlier Removal**

Figures 5 and 6 show the results obtained by using  $n_S = 1000$ ,  $n_P = 10$ ,  $n_R = 7$ , and *metric* =  $\ell_2$ -Norm and by varying  $p$ .



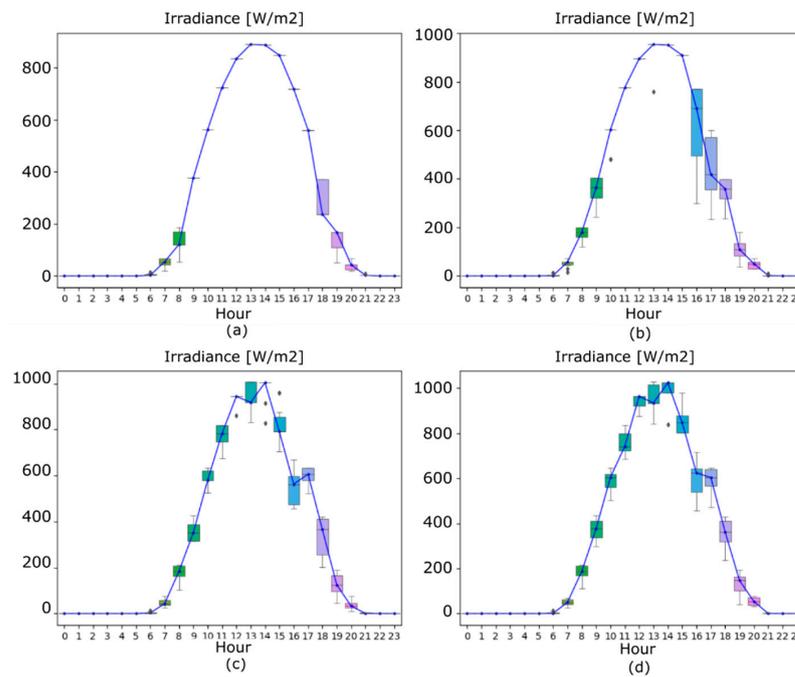
**Figure 5.** Median scenarios (blue) with box-whisker plot for scenarios generated for July using  $n_S = 1000$ ,  $n_P = 10$ ,  $n_R = 7$ , and  $metric = \ell_2$ -Norm and varying  $p$ : 1.5 (a), 1 (b), 0.5 (c), and 0.15 (d).



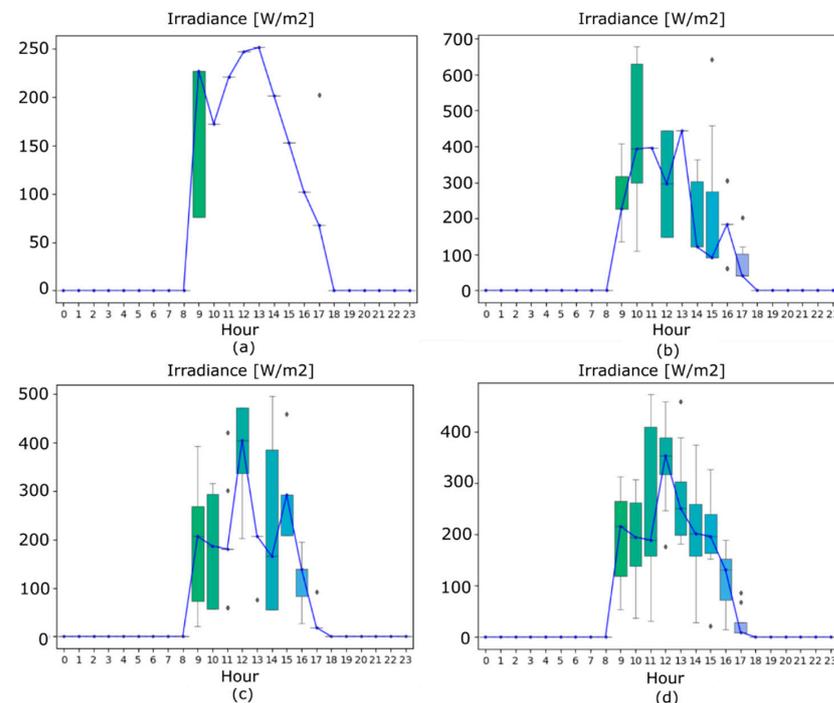
**Figure 6.** Median scenarios (blue) with box-whisker plot for scenarios generated for January using  $n_S = 1000$ ,  $n_P = 10$ ,  $n_R = 7$ , and  $metric = \ell_2$ -Norm and varying  $p$ : 1.5 (a), 1 (b), 0.5 (c), and 0.15 (d).

### 3.2. Number of Regions

Figures 7 and 8 show the results obtained by considering  $n_S = 1000$ ,  $n_P = 10$ , and  $metric = \ell_2$ -Norm without outlier removal and by varying  $n_R$ .



**Figure 7.** Median scenarios (blue) with box-whisker plot for scenarios generated for July using  $n_S = 1000$ ,  $n_P = 10$ , and  $metric = \ell_2$ -Norm without outlier removal and varying the number of regions ( $n_R$ ): 3 (a), 5 (b), 11 (c), and 21 (d).



