

Article

Solar Radiation Forecasting, Accounting for Daily Variability

Roberto Langella ¹, Daniela Proto ^{2,*} and Alfredo Testa ¹

¹ Department of Industrial and Information Engineering, Second University of Naples via Roma n. 29, Aversa 81031, Italy; roberto.langella@unina2.it (R.L.); alfredo.testa@unina2.it (A.T.)

² Department of Electrical Engineering and Information Technology, University of Naples Federico II, Via Claudio, n. 21, Napoli 80125, Italy

* Correspondence: danproto@unina.it; Tel.: +39-081-7683236

Academic Editors: John Ringwood and Anna Rita Di Fazio

Received: 30 October 2015; Accepted: 7 March 2016; Published: 15 March 2016

Abstract: Radiation forecast accounting for daily and instantaneous variability was pursued by means of a new bi-parametric statistical model that builds on a model previously proposed by the same authors. The statistical model is developed with direct reference to the Liu-Jordan clear sky theoretical expression but is not bound by a specific clear sky model; it accounts separately for the mean daily variability and for the variation of solar irradiance during the day by means of two corrective parameters. This new proposal allows for a better understanding of the physical phenomena and improves the effectiveness of statistical characterization and subsequent simulation of the introduced parameters to generate a synthetic solar irradiance time series. Furthermore, the analysis of the experimental distributions of the two parameters' data was developed, obtaining opportune fittings by means of parametric analytical distributions or mixtures of more than one distribution. Finally, the model was further improved toward the inclusion of weather prediction information in the solar irradiance forecasting stage, from the perspective of overcoming the limitations of purely statistical approaches and implementing a new tool in the frame of solar irradiance prediction accounting for weather predictions over different time horizons.

Keywords: solar radiation; solar irradiance; daily variability; instantaneous variability; statistical methods; parametric distributions; time series generation; forecasting

1. Introduction

Many photovoltaic (PV) applications, such as the sizing of stand-alone microgrids or sizing of energy storage systems for their inclusion in standalone (or grid connected) systems or day-ahead market offering, require forecasting PV production variability along different time scenarios, often also short and very short terms. For this purpose, it is necessary to forecast solar radiation on these time scenarios to account for weather variations during the day and even within each hour of the day.

Significant research is currently being devoted to the development of models to predict solar radiation and PV power. Common operational approaches to solar radiation forecasting include: (1) numerical weather prediction (NWP) models that infer local cloud information through the dynamic modeling of the atmosphere up to several days ahead [1]; (2) models using satellite remote sensing or ground-based sky measurements to infer the motion of clouds and project their impact in the future; (3) statistical time series models based on measured irradiance data applied for very short term forecasting in the range of minutes to hours [2–13].

In this context, daily variability of solar irradiance and its statistical characterization (and forecasting) is important. This has been treated in the relevant technical literature in [3–16], where the

clear sky index and the clearness index (or the normalized clearness index [12]) are typically used to describe solar irradiance and radiation.

In particular, in [14–16] the variability of solar irradiance is captured by observing the changes in clear sky index for a considered time interval. The primary usefulness of the clear sky index is the removal of diurnal and seasonal signals from a given set of radiation data to compute fluctuation power content ([17]).

Papers [3–13] refer to the clearness index and emphasize that the probability density distributions of the clearness index on short-term intervals, unlike those considered in longer intervals, present a multimodal nature [3,12]. In [13] it is stated that the instantaneous solar irradiance (in terms of clearness index) has bimodal probability distributions with peaks corresponding to clear and cloudy conditions. Regarding the statistical characterization of solar quantities, several attempts were proposed to identify the distribution functions that best fit solar data on a short time basis, some based on the Boltzmann statistics ([3,4]), others referring to the mixture of two normal distributions ([5,6]) or two bi-exponential probability density functions ([7,8]). These aspects make such methods mainly suitable for very short time forecasting with no exogenous inputs, as shown in [2], where five different techniques were assessed for one hour ahead and two hours ahead power forecasting of PV plants.

A common characteristic of the aforementioned models is that they refer to a unique parameter accounting for solar irradiance variations.

In [18] a statistical model referring to the clear sky index was proposed to characterize solar radiation, taking into account the daily variability in terms of two proper parameters. More specifically, solar irradiance was described as the sum of two components, characterized by specific quantities that take into account mean daily weather conditions and instantaneous variations of the solar irradiance, respectively. It was demonstrated that experimental characterization of the distributions of the two introduced quantities was suitable for a statistical characterization and for generating synthetic time series characterized by both daily and instantaneous variability starting from experimental probability distributions.

This paper introduces an evolution of the model proposed in [18]. The statistical model is developed with direct reference to the Liu-Jordan clear sky theoretical expression but is not bound by specific clear sky models; it accounts separately for the mean daily variability and for variation of solar irradiance during the day by means of two corrective parameters. The first parameter accounts for the mean daily variability, while the second accounts for the variation in solar irradiance during the day. This new proposal allows for a better understanding of the physical phenomena and improves the effectiveness of statistical characterization and subsequent simulation of the introduced parameters to generate synthetic solar irradiance time series, also starting from parametric analytical distributions. The analysis of the parametric analytical distributions that best fit the experimental distributions of the two-parameter data was also developed with reference to field measurements [19], obtaining opportune single parametric distributions or mixtures of more than one distribution. Moreover, the model's ability to combine with weather predictions was investigated, starting from the consideration that changing the variability time scale (e.g., introducing two or more sub-parts of the day) seems to allow for using it for both long-term and short-term scenarios. Thus, a further model improvement was developed by including weather prediction information in the forecasting stage, thus overcoming the limitation of a purely statistical approach. In the application proposed in this paper, the model was tested only for day-ahead forecasting, even if the prospect of utilizing it in other time scenarios, with various forecasting techniques, seems promising.

The remainder of the paper is organized as follows. Section 2 includes a description of the model; Section 3 describes the statistical analysis stage. Simulation and forecasting are developed in Sections 4 and 5 while Section 6 reports the results of field measurement analyses. Finally, conclusions are drawn in Section 7.

2. Proposed Model

The proposed model is an evolution of that proposed in [18], introducing a new expression decomposing the measured solar irradiance at time $n\Delta t$ (i.e., the n th sample of the j th day of the m th month), $R_{j,m}(n\Delta t)$, in two components:

$$R_{j,m}(n\Delta t) = S_{LJ_{j,m}}(n\Delta t) \cdot [C_m \cdot k_{j,m} + \varepsilon'_{j,m}(n\Delta t)] = S_{LJ_{j,m}}(n\Delta t) \cdot C_m \cdot [k_{j,m} + \varepsilon''_{j,m}(n\Delta t)] \quad (1)$$

where:

- $S_{LJ_{j,m}}(n\Delta t)$ is the “clear sky” theoretical model developed by Liu-Jordan at ground level [20,21];
- C_m is the monthly clear sky index for the m th month;
- $k_{j,m}$, mean daily variability for the j th day of the m th month, is a daily parameter, which plays the role of correction factor of C_m accounting for specific daily weather conditions;
- $\varepsilon'_{j,m}(n\Delta t)$, instantaneous variability for the n th time interval of the j th day of the m th month, is an additive “noise” component accounting for variations of the solar irradiance during a day;
- $\varepsilon''_{j,m}(n\Delta t)$ is equal to $\varepsilon'_{j,m}(n\Delta t)/C_m$.

It is worth noting that in Equation (1) any clear sky model (other than that proposed by Liu & Jordan) can be used without compromising the validity of the proposed model.

The second expression in Equation (1) is a useful alternative to the first and allows for establishing correspondences with popular simplified forms as it will appear clearly in the following sections (e.g., in Section 4, Equation (14)).

