Research Article

Improved Synthesis of Global Irradiance with One-Minute Resolution for PV System Simulations

Martin Hofmann,[1] Stefan Riechelmann,[2] Cristian Crisosto,[2] Riyad Mubarak,[2] and Gunther Seckmeyer[2]

[1] *Valentin Software GmbH, Stralauer Platz 34, 10243 Berlin, Germany*
[2] *Institute for Meteorology and Climatology, University Hannover, Herrenhäuser Straße 2, 30419 Hannover, Germany*

Correspondence should be addressed to Martin Hofmann; martin.hofmann@valentin-software.com

High resolution global irradiance time series are needed for accurate simulations of photovoltaic (PV) systems, since the typical volatile PV power output induced by fast irradiance changes cannot be simulated properly with commonly available hourly averages of global irradiance. We present a two-step algorithm that is capable of synthesizing one-minute global irradiance time series based on hourly averaged datasets. The algorithm is initialized by deriving characteristic transition probability matrices (TPM) for different weather conditions (cloudless, broken clouds and overcast) from a large number of high resolution measurements. Once initialized, the algorithm is location-independent and capable of synthesizing one-minute values based on hourly averaged global irradiance of any desired location. The one-minute time series are derived by discrete-time Markov chains based on a TPM that matches the weather condition of the input dataset. One-minute time series generated with the presented algorithm are compared with measured high resolution data and show a better agreement compared to two existing synthesizing algorithms in terms of temporal variability and characteristic frequency distributions of global irradiance and clearness index values. A comparison based on measurements performed in Lindenberg, Germany, and Carpentras, France, shows a reduction of the frequency distribution root mean square errors of more than 60% compared to the two existing synthesizing algorithms.

1. Introduction

The efficiency of PV modules depends mainly on the irradiance, amongst other secondary effects such as module temperature [1, 2]. The nonlinear dependency of the module efficiency on the irradiance and the influence of temperature on the module efficiency require simulations with a high temporal resolution.

For the understanding of the dynamic interaction of PV generator, storage systems, loads, and grids on a worldwide scale, one-minute data series of high quality in terms of realistic variability and frequency distributions are a key factor. Simulating those systems with hourly averaged values neglects significant behavior patterns like short time power enhancements [3].

To illustrate the importance of one-minute data for the simulation of PV systems, a 1 kWp PV example system with

PV generator, DC/AC inverter, and grid is analyzed at the location of HTW Berlin, Germany. DC/AC inverters are used in grid-connected PV systems as power processing interface between the PV power source (DC) and the electric grid (AC). The output power is very sensitive to the temporal variability of the solar radiation which is highest during broken clouds.

In some important markets (e.g., Germany), PV systems can be affected by grid connection restrictions that define the maximum AC power output of the inverter as a percentage of the installed PV power on the DC side, where the usual limit is around 70% [4]. In Figure 1 the power output of the PV example system is shown in a one-minute temporal resolution (grey) and in an hourly averaged temporal resolution (blue) for a day with broken clouds. An energy yield loss of 7% is calculated when the 70% restriction is applied to the hourly averaged power output. When applying the restriction to the
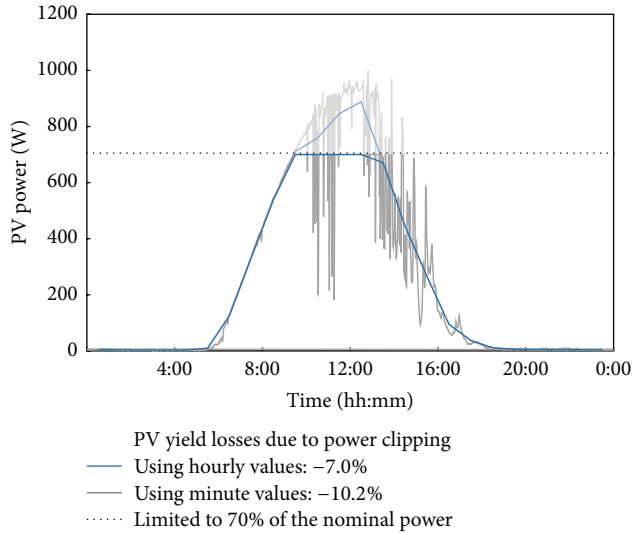
FIGURE 1: Power output of a 1 kWp PV system at HTW Berlin, Germany, on April 01, 2012, measured one-minute values (grey) and hourly averaged values (blue). The yield losses due to maximum power clipping (output power is cut above 700 W) are calculated.