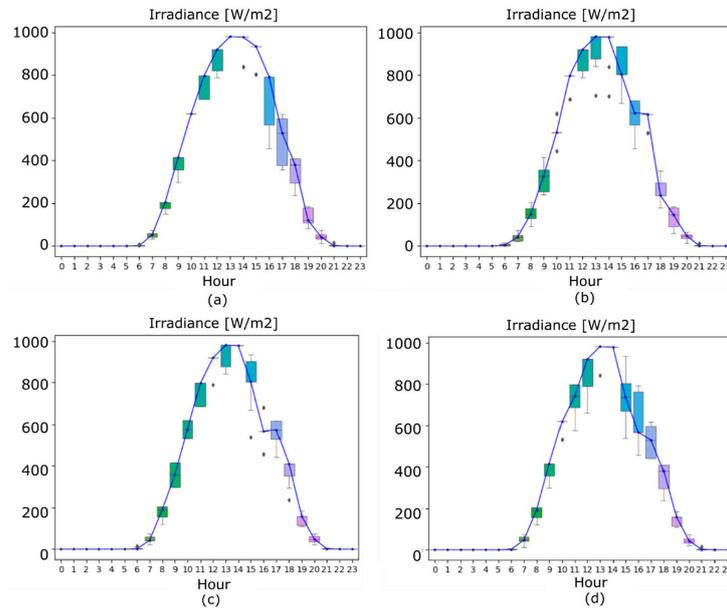
**Figure 8.** Median scenarios (blue) with box-whisker plot for scenarios generated for January using  $n_S = 1000$ ,  $n_P = 10$ , and  $metric = \ell_2$ -Norm without outlier removal and varying the number of regions ( $n_R$ ): 3 (a), 5 (b), 11 (c), and 21 (d).

### 3.3. Metric

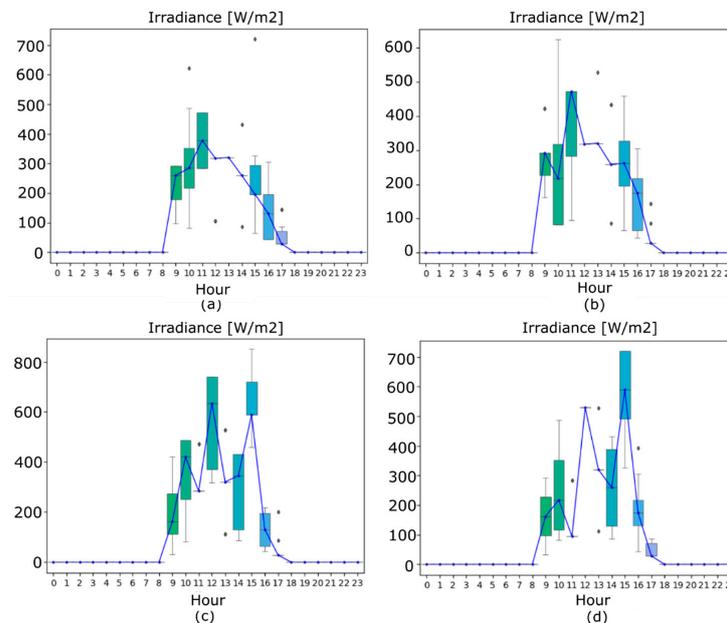
Figures 9 and 10 show the results obtained by using  $n_S = 1000$ ,  $n_P = 10$ , and  $n_R = 7$  without outlier removal and varying the metric used to compare two scenarios. For that,

several  $\ell_q$ -Norms were tested (Equation (8)):  $\ell_1$ -Norm ( $q = 1$ ),  $\ell_2$ -Norm ( $q = 2$ ),  $\ell_4$ -Norm ( $q = 4$ ), and  $\ell_\infty$ -Norm (Equation (19)).

$$c_T(s^k, s^u) = \max_t |s_t^k - s_t^u| \tag{19}$$



**Figure 9.** Median scenarios (blue) with box-whisker plot for scenarios generated for July using  $n_S = 1000$ ,  $n_P = 10$ , and  $n_R = 7$  without outlier removal and varying the type of metric:  $\ell_1$ -Norm (a),  $\ell_2$ -Norm (b),  $\ell_4$ -Norm (c), and  $\ell_\infty$ -Norm (d).

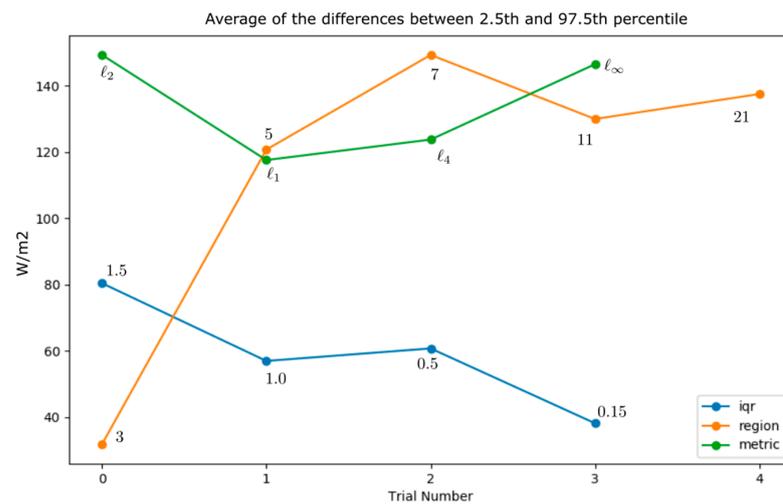


**Figure 10.** Median scenarios (blue) with box-whisker plot for scenarios generated for January using  $n_S = 1000$ ,  $n_P = 10$ , and  $n_R = 7$  without outlier removal and varying the type of metric:  $\ell_1$ -Norm (a),  $\ell_2$ -Norm (b),  $\ell_4$ -Norm (c), and  $\ell_\infty$ -Norm (d).