It is worth noting that the following constraints apply for $k_{j,m}$ and $\varepsilon'_{j,m}(n\Delta t)$ values:

$$0 \leq k_{j,m} \leq \frac{1}{C_m} \quad (2)$$

$$-C_m \cdot k_{j,m} \leq \varepsilon'_{j,m}(n\Delta t) \leq 1 - C_m \cdot k_{j,m} \quad (3)$$

The parameter C_m is calculated for the m th month as:

$$C_m = \frac{W_{o_m}}{W_{LJ_m}} \quad (4)$$

where W_{o_m} is the historical m th month average value of the daily solar radiation issued by meteorological services [19,22]; W_{LJ_m} is the m th month average value of the daily theoretical Liu-Jordan radiation, that is:

$$W_{LJ_m} = \frac{\Delta t}{N_m} \cdot \sum_{j=1}^{N_m} \left[\sum_{n=1}^{N_n} S_{LJ_{j,m}}(n\Delta t) \right] \quad (5)$$

being N_m the number of days of the m th month and N_n the number of Δt of a day.

The formulation corresponding to Equation (1) given in [18] is:

$$R_{j,m}(n\Delta t) = S_{W_{j,m}}(n\Delta t) \cdot k_{j,m} + \varepsilon_{j,m}(n\Delta t) \quad (6)$$

and is based on:

- $S_{W_{j,m}}(n\Delta t)$, which is the historically expected irradiance at ground level for n th sample of the j th day of the m th month, in absence of daily variability and instantaneous weather variability respect to historical behavior, that is:

$$S_{W_{j,m}}(n\Delta t) = S_{LJ_{j,m}}(n\Delta t) \cdot C_m \quad (7)$$

- $k_{j,m}$ as in Equation (1),
- $\varepsilon_{j,m}(n\Delta t)$ given by:

$$\varepsilon_{j,m}(n\Delta t) = S_{LJ_{j,m}}(n\Delta t) \cdot [\varepsilon'_{j,m}(n\Delta t)] \quad (8)$$

Figure 1a,b show, for the 11th and 12th of August, respectively, of 2004 for a site in the south of Italy, the profiles of the measures [19] of solar irradiance, $R_{j,m}(n\Delta t)$, of its components $k_{j,m} \cdot S_{W_{j,m}}(n\Delta t)$ and $\epsilon_{j,m}(n\Delta t)$ and of the expected irradiance at ground level, $S_{W_{j,m}}(n\Delta t)$ according to Equation (6). Of course, a similar interpretation of the same figure can be effected in terms of the components given by Equation (1).

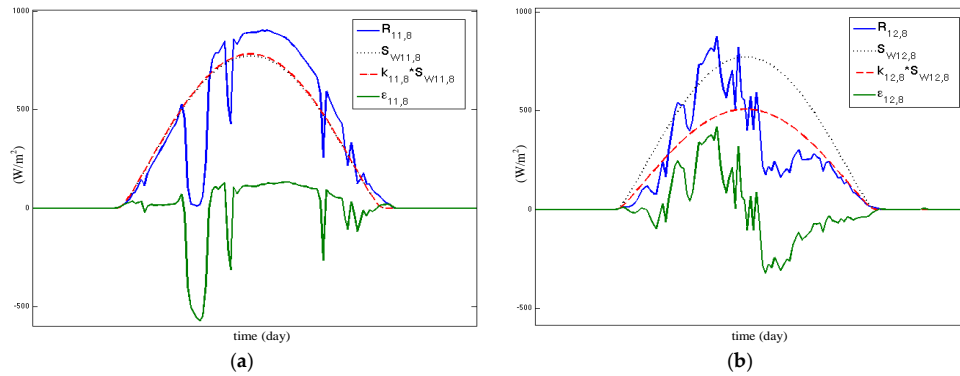


Figure 1. Profile of solar irradiance and its components, according to Equation (6), on (a) 11 August 2004; and (b) 12 August 2004.

In the figure, the regularity of the red and dotted black curves can be observed; they have the same shape as the theoretical Liu-Jordan profile multiplied by C_m ; the red curve derives from the product of the black curve and $k_{j,m}$. The measured solar irradiance, $R_{j,m}(n\Delta t)$, (blue curve), is affected by irregularities deriving from the additional “noise” $\epsilon_{j,m}(n\Delta t)$ (green curve).

The 11th of August was characterized by a daily radiation (area of the red curve) almost equal to that expected (area of the black curve); the irregularities appear concentrated in some specific parts of the day. The 12th of August was characterized by a daily radiation (area of the red curve) sensibly lower than that expected (area of the black curve); the irregularities appear to be scattered along the day with morning hours better than those of the afternoon.

The application of the proposed model includes two stages: (1) statistical analysis of irradiance quantities summarized in the flow chart of Figure 2 and detailed in Section 3; (2) simulation and forecast of the solar irradiance summarized in the flow chart of Figure 3 and detailed in Section 4.

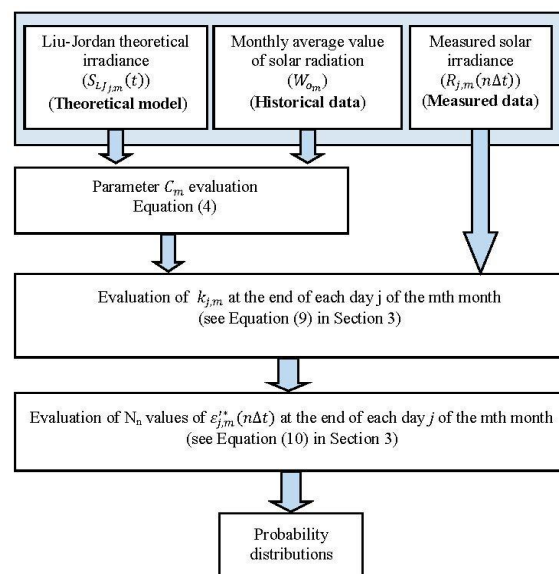


Figure 2. Flow chart of the “Statistical Analysis” stage.

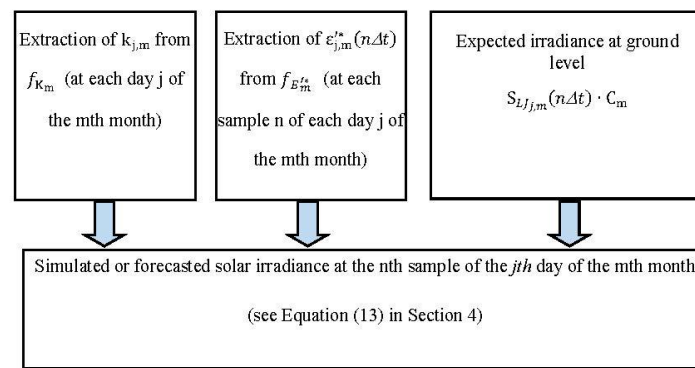


Figure 3. Flow chart of the “Simulation of solar irradiance” stage.

3. Statistical Analysis

The statistical characterization of the parameters $k_{j,m}$ and $\epsilon_{j,m}^*(n\Delta t)$, which refer to the j th day of the m th month, is performed by identifying their distributions at each month m . This stage includes the derivation of the distributions of the parameters and their statistical characterization.

Inputs are the historical data of solar irradiance measured at the specific location where the procedure is applied.

3.1. Experimental Distributions

Starting from the measurements of solar irradiance $R_{j,m}(n\Delta t)$, the $k_{j,m}$ parameter can be evaluated from Equation (1) as follows:

$$k_{j,m} = \frac{\sum_1^{N_n} R_{j,m}(n\Delta t) - \sum_1^{N_n} \epsilon_{j,m}^*(n\Delta t) \cdot S_{LJ,m}(n\Delta t)}{C_m \sum_1^{N_n} S_{LJ,m}(n\Delta t)} = \frac{\sum_1^{N_n} R_{j,m}(n\Delta t)}{C_m \sum_1^{N_n} S_{LJ,m}(n\Delta t)} \quad (9)$$

with N_n number of samples in a day and $\sum_1^{N_n} \epsilon_{j,m}^*(n\Delta t) \cdot S_{LJ,m}(n\Delta t) = 0$.