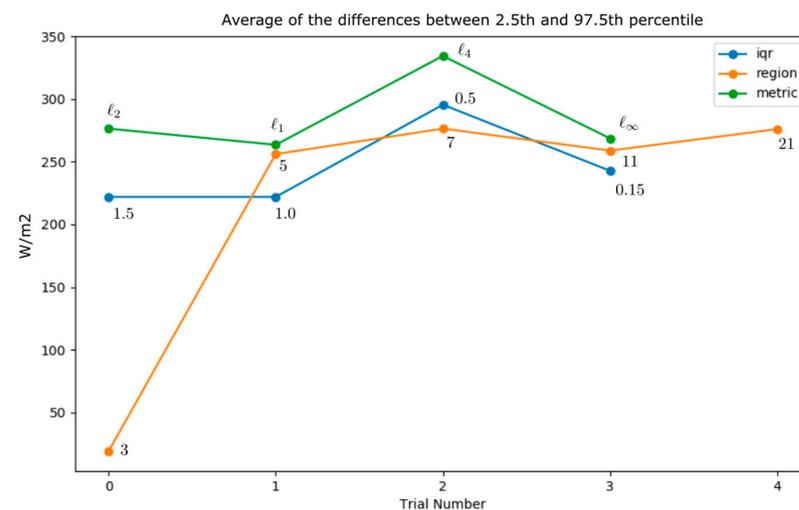
### 4. Discussion

In order to show the impact of the parameters  $p$ ,  $n_R$ , and  $metric$  on the preserved scenarios (in terms of smoothness and variability in the same hour), a sensitivity analysis was performed.

For the estimation of the variability in the preserved scenarios, the average (for all hours excluding those in which irradiance is always zero) of the difference between the 97.5th and 2.5th percentiles of the preserved scenarios was considered. The trends of the average of the difference between the 97.5th and 2.5th percentiles for the different trials described in Sections 3.1–3.3 for July (summer) and January (winter) are shown in Figures 11 and 12, respectively.



**Figure 11.** Trend of average of the difference between 97.5th and 2.5th percentiles for preserved scenarios for the summer: (blue)  $n_R = 7$  and  $metric = l_2$ -Norm and varying  $p$ : 1.5, 1, 0.5, and 0.15; (green)  $n_R = 7$ , without outlier removal and varying  $metric$ :  $l_2$ -Norm,  $l_1$ -Norm,  $l_4$ -Norm, and  $l_\infty$ -Norm; (orange)  $metric = l_2$ -Norm, without outlier removal and varying  $n_R$ : 3, 5, 7, 11, and 21. The text in the graph indicates the values assumed by the varied parameters.



**Figure 12.** Trend of average of the difference between 97.5th and 2.5th percentiles for preserved scenarios for the winter: (blue)  $n_R = 7$  and  $metric = l_2$ -Norm and varying  $p$ : 1.5, 1, 0.5, and 0.15; (green)  $n_R = 7$ , without outlier removal and varying  $metric$ :  $l_2$ -Norm,  $l_1$ -Norm,  $l_4$ -Norm, and  $l_\infty$ -Norm; (orange)  $metric = l_2$ -Norm, without outlier removal and varying  $n_R$ : 3, 5, 7, 11, and 21. The text in the graph indicates the values assumed by the varied parameters.

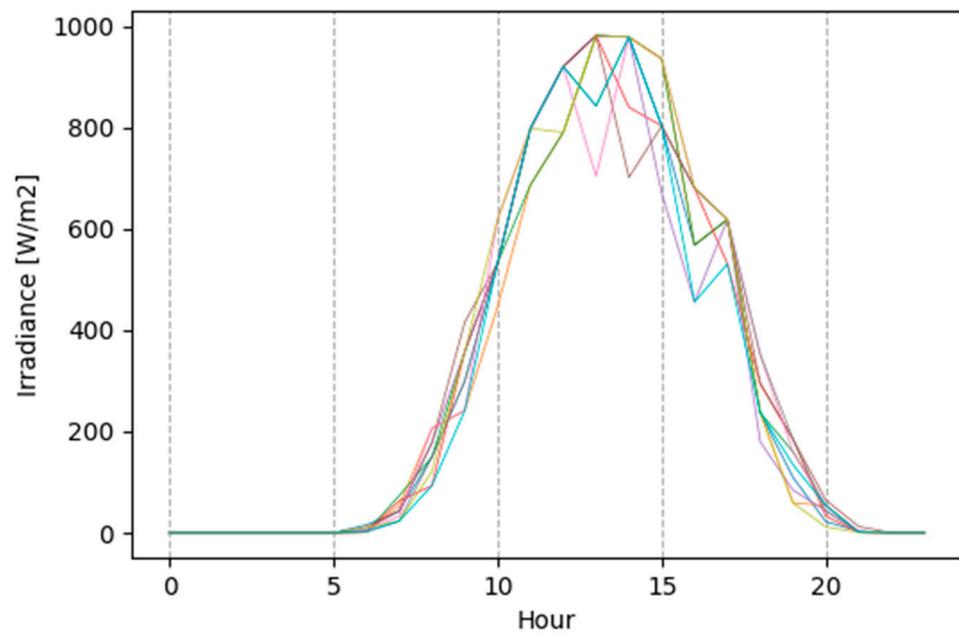
Regarding the outlier removal process, when reducing  $p$ , more samples for each hour are considered outliers (and hence removed). For the solar irradiance distributions with small hourly IQRs and many values outside IQRs, as in July, where the average IQR is about  $111 \text{ W/m}^2$  (Figure 2), removing outliers and reducing the value of  $p$  lead to generated scenarios with small variability (Figure 5 and blue line in Figure 11). For the solar irradiance distributions where hourly IQRs are high and few observations are outside IQRs, as in January, where the average IQR is about  $460 \text{ W/m}^2$  (Figure 3), the outlier removal process does not have, in general, a high impact (Figure 6 and the blue line in Figure 12).

Regarding the impact of the number of regions ( $n_R$ ), the variability is increased for both seasons when  $n_R$  is increased from 3 to 7 (Figures 7 and 8 and orange lines in Figures 11 and 12). With a value of  $n_R$  that is higher than 7, the variability is quite stable. This can be explained by the fact that, even though the support of the beta distribution has been divided using more regions, more of them have a low probability of occurrence and, hence, do not impact the resulting variability.

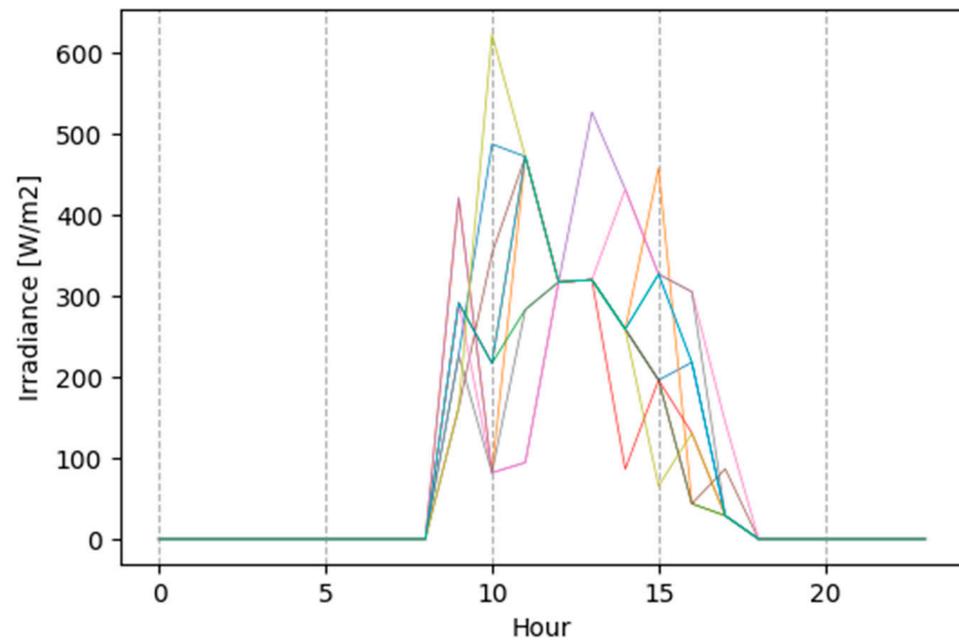
Regarding the impact of several metrics for comparing two scenarios during the reduction process (Figures 9 and 10 and green lines in Figures 11 and 12), it is possible to observe that very different results are obtained from different metrics in the two considered seasons (e.g.,  $\ell_2$ -Norm produces scenarios with high variability in summer and low variability in winter,  $\ell_4$ -Norm produces scenarios with lower variability than  $\ell_2$ -Norm in summer and higher variability than  $\ell_2$ -Norm in winter, etc.), and the variability obtained with  $\ell_\infty$ -Norm is close to that obtained with  $\ell_2$ -Norm.

The user of the method can select the best combination of the presented parameters in order to obtain the best trade-off between the variability among preserved scenarios and their plausibility. To verify the plausibility of the preserved scenarios, the user can plot them in the box-whisker plot obtained from the real solar irradiance data (Figures 1 and 2) and verify that the generated values of hourly solar irradiance for the particular scenario do not deviate too much from the boxes and whiskers. (In the box-whisker plot, the box describes the range between  $Q_1$  and  $Q_3$ , the upper whisker is on  $Q_3 + 1.5 \cdot IQR$ , and the lower whisker is on  $Q_1 - 1.5 \cdot IQR$ . The values outside the range defined by the whiskers can be considered outliers). In this case, this means that the generated values of hourly solar irradiance for that scenario can be considered inliers with respect to the observed data.