The parameter $k_{j,m}$ —which is different from the clearness index typically used for characterizing the solar radiation—assumes values lower than 1 on cloudy days, when the total energy registered is lower than the average energy expected in that month according to historical data, and assumes values greater than 1 on sunny days. Its monthly expected value, for a huge number of observations, is equal to 1.

With reference to $\epsilon_{j,m}^*(n\Delta t)$, the following relation can be derived from Equation (1):

$$\epsilon_{j,m}^*(n\Delta t) = \frac{R_{j,m}(n\Delta t) - S_{LJ,m}(n\Delta t) \cdot C_m \cdot k_{j,m}}{S_{LJ,m}(n\Delta t)} \quad (10)$$

Note that the values of all the $N_n + 1$ parameters for each day can be evaluated only at the end of the day—that is, when the measurements $R_{j,m}(n\Delta t)$ for all the N_n samples of the day have been carried out. Once the data for $k_{j,m}$ and $\epsilon_{j,m}^*(n\Delta t)$ have been collected, a statistical analysis can be performed with reference to the m th month of the year, obtaining the experimental joint distribution f_{K_m, E_m^*} or the experimental marginal distributions f_{K_m} and $f_{E_m^*}$. Details on the statistical behavior of these quantities are provided in Section 3.2.

In order to statistically characterize the above parameters, the procedure was tested in [18] to derive the daily joint probability density function (*pdf*) $f_{k_{j,m}, \epsilon_{j,m}}$ or, as an alternative, the marginal *pdfs* (f_{K_m}, f_{E_m}). Considering the marginal is preferable due to the scarcity of available data and to practical reasons without relevant effects on the accuracy. For both the parameters, $k_{j,m}$ and $\epsilon_{j,m}(k\Delta t)$, the analysis was limited to the case of experimental probability mass functions (*pmf*) as obtainable from the data observed.

3.2. Parametric Distributions

We focused our attention on the analysis of the parametric *pdfs* that best fit the distributions of the data experimentally obtained. This approach, in fact, has the advantage of simply referring to specific values of the parameters of the distributions of the random variables K_m and E_m^{t*} , thus making the data storage very efficient and the simulation/forecast procedure extremely straightforward.

In the following subsections, the parametrical characterizations of the distributions of $k_{j,m}$ and $\varepsilon_{j,m}^{t*}(k\Delta t)$ are shown separately.

3.2.1. Daily Variability Parameter

From the observation of the *pmfs* of K_m at each month of the year, it emerged that they always exhibited a multimodal nature. It also emerged that the different seasons were typically characterized by very different behaviors of the *pmfs*. Some months of the year, for example, are characterized by scattered distributions in correspondence of low values of $k_{j,m}$ which appear almost uniformly distributed, whereas an increased concentration of the occurrences can be observed for higher values of $k_{j,m}$, with the majority of occurrences around a specific value, analogous to Gaussian behavior. In other cases, the multimodality is characterized by shapes that are more similar to either Gaussian or Weibull distributions, with modes appearing for both the low and high values of $k_{j,m}$. In the most general case, the heterogeneity of the distributions of $k_{j,m}$ at each month could be represented by the mixture of two or more distributions that can be expressed as weighted sum of *pdfs*:

$$f(x) = \sum_{i=1}^N w_i \cdot f_i(x) \tag{11}$$

where $f_i(x)$ are proper *pdfs*, each characterized by specific values of its parameters and w_i are opportune weighting factors.

In our analysis, several options were considered for fitting the distributions of data with probability density curves. For each of them, the root mean square error (RMSE) was evaluated between $f(x)$ and the experimental *pmf* in order to choose the most suitable parametric distribution to reproduce the selected data. In all the cases, it emerged that a mixture of two *pdfs* could be an acceptable trade-off between accuracy and simplicity.

The distributions used in this paper to compose the mixtures were Gaussian, Weibull, and uniform, whose analytical expressions are reported in Table 1.

Table 1. Density functions of the distributions chosen for the analysis.

Name	Density Function	Parameters
Weibull	$y = p_2 p_1^{-p_2} x^{p_2-1} e^{-\left(\frac{x}{p_1}\right)^{p_2}} \quad x \in [0, \infty)$	p_1 scale parameter p_2 shape parameter
Gaussian	$y = \frac{1}{p_2\sqrt{2\pi}} e^{-\frac{(x-p_1)^2}{2p_2^2}} \quad x \in (-\infty, +\infty)$	p_1 mean p_2 standard deviation
Uniform	$y = \frac{1}{p_2-p_1} \quad x \in [p_1, p_2]$	p_1 lower endpoint (minimum) p_2 upper endpoint (maximum)

3.2.2. Instantaneous Variability Parameter

A mono-modal nature of the distribution of the instantaneous variability parameter, $\varepsilon_{j,m}^{t*}(n\Delta t)$, was observed for all months of the year. The behavior was found to be very well approximated by using a *T-location scale* distribution:

$$\frac{\Gamma\left(\frac{p_3+1}{2}\right)}{p_2\sqrt{p_3\pi}\Gamma\left(\frac{p_3}{2}\right)} \left[\frac{p_3 + \left(\frac{x-p_1}{p_2}\right)^2}{p_3} \right]^{-\left(\frac{p_3+1}{2}\right)} \tag{12}$$

where p_1 is the location parameter, p_2 is the scale parameter, and p_3 is the degree of freedom.

The Gaussian distribution was also tested but the results always appeared to be worse than those obtained with the T-location scale.

4. Simulation of Solar Radiation

Simulation results are illustrated for the purely statistical case.

Starting from the *pdfs* of the two variability parameters, $k_{j,m}$ and $\varepsilon_{j,m}^{i*}(n\Delta t)$, it is possible to simulate the solar irradiance with reference to each $n\Delta t$ of the j th day of the m th month as:

$$\hat{R}_{j,m}(n\Delta t) = S_{LJ_{j,m}}(n\Delta t) \cdot [C_m \cdot k_{j,m} + \varepsilon_{j,m}^{i*}(n\Delta t)] \tag{13}$$

where $k_{j,m}$ and $\varepsilon_{j,m}^{i*}(n\Delta t)$ assume the values obtained by means of opportune extractions from the marginal parametric distributions f_{K_m} and $f_{E_m^*}$; C_m and $S_{LJ_{j,m}}(n\Delta t)$ assume values deriving from their definitions. $\hat{R}_{j,m}(n\Delta t)$, evaluated according to Equation (13), is able to reproduce synthetic time series of solar irradiance whose variability is similar to that observed in the building stage of f_{K_m} and $f_{E_m^*}$.

It is useful to recall that the expression typically used in the design stage to evaluate the power producibility is based on the simple expression:

$$\hat{R}_{j,m}(n\Delta t) = S_{LJ_{j,m}}(n\Delta t) \cdot C_m \tag{14}$$

able to reproduce the expected value of the irradiance without taking into account the phenomena of daily and instantaneous variability.

5. Forecasting Solar Radiation Including Weather Predictions

Forecasting solar radiation including weather predictions requires a more complex procedure than the simple use of Equation (13), as illustrated in Figure 4.

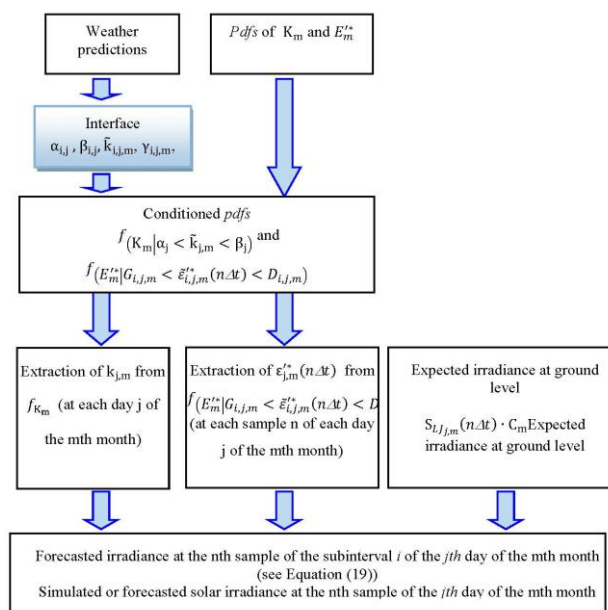


Figure 4. Flow chart of the “Forecasting of solar irradiance including weather predictions” stage.