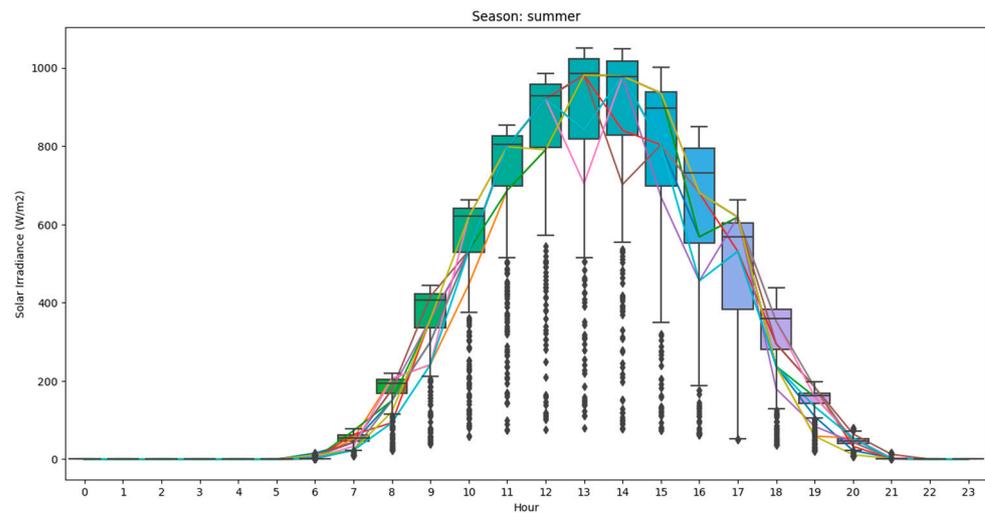
The scenarios obtained with  $n_S = 1000$ ,  $n_P = 10$ , and  $n_R = 7$ ,  $\ell_2$ -Norm metric and without the outlier removal step are shown in Figure 13 for summer and Figure 14 for winter. Figures 15 and 16 show the preserved scenarios plotted over the box-whisker plot of the observed solar irradiance. In these images, it is possible to see that the preserved scenarios are almost completely contained in the boxes (representing the IQR of the observed solar irradiance), and almost all values that are outside the boxes are confined to the variability range observed in the real data (the range between the two whiskers) and hence can be considered plausible.



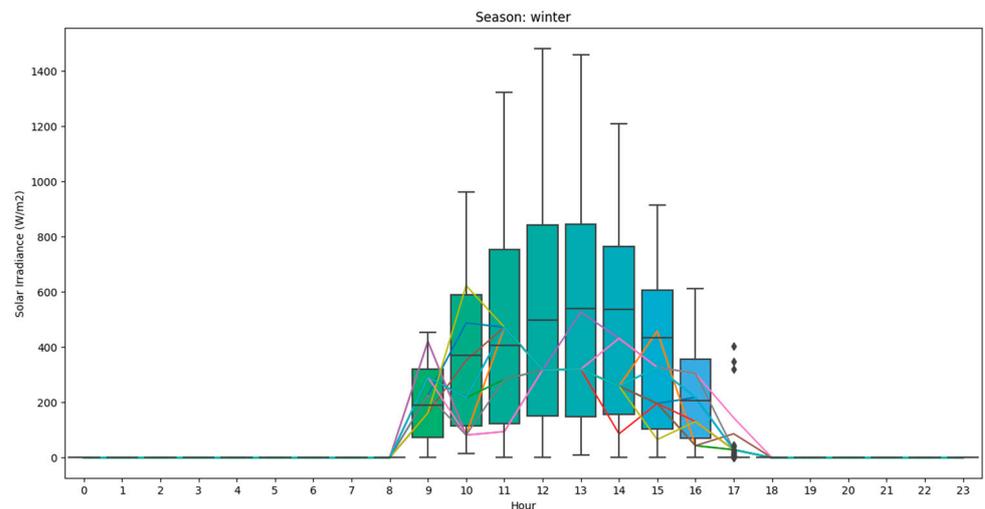
**Figure 13.** Ten generated scenarios reduced from one thousand initial scenarios using Fast-Forward algorithm with 7 regions,  $\ell_2$ -Norm metric, and without outlier removal. The historical data are related to July for the city of Turin (Italy) from 2005 to 2016. Different line colors represent different scenarios.



**Figure 14.** Ten generated scenarios reduced from one thousand initial scenarios using Fast-Forward algorithm with 7 regions,  $\ell_2$ -Norm metric, and without outlier removal. The historical data are related to January for the city of Turin (Italy) from 2005 to 2016. Different line colors represent different scenarios.



**Figure 15.** Ten preserved scenarios plotted over the box-whisker plot of the observed solar irradiance for the days in July from 2005 to 2016. Different line colors represent different scenarios.



**Figure 16.** Ten preserved scenarios plotted over the box-whisker plot of the observed solar irradiance for the days in January from 2005 to 2016. Different line colors represent different scenarios.

## 5. Conclusions

In this work, a comprehensive tool to generate solar irradiance profiles is presented. The proposed approach is based on a scenario generation process aimed at generating 24 h solar irradiance scenarios using the historical data of solar irradiance for a specific location.

In the case study, the proposed method was applied to generate a set of daily solar irradiance scenarios for the months of January and July for the city of Turin (Italy). The Roulette Wheel mechanism was used to generate the initial set of scenarios, and the Fast-Forward method for the reduction process was applied to preserve the most representative scenarios and reduce the computational efforts associated with the potential stochastic operation optimization phase. The results demonstrate the flexibility of the method in generating scenarios for solar irradiance and in assessing their plausibility. These characteristics make the proposed approach an effective tool to be used for the stochastic operation optimization of DER.

Moreover, the results of the sensitivity analysis show the influence of the variation in the key parameters on the results in terms of increasing the variability and/or the smoothness of the generated scenarios, which could be very effective in estimating the

behavior of the stochastic operation optimization of DER in the presence of more fluctuating but plausible solar irradiance patterns.

Given the generality of the proposed method, it can be easily adapted to model solar irradiance profiles for different locations and use cases, and hence, it can serve as a guide to users for the definition of scenarios with specific characteristics. Moreover, the proposed pipeline can be implemented as a Web Service queryable by users in order to generate solar irradiance scenarios with their probability of occurrence, which is fundamental for the stochastic optimization of DER.

**Author Contributions:** Conceptualization, A.B., M.C. and M.D.S.; methodology, A.B.; software, A.B.; data curation, A.B.; writing, review and editing, A.B., M.C., M.D.S., G.G. and M.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded under Project 1.7 “Tecnologie per la penetrazione efficiente del vettore elettrico negli usi finali” within the Italian Research Program ENEA-MISE 2022–2024 “Ricerca di Sistema Elettrico”.

**Data Availability Statement:** The data used for this study were gathered using PVGIS as described in the Section 2.1.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Networks
DER	Distributed energy resources
FL	Fuzzy Logic
IGDT	Information Gap Decision Theory
IQR	Interquartile Range
MF	Membership function
PDF	Probability density function
PV	Photovoltaic
PVGIS	Photovoltaic Geographical Information System
RES	Renewable energy sources

## Nomenclature

$\alpha_{t,r}$	Probability of occurrence of a particular region $r$ at time $t$
$\hat{\alpha}_{t,r}$	Normalized probability of occurrence of a particular region $r$ at time $t$
$\pi_k$	Probability of occurrence of scenario $s^k$
$\Gamma(x)$	Gamma function
$c_T$	Metric used to compute the distance between two scenarios
$h_{t,r}$	Height of region $r$ at time $t$
$n_p$	Number of preserved scenarios
$n_R$	Number of regions (bins) used to divide the support of the distribution
$n_S$	Number of generated scenarios
$p$	Parameter used to define outliers
$r$	Number of considered regions ( $r \in \{1, \dots, n_R\}$ )
$s^k$	$k$ -th scenario (signal containing 24 irradiance values)
$t$	Hour of the day ( $t \in \{0, \dots, 23\}$ )
$w_{t,r}$	Width of region $r$ at time $t$
$z_u^{[m]}$	Weighted distance of scenario $s^u$ from all other scenarios in step $m$
$B(a,b)$	Beta function with parameters $a$ and $b$
$C$	Matrix containing the distances between all pairs of scenarios
$C_{k,u}^{[m]}$	$(k,u)$ th entry of matrix $C$ , representing the distance between scenarios $s^k$ and $s^u$ in step $m$
$J(j)$	Set of indexes of the removed scenarios that have $s^j$ as the nearest preserved scenario
$J^{[m]}$	List of indexes of deleted scenarios in step $m$
$Q_1$	1st quartile, 25th percentile of observed values
$Q_3$	3rd quartile, 75th percentile of observed values
$W_{k,t,r}$	Binary variable that describes whether region $r$ of scenario $s^k$ is selected at time $t$