Here reference is made to day-ahead forecasting, but the procedure can be implemented in different time scenarios depending on the specific weather prediction model used. For this purpose, it has to be noted that in principle any weather prediction model can be used; of course, the critical aspect is to provide an interface procedure to convert the information coming from weather prediction into corresponding proper conditioning of the statistical distributions of $k_{j,m}$ and $\varepsilon_{j,m}'^*$ that our statistical approach utilizes.

In what follows, the aforementioned critical aspect is considered to be already solved and the availability of properly conditioned *pdfs* is assumed. It is worth noting that, in this framework, the statistical approach proposed in this paper could play the role of representing the variability characteristic observed in the past, improving, as a statistical post-processing, the outputs of NWP models.

First of all, $k_{j,m}$ values should be extracted according to day-ahead weather predictions with reference to the conditioned *pdfs* expressed, for instance, as:

$$f_{(K_m | \alpha_j < \tilde{k}_{j,m} < \beta_j)} \quad (15)$$

where $\tilde{k}_{j,m}$ is the weather prediction for $k_{j,m}$ and α_j and β_j the edges of the interval of confidence of $\tilde{k}_{j,m}$. The conditioned *pdfs* can be easily derived from the original unconditioned f_{K_m} .

More complex is the inclusion of weather prediction for $\varepsilon_{j,m}'^*$ ($n\Delta t$) values. A possible procedure can be:

1. division of the daylight hours in N_i subintervals;
2. expression of weather prediction for each i th subinterval of duration Δi as an opportune value of $\tilde{k}_{i,j,m}$, $\alpha_{i,j}$ and $\beta_{i,j}$;
3. evaluation of $\tilde{k}_{j,m}$ as opportune weighted mean of $\tilde{k}_{i,j,m}$:

$$\tilde{k}_{j,m} = \frac{\sum_{i=1}^{N_i} \tilde{k}_{i,j,m} \cdot \Delta i}{\sum_{i=1}^{N_i} \Delta i} \quad (16)$$

4. evaluation of opportune intervals of variability for $\varepsilon_{j,m}'^*$ ($n\Delta t$) for each i th subinterval, as:

$$-\gamma_{i,j,m}(C_m \tilde{k}_{i,j,m}) \leq \varepsilon_{i,j,m}'^*(n\Delta t) \leq \delta_{i,j,m}(1 - C_m \tilde{k}_{i,j,m}) \quad (17)$$

with $\gamma_{i,j,m}$ and $\delta_{i,j,m}$ assuming values minor or equal to 1 deriving from weather prediction;

5. Extraction of the $k_{i,j,m}$ values for each of the N_i subintervals of the j th day according to the day-ahead weather prediction from conditioned distributions $f_{(k_m | \alpha_{i,j} < \tilde{k}_{i,j,m} < \beta_{i,j})}$ analogously to Equation (15);
6. Extraction of the $\varepsilon_{i,j,m}'^*$ ($n\Delta t$) values according to day-ahead weather predictions, for instance, with reference to the conditioned *pdfs* expressed as:

$$f_{(E_m'^* | G_{i,j,m} < \tilde{\varepsilon}_{i,j,m}'^*(n\Delta t) < D_{i,j,m})} \quad (18)$$

with $G_{i,j,m} = -\gamma_{i,j,m}(C_m \tilde{k}_{i,j,m})$ and $D_{i,j,m} = \delta_{i,j,m}(1 - C_m \tilde{k}_{i,j,m})$

For the i th subinterval the forecasted irradiance is given by:

$$\hat{R}_{i,j,m}(n\Delta t) = S_{Lj,m}(n\Delta t) \cdot [C_m \cdot k_{i,j,m} + \varepsilon_{i,j,m}'^*(n\Delta t)]. \quad (19)$$

6. Field Measurement Analysis

In order to test the proposed model, the data, related to six-year observations (years 2004–2009) from a location in the south of Italy whose latitude and longitude are $40^\circ 48.8'$, $14^\circ 20.3'$ [19], were

analyzed. The years 2004–2008 were used to calibrate the model (identification period) while the year 2009 was used to validate the model (validation period). The data used represent the global horizontal solar irradiance and are available on a 10 min basis, already filtered in order to clean values deriving from measurements errors or absence of measurements. Due to the characteristics of the proposed method, the data were utilized without any *ad hoc* time series pre-processing finalized to extract trends or seasonality. In the following subsections, the results of the fitting of the aforementioned data (*i.e.*, K_m and E_m^{*}) are reported first. Then, the results of the application of the model for forecasting purposes are shown together with an evaluation of the method's performance.

6.1. Parametric Distribution Fitting

The results are reported separately for mean daily variability and instantaneous variability.

6.1.1. Mean Daily Variability

Table 2 reports the parameters of the distributions and weights of their mixture chosen to fit the K_m data with reference to all months of the year. The values of the corresponding RMSEs are also reported in the table. The analysis was effected for all months of the year, each characterized by a proper combination of distributions. Several attempts were made to find the best combination of the above parametric distributions. The choice fell on those characterized by the minor value of the RMSE.

Table 2. Mixtures of distributions resulting from the analysis of K_m (p'_1 , p'_2 , p''_1 and p''_2 are defined in Table 1 and w_1 and w_2 in Equation (11)). Root mean square error: RMSE.

Month	1 st Distribution	p'_1	p'_2	Weight (w_1)	2 st Distribution	p''_1	p''_2	Weight (w_2)	Norm. RMSE Mean (%)
January	Uniform	0.0977	1.371	0.7800	Gaussian	1.310	0.0688	0.2200	5.7%
February	Uniform	0.0007	1.391	0.7572	Gaussian	1.308	0.0832	0.2428	4.3%
March	Gaussian	0.6219	0.3122	0.4671	Weibull	1.2556	9.758	0.5329	4.8%
April	Gaussian	0.8279	0.2943	0.5740	Gaussian	1.283	0.0652	0.4260	6.6%
May	Gaussian	0.9053	0.2387	0.5626	Gaussian	1.238	0.0413	0.4374	7.7%
June	Weibull	0.9127	5.153	0.3136	Weibull	1.173	25.81	0.6864	8.5%
July	Weibull	1.141	23.33	0.9119	Gaussian	0.8553	0.2041	0.0881	9.7%
August	Gaussian	0.8664	0.1914	0.2526	Gaussian	1.159	0.0399	0.7474	8.8%
September	Weibull	0.8978	3.942	0.3323	Gaussian	1.213	0.0587	0.6677	5.1%
October	Weibull	0.8938	3.606	0.4896	Gaussian	1.215	0.0664	0.5104	6.9%
November	Gaussian	0.6605	0.2999	0.6448	Gaussian	1.277	0.0804	0.3552	5.6%
December	Weibull	0.7376	2.187	0.5712	Weibull	1.322	14.17	0.4288	4.1%

In order to provide a graphical example, Figure 5 shows the *pmfs* and their fitting parametric distributions for the months of January, April, July, and October (representatives of the four seasons) where the experimental *pmfs* were fitted by the mixture of (1) Uniform and Gaussian distribution in the case of January; (2) two Gaussian distributions in the case of April; and (3) Weibull and Gaussian distributions in the cases of both July and October.

From the analysis of the figure, the bimodal nature of f_{K_m} clearly appears. It can also be observed how the selected mixtures of parametric distributions quite accurately approximate the *pmfs*. Figure 5 also reveals that, in some months, more than two modes could be identified for the *pmfs*. The results of the analysis performed, however, revealed that mixing more than two parametric distributions doesn't significantly improve the statistical characterization of K_m .