## References

1. 'Fit1. Fit for 55': Delivering the EU's 2030 Climate Target on the Way to Climate neutralitycom(2021) 550 Final, Brussels, 14.7.2021. Available online: [https://ec.europa.eu/info/sites/default/files/chapeau\\_communication.pdf](https://ec.europa.eu/info/sites/default/files/chapeau_communication.pdf) (accessed on 30 October 2022).
2. Breyer, C.; Khalili, S.; Bogdanov, D. Solar photovoltaic capacity demand for a sustainable transport sector to fulfil the Paris Agreement by 2050. *Prog. Photovolt. Res. Appl.* **2019**, *27*, 978–989. [[CrossRef](#)]
3. Sæle, H.; Morch, A.; Buonanno, A.; Caliano, M.; di Somma, M.; Papadimitriou, C. Development of Energy Communities in Europe. In Proceedings of the 18th International Conference on the European Energy Market (EEM), Ljubljana, Slovenia, 13–15 September 2022.
4. Fateh, D.; Eldoromi, M.; Birjandi, A.A.M. Uncertainty Modeling of Renewable Energy Sources. In *Scheduling and Operation of Virtual Power Plants*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 193–208.
5. Brenna, M.; Foiadelli, F.; Zaninelli, D.; Graditi, G.; Di Somma, M. The Integration of Electric Vehicles in Smart Distribution Grids with Other Distributed Resources. In *Distributed Energy Resources in Local Integrated Energy Systems*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 315–345. [[CrossRef](#)]
6. Papadimitriou, C.; Patsalides, M.; Venizelos, V.; Therapontos, P.; Efthymiou, V. Assessing Renewables Uncertainties in the Short-Term (Day-Ahead) Scheduling of DER. In *Distributed Energy Resources in Local Integrated Energy Systems*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 347–387.
7. Zakaria, A.; Ismail, F.B.; Lipu, M.H.; Hannan, M. Uncertainty models for stochastic optimization in renewable energy applications. *Renew. Energy* **2019**, *145*, 1543–1571. [[CrossRef](#)]
8. Li, Y.; Han, M.; Yang, Z.; Li, G. Coordinating Flexible Demand Response and Renewable Uncertainties for Scheduling of Community Integrated Energy Systems with an Electric Vehicle Charging Station: A Bi-level Approach. *IEEE Trans. Sustain. Energy* **2021**, *12*, 2321–2331. [[CrossRef](#)]
9. di Somma, M.; Graditi, G.; Heydarian-Forushani, G.; Shafie-khah, M.; Siano, P. Stochastic optimal scheduling of distributed energy resources with renewables considering economic and environmental aspects. *Renew. Energy* **2018**, *116*, 272–287. [[CrossRef](#)]
10. Zhou, Z.; Zhang, J.; Liu, P.; Li, Z.; Georgiadis, M.C.; Pistikopoulos, E.N. A two-stage stochastic programming model for the optimal design of distributed energy systems. *Appl. Energy* **2012**, *103*, 135–144. [[CrossRef](#)]
11. Hong, F.; Cuiying, W.; Lu, L.; Xuan, L. Review of Uncertainty Modeling for Optimal Operation of Integrated Energy System. *Front. Energy Res.* **2022**, *9*, 641337. [[CrossRef](#)]
12. Di Somma, M.; Buonanno, A.; Caliano, M.; Graditi, G.; Piazza, G.; Bracco, S.; Delfino, F. Stochastic Operation Optimization of the Smart Savona Campus as an Integrated Local Energy Community Considering Energy Costs and Carbon Emissions. *Energies* **2022**, *15*, 8418. [[CrossRef](#)]
13. Ashtari, A.; Bibeau, E.; Shahidinejad, S.; Molinski, T. PEV Charging Profile Prediction and Analysis Based on Vehicle Usage Data. *IEEE Trans. Smart Grid* **2011**, *3*, 341–350. [[CrossRef](#)]
14. Ali, S.; Lee, S.-M.; Jang, C.-M. Statistical analysis of wind characteristics using Weibull and Rayleigh distributions in Deokjeok-do Island—Incheon, South Korea. *Renew. Energy* **2018**, *123*, 652–663. [[CrossRef](#)]
15. Shoaib, M.; Siddiqui, I.; Amir, Y.M.; Rehman, S.U. Evaluation of wind power potential in Baburband (Pakistan) using Weibull distribution function. *Renew. Sustain. Energy Rev.* **2017**, *70*, 1343–1351. [[CrossRef](#)]
16. Labeeuw, W.; Deconinck, G. Customer sampling in a smart grid pilot. In Proceedings of the 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 22–26 July 2012.
17. Li, J.; Zhou, J.; Chen, B. Review of wind power scenario generation methods for optimal operation of renewable energy systems. *Appl. Energy* **2020**, *280*, 115992. [[CrossRef](#)]
18. Aien, M.; Hajebrahimi, A.; Fotuhi-Firuzabad, M. A comprehensive review on uncertainty modeling techniques in power system studies. *Renew. Sustain. Energy Rev.* **2016**, *57*, 1077–1089. [[CrossRef](#)]
19. Alipour, M.; Jalali, M.; Abapour, M.; Tohidi, S. Uncertainty Modeling in Operation of Multi-carrier Energy Networks. In *Planning and Operation of Multi-Carrier Energy Networks. Power Systems*; Springer: Berlin/Heidelberg, Germany, 2021.
20. Teke, A.; Yıldırım, B.H.; Çelik, Ö. Evaluation and performance comparison of different models for the estimation of solar radiation. *Renew. Sustain. Energy Rev.* **2015**, *50*, 1097–1107. [[CrossRef](#)]
21. Khatib, T.; Mohamed, A.; Sopian, K. A review of solar energy modeling techniques. *Renew. Sustain. Energy Rev.* **2012**, *16*, 2864–2869. [[CrossRef](#)]
22. Yorukoglu, M.; Celik, A.N. A critical review on the estimation of daily global solar radiation from sunshine duration. *Energy Convers. Manag.* **2006**, *47*, 2441–2450. [[CrossRef](#)]
23. Besharat, F.; Dehghan, A.A.; Faghih, A.R. Empirical models for estimating global solar radiation: A review and case study. *Renew. Sustain. Energy Rev.* **2013**, *21*, 798–821. [[CrossRef](#)]
24. Elminir, H.; Azzam, Y.; Younes, F. Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models. *Energy* **2007**, *32*, 1513–1523. [[CrossRef](#)]
25. Rahimikhoob, A. Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. *Renew. Energy* **2010**, *35*, 2131–2135. [[CrossRef](#)]
26. Aguiar, R.; Collares-Pereira, M. Statistical properties of hourly global radiation. *Sol. Energy* **1992**, *48*, 157–167. [[CrossRef](#)]
27. Duzen, H.; Aydin, H. Sunshine-based estimation of global solar radiation on horizontal surface at Lake Van region (Turkey). *Energy Convers. Manag.* **2012**, *58*, 35–46. [[CrossRef](#)]

28. Teke, A.; Yıldırım, H.B. Estimating the monthly global solar radiation for Eastern Mediterranean Region. *Energy Convers. Manag.* **2014**, *87*, 628–635. [[CrossRef](#)]
29. Korachagaon, I. Estimating Global Solar Radiation Potential for Brazil by Irrana-Bapat's Regression Models. *Int. J. Emerg. Technol. Adv.* **2012**, *2*, 178–186.
30. Khatod, D.K.; Pant, V.; Sharma, J. Evolutionary programming based optimal placement of renewable distributed generators. *IEEE Trans. Power Syst.* **2012**, *28*, 683–695. [[CrossRef](#)]
31. Salameh, Z.; Borowy, B.; Amin, A. Photovoltaic module-site matching based on the capacity factors. *IEEE Trans. Energy Convers.* **1995**, *10*, 326–332. [[CrossRef](#)]
32. Li, Y.; Zio, E. Uncertainty analysis of the adequacy assessment model of a distributed generation system. *Renew. Energy* **2012**, *41*, 235–244. [[CrossRef](#)]
33. Karaki, S.; Chedid, R.; Ramadan, R. Probabilistic performance assessment of autonomous solar-wind energy conversion systems. *IEEE Trans. Energy Convers.* **1999**, *14*, 766–772. [[CrossRef](#)]
34. Khan, M.F.N.; Malik, T.N. Probabilistic generation model for optimal allocation of PV DG in distribution system with time-varying load models. *J. Renew. Sustain. Energy* **2017**, *9*, 065503. [[CrossRef](#)]
35. Hung, D.Q.; Mithulananthan, N.; Lee, K.Y. Determining PV Penetration for Distribution Systems With Time-Varying Load Models. *IEEE Trans. Power Syst.* **2014**, *29*, 3048–3057. [[CrossRef](#)]
36. Hagan, K.E.; Oyebanjo, O.O.; Masaud, T.M.; Chaloo, R. A probabilistic forecasting model for accurate estimation of PV solar and wind power generation. In Proceedings of the 2016 IEEE Power and Energy Conference at Illinois (PECI), Urbana, IL, USA, 19–20 February 2016; pp. 1–5. [[CrossRef](#)]
37. Afzaal, M.U.; Sajjad, I.A.; Awan, A.B.; Paracha, K.N.; Khan, M.F.N.; Bhatti, A.R.; Zubair, M.; Rehman, W.U.; Amin, S.; Haroon, S.S.; et al. Probabilistic Generation Model of Solar Irradiance for Grid Connected Photovoltaic Systems Using Weibull Distribution. *Sustainability* **2020**, *12*, 2241. [[CrossRef](#)]
38. Roche, D.M.; Michalewicz, Z. Genetic Algorithms + Data Structures = Evolution Programs. *J. Am. Stat. Assoc.* **2000**, *95*, 347. [[CrossRef](#)]
39. Growe-Kuska, N.; Heitsch, H.; Romisch, W. Scenario Reduction and Scenario Tree Construction for Power Management Problems. In Proceedings of the 2003 IEEE Bologna Power Tech Conference Proceedings, Bologna, Italy, 23–26 June 2003; pp. 1–7.
40. Heitsch, H.; Romisch, W. Scenario Reduction Algorithms in Stochastic Programming. *Comput. Optim. Appl.* **2003**, *24*, 187–206. [[CrossRef](#)]
41. Buonanno, A.; Caliano, M.; Di Somma, M.; Graditi, G.; Valenti, M. Comprehensive method for modeling uncertainties of solar irradiance for PV power generation in smart grids. In Proceedings of the 2021 International Conference on Smart Energy Systems and Technologies (SEST), Vaasa, Finland, 6–8 September 2021.
42. PVGIS. Available online: [https://re.jrc.ec.europa.eu/pvg\\_tools/en/tools.html](https://re.jrc.ec.europa.eu/pvg_tools/en/tools.html) (accessed on 30 October 2022).
43. Tukey, J.W. *Exploratory Data Analysis*; Addison-Wesley: Boston, MA, USA, 1977.
44. Murphy, K.P. *Probabilistic Machine Learning—An Introduction*; The MIT Press: Cambridge, MA, USA, 2022.
45. Pourghasem, P.; Sohrabi, F.; Abapour, M.; Mohammadi-Ivatloo, B. Stochastic multi-objective dynamic dispatch of renewable and CHP-based islanded microgrids. *Electr. Power Syst. Res.* **2019**, *173*, 193–201. [[CrossRef](#)]
46. Rubinstein, R.Y.; Kroese, D.P. *Simulation and the Monte Carlo Method*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2016.