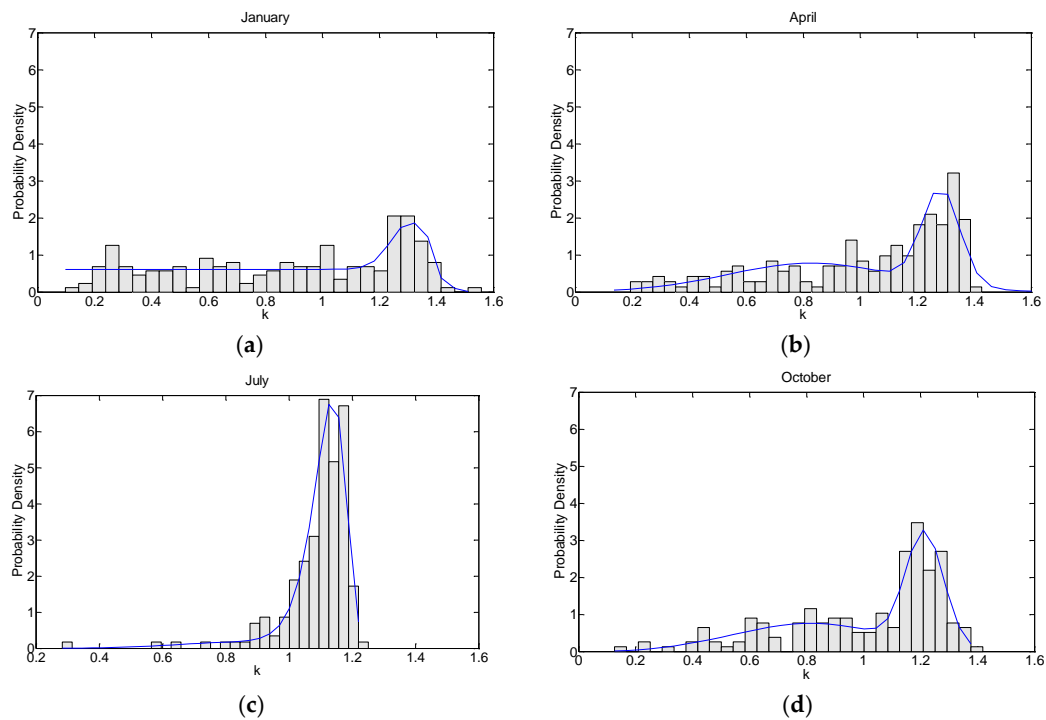


Figure 5. Discrete probability distributions of K_m and their fitting probability density functions (*pdfs*) with reference to the months of (a) January; (b) April; (c) July; and (d) October.

6.1.2. Instantaneous Daily Variability

The analysis performed for E_m^{*} , for all months of the year, allowed us to identify the parameter values associated with the t location-scale distribution of each month. Table 3 synthetically reports the results of the analysis.

Table 3. Parameters of the t location-scale distribution resulting from the analysis of E_m^{*} .

Month	t Location-Scale Parameters		
	Location (p1)	Scale (p2)	Degree of Freedom (p3)
January	−0.00181672	0.139726	2.29907
February	−0.00367767	0.152067	2.82915
March	−0.00259997	0.167867	3.68524
April	−0.0059413	0.159816	3.43902
May	0.00258189	0.144355	2.82872
June	0.0120793	0.0847774	1.42587
July	0.0137682	0.0448222	1.06117
August	0.00911046	0.0483726	1.11224
September	0.00957822	0.0836522	1.36912
October	0.000583434	0.124882	2.08324
November	−0.00760512	0.149513	2.27951
December	−0.00882419	0.144097	2.15243

The *pmfs* and their fitting parametric distributions for the months of January, April, July, and October are reported in Figure 6, showing the suitability of this distribution for E_m^{*} .

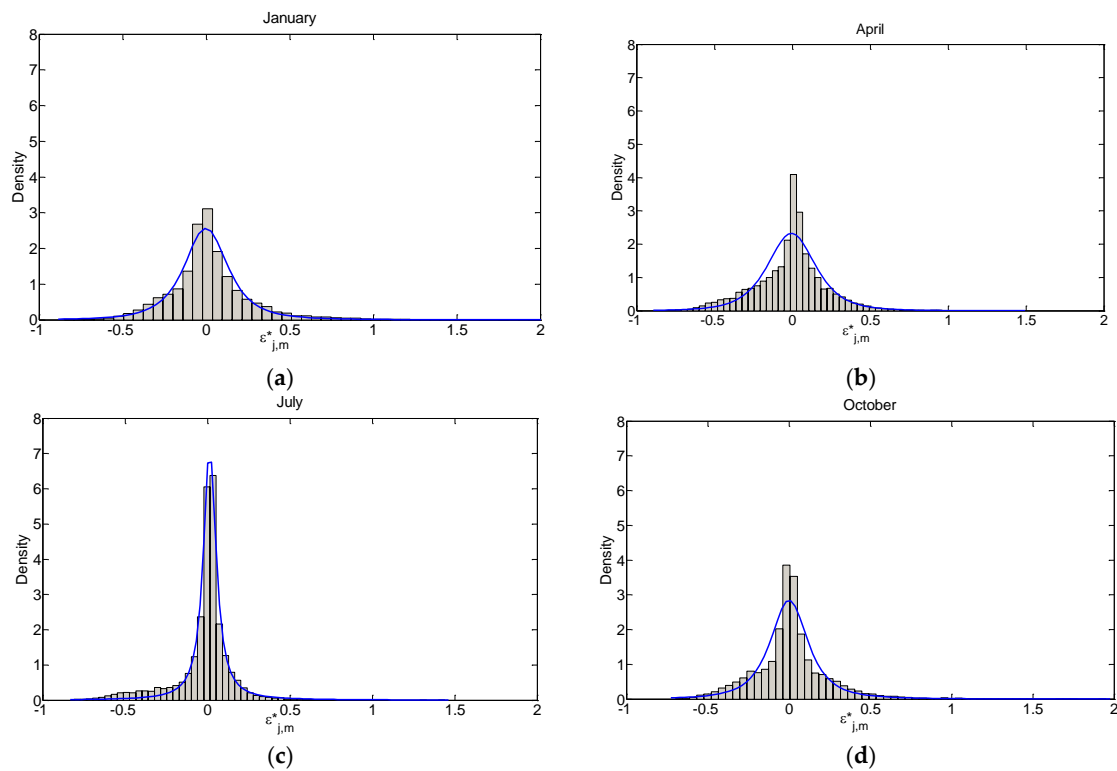


Figure 6. Discrete probability distributions of $E_m'^*$, and their fitting *pdfs* with reference to the months of (a) January; (b) April; (c) July; and (d) October.

6.2. Forecasting Application

Starting from the parametric *pdfs* of K_m and $E_m'^*$, the forecasting procedure described in Section 4 was applied to several tests performed with reference to all months of the year. For the sake of brevity, in what follows, only the results referring to the day-ahead forecasts for the months of January, April, July, and October are reported.

The results of the application of Parametric Marginal distribution functions (PARAMETRIC) have been compared with those obtained by means of Experimental Marginal distribution functions (EXPERIMENTAL) [18] and the Persistence method (PERS, one of the most commonly used reference methods [23]).

In order to compare the forecast methods performances, the mean average % error, $MAE_{j,m}$, the mean bias % error, $MBE_{j,m}$, and the root mean square % error, $RMSE_{j,m}$, have been calculated for each day of the m th month:

$$MAE_{j,m} = \frac{1}{N_n} \frac{\sum_{n=1}^{N_n} |\hat{R}_{j,m}(n\Delta t) - R_{j,m}(n\Delta t)|}{MR_{j,m}} \cdot 100 \tag{20}$$

$$MBE_{j,m} = \frac{1}{N_n} \frac{\sum_{n=1}^{N_n} (\hat{R}_{j,m}(n\Delta t) - R_{j,m}(n\Delta t))}{MR_{j,m}} \cdot 100 \tag{21}$$

$$RMSE_{j,m} = \frac{\sqrt{\frac{1}{N_n} \sum_{n=1}^{N_n} (\hat{R}_{j,m}(nt) - R_{j,m}(nt))^2}}{MR_{j,m}} \cdot 100 \tag{22}$$

with $MR_{j,m}$ being the daily mean value of the measured solar irradiance evaluated over the $N_{s,j,m}$ samples of the day with solar irradiance different from zero:

$$MR_{j,m} = \frac{\sum_{n=1}^{N_{s,j,m}} R_{j,m}(n\Delta t)}{N_{s,j,m}} \quad (23)$$

Skill Scores, evaluated with reference to the Persistence method of MAE and RMSE, have been also evaluated by means of the following expressions where METH refers either to EXPERIMENTAL or PARAMETRIC forecasting methods:

$$SS_{MAE,j,m}^{METH} = 1 - \frac{MAE_{j,m}^{METH}}{MAE_{j,m}^{PERS}} \quad (24)$$

$$SS_{RMSE,j,m}^{METH} = 1 - \frac{RMSE_{j,m}^{METH}}{RMSE_{j,m}^{PERS}} \quad (25)$$

Figure 7 reports $MAE_{j,m}$ values for the two different forecast methods with reference to each day of: (a) January, (b) April, (c) July, and (d) October.

It is possible to observe that:

- For all the months considered, the daily error varies in a quite wide range (e.g., in October from less than 10% on the 20th to almost 50% on the 13th);
- The performances of the methods are very close to each other.

Monthly mean values of the normalized MAE, normalized MBE, and normalized RMSE for the two methods are reported in Table 4, together with the corresponding values for the Persistence method, for the same months considered in Figure 7.

Table 5 reports Skill Scores obtained using the Persistence method as reference of the monthly mean of $MAE_{j,m}$ and of $RMSE_{j,m}$ for the months given in Figure 7. EXPERIMENTAL refers to the use of the experimental marginal distribution functions proposed in [18] and PARAMETRIC refers to the use of the parametric distribution functions proposed in this paper.

It is possible to observe that:

- Both methods presented by the authors always have better performance than the Persistence method (Table 4);
- The use of parametric distributions gives almost the same performance as the experimental distributions (Table 4);
- Skill Scores for both MAE and RMSE quantify the improvement of the performance with respect to the Persistence method (Table 5);
- The performance of the proposed methods is also better for the month of July, when more stable weather conditions allow the Persistence method to have good performance (Table 5).

Table 4. Monthly mean values of $MAE_{j,m}$, $MBE_{j,m}$ and of $RMSE_{j,m}$ for the months in Figure 7.

Month	EXPERIMENTAL			PARAMETRIC			PERSISTENCE		
	MAE (%MR)	MBE (%MR)	RMSE (%MR)	MAE (%MR)	MBE (%MR)	RMSE (%MR)	MAE (%MR)	MBE (%MR)	RMSE (%MR)
January	28	0	36	28	0	35	72	29	94
April	35	-1	44	35	0	45	68	21	90
July	12	-1	16	12	0	15	15	2	21
October	21	-1	27	22	-1	27	38	9	52

Table 5. Skill Scores, referring to the Persistence method, of the monthly mean of $MAE_{j,m}$, and of $RMSE_{j,m}$ for the months in Figure 7.

Month	EXPERIMENTAL		PARAMETRIC	
	SS_{MAE} (pu)	SS_{RMSE} (pu)	SS_{MAE} (pu)	SS_{RMSE} (pu)
January	0.6	0.6	0.6	0.6
April	0.5	0.5	0.5	0.5
July	0.2	0.2	0.2	0.3
October	0.4	0.4	0.5	0.5

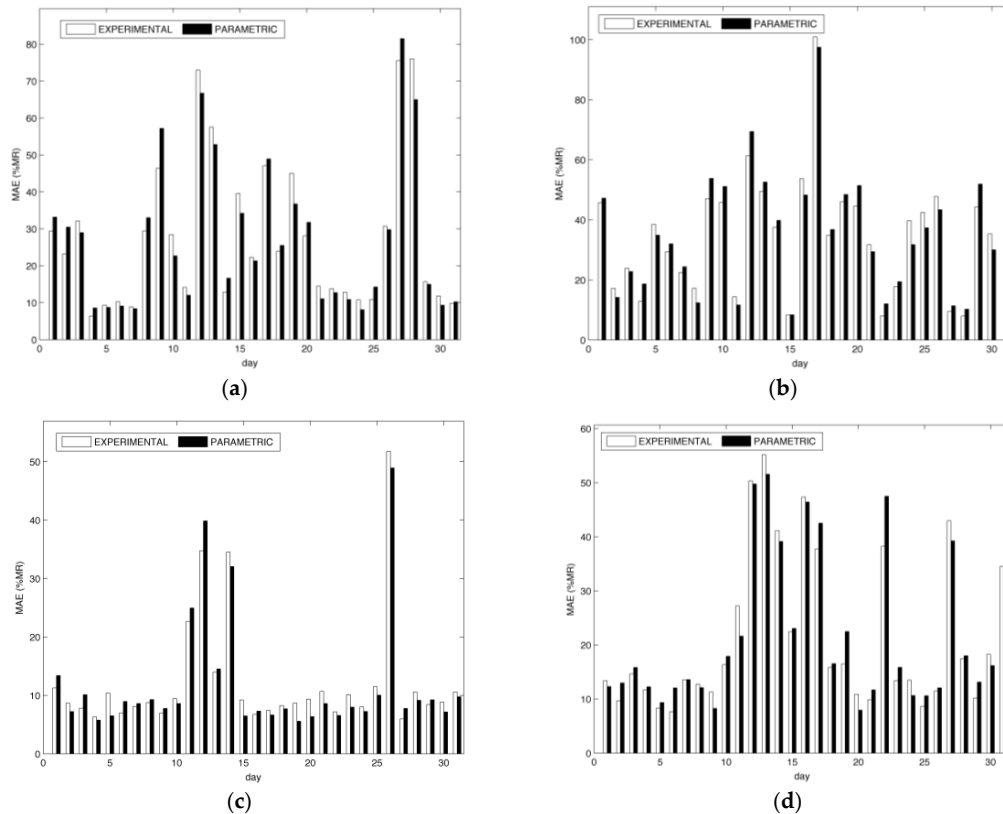


Figure 7. $MAE_{j,m}$ values versus the day of the month for the two different forecast methods: EXPERIMENTAL and PARAMETRIC. (a) January, (b) April, (c) July, and (d) October.

Going back to Figure 7, it is possible to observe that only some days of the month show MAE values appreciably higher than the rest of the month. In fact, looking at the monthly mode values of the MAE metric reported in Table 6 for the four months analyzed, it is evident that the mode values are appreciably lower than the mean values (see Table 4).

Table 6. Monthly mode values of $MAE_{j,m}$ for the months in Figure 7.

Month	EXPERIMENTAL	PARAMETRIC
	MAE (%MR)	MAE (%MR)
January	6	8
April	8	8
July	6	6
October	8	8

In order to understand the reason why high error values appear in some specific days (see Figure 7), the solar irradiance on such days has been further analyzed.

As an example, Figure 8a reports, with reference to the 22nd of October (that is, one of the days characterized by a high error value), the measured solar irradiance, $R_{22,10}$, the forecasted solar irradiance, $\hat{R}_{22,10}$, the expected clear sky theoretical solar irradiance, $S_{LJ22,10}$, and the product of the factor, $k_{22,10}$ by $S_{W22,10}$ versus the time. It is evident that the daily radiation is lower than that expected from the historical data available for the specific site ($k_{22,10}$ is lower than 1); furthermore, 2/3 of daylight hours are characterized by very low irradiance values while the remaining 1/3 of the time is characterized by high values. So this particular day is characterized by a very strong correlation between the k and ε values. This highlights an intrinsic limitation of purely statistical forecasting, which is able to follow the average behavior of the solar irradiance but randomly generates statistically independent epsilon values through the whole daylight period.

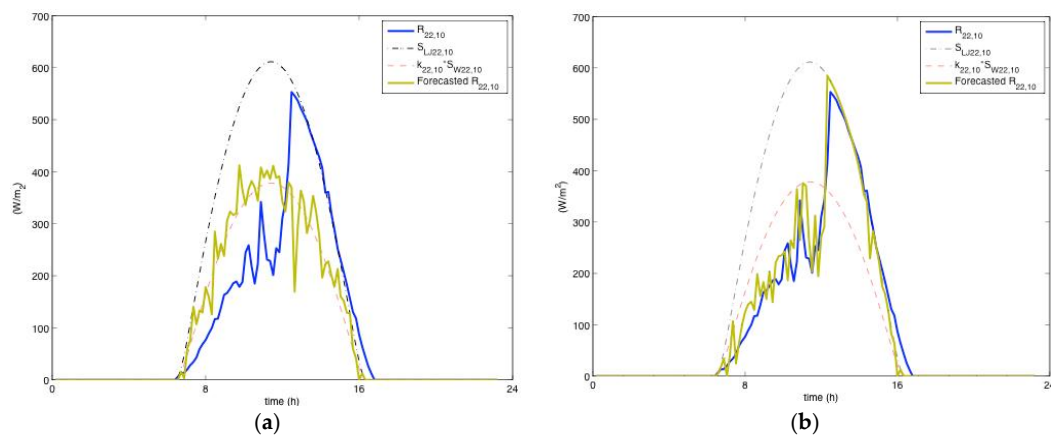


Figure 8. 22nd October: Measured solar irradiance, $R_{22,10}$, forecasted solar irradiance, $\hat{R}_{22,10}$, clear sky theoretical solar irradiance, $S_{LJ22,10}$, and product of the factor, $k_{22,10}$ by $S_{W22,10}$ versus the time: (a) pure statistical forecasting; and (b) forecasting accounting for weather predictions.

It is worth noting that the behavior of the analyzed 22nd of October helps us understand the reason why the values of the monthly modes are lower than those of the monthly means (Table 4). In fact, looking at Figure 7, it is possible to note that the majority of the days for all months are characterized by low error (resulting in a low value of the monthly mode) while only some days of all the months are characterized by very high error, resulting in an increased value of the monthly mean.

Results with a sensibly higher level of accuracy were obtained by applying the procedures described in Section 5, which is able to take into account weather predictions.

Figure 8b reports the results referring to the 22nd of October obtained from the study of weather predictions able to predict the two-time behavior of the day, with a level of approximation of \tilde{k} of $\pm 10\%$.

The figure gives an immediate visual idea of the improvement in forecasting solar irradiance through the hours of the day; the corresponding $MAE_{22,10}$ results reduced to 10% from the initial value 47% of Figure 8d). As for the other two metrics introduced, $MBE_{22,10}$ remains almost constant around 5% while $RMSE_{22,10}$ results reduced to 13% from 58%. By analysis of the figure, the strength of the procedures shown in Section 5 clearly appears to be predicting the different behavior during the day, not only in terms of radiation level but also in terms of variability of instantaneous solar irradiance. The first part of the day, in fact, is characterized by low values of mean solar irradiance (low $\tilde{k}_{1,22,10}$, see Equation (16)) and a significant variability value. The second part of the day, instead, is characterized by a regular profile of the solar irradiance with a high value of mean solar irradiance (high $\tilde{k}_{2,22,10}$, see Equation (16)). This behavior is quite well captured by the forecasting procedure, of which the accuracy is evident, especially in case of accurate weather predictions.

This shows the difference between a purely statistical method (without taking into account any weather predictions) and the procedure completed with information on weather conditions. The former seems to be a practical solution for simulating synthetic time series in the long run (e.g., for isolated

microgrids design purposes). The last seems to be a practical solution especially for those applications where information on instantaneous variations of solar irradiance is required (e.g., for day-ahead forecasts).

For comparative purposes, the results of the proposed model have been compared with the evaluations performed in [24], which refer to a benchmarking exercise organized within the framework of the European project “WIRE” [25] with the purpose of evaluating the performance of state-of-the-art models for short-term renewable energy forecasting. More specifically, 10 different solar power forecasting methods were applied to historical data for 2010 and 2011 and with reference to the suburbs of the city of Milan in Northern Italy and to the suburban area of Catania in Southern Italy. The location analyzed in this paper (Portici, Italy) is geographically situated between these two realities. Hence, though based on different data, this non-customized comparison can give an idea of the performance of the proposed model. For the evaluation in [24], among the metrics adopted, the MAE, normalized by the mean power (MP) measured during the test period, was reported. Table 7 reports the forecasting methods used by the participants in the project [25] and their error ranges in terms of MAE (%MP) for Milano (IT) and Catania (IT). Error ranges refer to minimum and maximum values regardless of the location of the best scores reported in Figures 7 and 8 of [24]. Mean normalized MAE (%MR) for the methods reported in Table 4 (evaluated for the city of Portici, IT) are reported in Table 8. The commensurateness of the proposed method with all of the analyzed approaches clearly appears.

Table 7. Normalized MAE (%MP) ranges for the methods applied in [24] to Milano (IT) and Catania (IT).

id	Method	Normalized MAE (%MP) Range	
		Min.	Max.
1	WIRE model data + linear regression (random forest), (Milano and Catania)	20%	32%
2	Own meteorological model + output correction using tendency of past production (Milano and Catania)	23%	40%
3	GFS + Model Output Statistics + conversion to power (Milano) WIRE data + conversion to power (Catania)	29%	38%
4	WIRE data + Support Vector Machines (Milano and Catania)	19%	65%
5	Linear regression between GHI and solar power (Milano and Catania)	27%	58%
6	WIRE data + ANN (Multilayer Perceptron with Standard Back Propagation and Logistic Functions), (Milano and Catania)	42%	56%
7	WIRE data + quantile regression to estimate clear sky production, irradiation and medium temperature + linear regression to explain the rate of clear sky production observed (Milano and Catania)	15%	30%
8	WIRE data + linear regression model (Milano and Catania)	23%	42%
9	Combination of WIRE data and WRF ARW model version 2.2.1 using initial and boundary conditions from NCEP GFS + Gaussian Generalized Linear Model (Catania)	17%	-

Table 8. Mean normalized MAE (%MR) for the methods reported in Table 4 applied to Portici (IT).

Method	Normalized MAE (%MR)
EXPERIMENTAL	24%
PARAMETRIC	25%
PERSISTENCE	48%

7. Conclusions

Radiation forecasting, accounting for daily and instantaneous variability, was pursued by means of a new bi-parametric model that builds on a model previously proposed by the same authors. The statistical model is developed with direct reference to the Liu-Jordan clear sky theoretical expression but is not bound by specific clear sky models; it accounts separately for the mean daily variability and for the variation of solar irradiance during the day by means of two corrective

parameters. This new proposal allows for a better understanding of the physical phenomena and improves the effectiveness of statistical characterization and subsequent simulation of the introduced parameters to generate a synthetic solar irradiance time series. Furthermore, the analysis of the parametric distributions that best fit the experimental distributions of the two-parameter data was developed by obtaining opportune parametric distributions or mixtures of more than one distribution. Finally, the model was further improved by including weather prediction information in the simulation and forecasting stage, thereby overcoming the limitations of a purely statistical approach.

The main outcomes are:

- The introduction of single parametric distributions and of mixtures of parametric distributions seems to offer, with reference to specific geographical areas, general models easy to handle in both data acquisition and subsequent simulation stages.
- The model is suitable for inclusion of weather prediction in the solar radiation forecast stage.

Acknowledgments: The author gratefully acknowledges financial support from Italian Ministry of University and Research under Grant PON03PE 00178 1.

Author Contributions: The authors contributed equally to this work.

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

Nomenclature

$m = 1, 2, \dots, 12$	month of the year
$j = 1, 2, \dots, N_m$	day of the m th month
N_m	number of days of the m th month
N_n	number of samples in a day
$N_{s,j}$	number of samples of the day with solar irradiance different from zero
$S_{LJ,j,m}(n\Delta t)$	Liu-Jordan theoretical solar irradiance at ground level at time $n\Delta t$ of the j th day of the m th month (W/m^2)
$W_{LJ,m}$	average daily solar radiation (or solar irradiation) of the theoretical Liu-Jordan model during the m th month (Wh/m^2)
$S_{W,j,m}(n\Delta t)$	expected irradiance at ground level based on historical data at time $n\Delta t$ of the j th day of the m th month (W/m^2)
C_m	monthly clear sky index for the m th month based on historical data
$W_{o,m}$	expected daily solar radiation during the m th month (Wh/m^2)
$R_{j,m}(n\Delta t)$	measured solar irradiance at time $n\Delta t$ of the j th day of the m th month (W/m^2)
$k_{j,m}$	ratio between the average daily $R_{j,m}$ and the average daily $S_{W,j,m}$
$\varepsilon_{j,m}(n\Delta t)$	absolute deviation of $R_{j,m}(n\Delta t)$ from $S_{LJ,j,m}(n\Delta t) * C_m * k_{j,m}$
$\varepsilon_{j,m}'^*(n\Delta t)$	deviation of $R_{j,m}(n\Delta t)$ from $S_{LJ,j,m}(n\Delta t) * C_m * k_{j,m}$ divided by $S_{LJ,j,m}(n\Delta t)(W/m^2)$
$\varepsilon_{j,m}''^*(n\Delta t)$	deviation of $R_{j,m}(n\Delta t)$ from $S_{LJ,j,m}(n\Delta t) * C_m * k_{j,m}$ divided by $C_m * S_{LJ,j,m}(n\Delta t)(W/m^2)$
K_m	random variable whose determinations are $k_{j,m}$
E_m	random variable whose determinations are $\varepsilon_{j,m}(n\Delta t)$
$E_m'^*$	random variable whose determinations are $\varepsilon_{j,m}'^*(n\Delta t)$
$\hat{R}_{j,m}(n\Delta t)$	simulated solar irradiance at time $n\Delta t$ of the j th day of the m th month (W/m^2)

References

1. Remund, J.; Perez, R.; Lorenz, E. Comparison of solar radiation forecasts for the USA. In Proceedings of the 23rd European Photovoltaic Solar Energy Conference, Valencia, Spain, 1–5 September 2008.
2. Pedro, H.T.C.; Coimbra, C.F.M. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Sol. Energy* **2012**, *86*, 2017–2028. [[CrossRef](#)]
3. Tovar, J.; Olmo, F.J.; Alados-Arboledas, L. One-minute global irradiance probability density distributions conditioned to the optical air mass. *Sol. Energy* **1998**, *62*, 387–393. [[CrossRef](#)]

4. Tovar, J.; Olmo, F.J.; Batlles, F.J.; Alados-Arboledas, L. Dependence of one-minute global irradiance probability density distributions on hourly irradiation. *Energy* **2001**, *26*, 659–668. [[CrossRef](#)]
5. Sim, M.L.; Kung, F.W.L.; Ghassemlooy, Z. Statistical analysis and modelling of one-minute global solar irradiance for a tropical country. In Proceedings of the 2nd International Symposium on Environment Friendly Energies and Applications (EFEA), Newcastle upon Tyne, UK, 25–27 June 2012; pp. 243–248.
6. Jurado, M.; Caridad, J.M.; Ruiz, V. Statistical distribution of the clearness index with radiation data integrated over five minute intervals. *Sol. Energy* **1995**, *55*, 469–473. [[CrossRef](#)]
7. Ibáñez, M.; Rosell, J.I.; Beckman, W.A. A bi-variable probability density function for the daily clearness index. *Sol. Energy* **2003**, *75*, 73–80. [[CrossRef](#)]
8. Ibáñez, M.; Beckman, W.A.; Klein, S.A. Frequency Distributions for Hourly and Daily Clearness Indices. *J. Sol. Energy Eng.* **2001**, *124*, 28–33. [[CrossRef](#)]
9. Skartveit, A.; Olseth, J.A. The probability density and autocorrelation of short term global and beam irradiance. *Sol. Energy* **1992**, *49*, 447–487. [[CrossRef](#)]
10. Woyte, A.; Belmans, R.; Nijs, J. Fluctuations in instantaneous clearness index: Analysis and statistics. *Sol. Energy* **2007**, *81*, 195–206. [[CrossRef](#)]
11. Suehrcke, H.; McCormick, P.G. The Frequency Distribution of Instantaneous Insolation Values. *Sol. Energy* **1988**, *40*, 413–422. [[CrossRef](#)]
12. Hollands, K.G.T.; Suehrcke, H. A three-state model for the probability distribution of instantaneous solar radiation, with applications. *Sol. Energy* **2013**, *96*, 103–112. [[CrossRef](#)]
13. Widén, J.; Carpman, N.; Castellucci, V.; Lingfors, D.; Olauson, J.; Remouit, F.; Bergkvist, M.; Grabbe, M.; Waters, R. Variability assessment and forecasting of renewables: A review for solar, wind, wave and tidal resources. *Renew. Sustain. Energy Rev.* **2015**, *44*, 356–375. [[CrossRef](#)]
14. Hoff, T.E.; Perez, R. Modeling PV fleet output variability. *Sol. Energy* **2012**, *86*, 2177–2189. [[CrossRef](#)]
15. Perez, R.; Hoff, T.E. Solar Resource Variability. In *Solar Energy Forecasting and Resource Assessment*; Elsevier: Philadelphia, PA, USA, 2013; pp. 133–148.
16. David, M.; Andriamasomanana, F.H.R.; Liandrat, O. Spatial and Temporal Variability of PV Output in an Insular Grid: Case of Reunion Island. *Energy Procedia* **2014**, *57*, 1275–1282. [[CrossRef](#)]
17. Engerer, N.A.; Mills, F.P. KPV: A clear-sky index for photovoltaics. *Sol. Energy* **2014**, *105*, 679–693. [[CrossRef](#)]
18. Langella, R.; Proto, D.; Testa, A. A statistical model of solar radiation daily variability. In Proceedings of the International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Durham, UK, 7–10 July 2014.
19. ENEA Fonti Rinnovabili: Atlante Italiano della Radiazione Solare. Available online: <http://www.solaritaly.enea.it/CalcComune/Calcola.php> (accessed on 10 March 2016).
20. Liu, B.Y.H.; Jordan, R.C. The interrelationship and characteristics distribution of direct, diffuse, and total solar radiation. *Sol. Energy* **1960**, *4*, 1–19. [[CrossRef](#)]
21. Liu, B.Y.H.; Jordan, R.C. A rational procedure for predicting the long term average performance of flat plate solar energy collectors. *Sol. Energy* **1965**, *7*, 53–74. [[CrossRef](#)]
22. European Solar Radiation Atlas. Available online: <http://www.soda-is.com/esra/reference.html> (accessed on 10 March 2016).
23. Jolliffe, I.T.; Stephenson, D.B. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*; John Wiley & Sons, Ltd.: London, UK, 2011.
24. Sperati, S.; Alessandrini, S.; Pinson, P.; Kariniotakis, G. The “Weather Intelligence for Renewable Energies” Benchmarking Exercise on Short-Term Forecasting of Wind and Solar Power Generation. *Energies* **2015**, *8*, 9594–9619. [[CrossRef](#)]
25. COST Action ES1002, WIRE: Weather Intelligence for Renewable Energies Official Website. Available online: <http://www.wire1002.ch/> (accessed on 10 March 2016).