FIGURE 2: Yearly energy yield losses of a 1 kWp PV system at HTW Berlin, Germany, for various inverter sizing factors (the relation between installed PV power on the DC side and nominal AC inverter output) and a maximum power clipping value of 70% (output power is cut at 70% of the installed DC power). Using hourly averaged values for the simulation of PV systems leads to a significant underestimation of the yearly yield losses. With hourly averaged values (grey), the total energy loss is at 1.3% while the more precise simulation with one-minute values (blue) returns a total energy loss of 3.9%. In addition, this figure illustrates that the optimal inverter sizing factor (here, 143%) for systems with maximum power clipping is the reciprocal of the clipping value (70%).

one-minute power output values, an energy yield loss of 10% is calculated.

Following Vanicek et al. in his contribution on the energy yield losses as a function of inverter dimensioning [3], we analyzed the dependency of energy yield losses due to maximum power clipping for PV inverters. Figure 2 shows that these losses are dependent on the inverter sizing factor as well and increase significantly when using one-minute instead of hourly averaged time series. In sum, energy losses due to inverter undersizing and maximum power clipping add up to a constant value within the inverter dimensioning range until the reciprocal of the power clipping value is reached (143%). This threshold marks the optimum inverter sizing factor for PV inverters with maximum power clipping, since losses will not decrease when using a larger inverter. With hourly averaged values (grey), the total energy loss is at 1.3% while the more precise simulation with one-minute values (blue) returns a total energy loss of 3.9%. These examples indicate that the use of hourly averaged irradiance datasets can result in falsified yield predictions.

While there exist several commercial providers and free sources of meteorological data in a resolution of one hour (e.g., Meteotest, SolarGIS, and TMY), covering nearly the whole earth, the availability of measured irradiance data with a resolution of less than an hour is very limited. This limited availability leads to the necessity to synthesize one-minute time series from hourly averaged data.

Several algorithms were developed in the past in order to synthesize one-minute global irradiance datasets with realistic variability and frequency distributions from hourly averaged datasets. The supposedly most established algorithms were developed by Aguiar and Collares-Pereira [5, 6], Skartveit and Olseth [7], and Glasbey [8]. Like many similar algorithms, t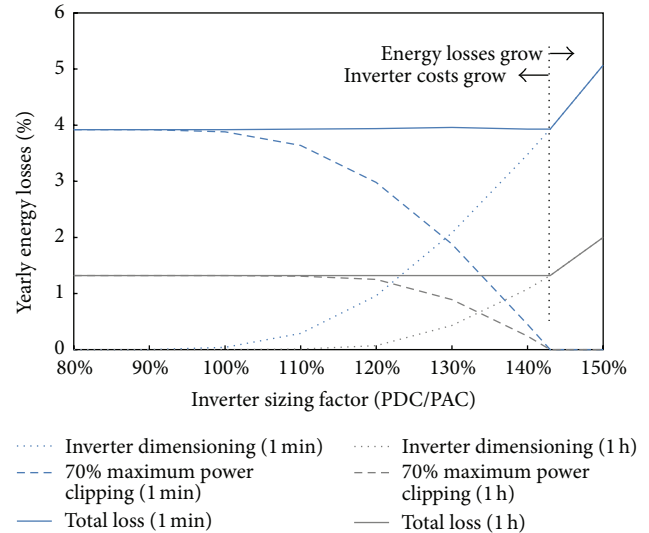he aim of those approaches is the reproduction of the characteristic frequency distributions of the solar irradiance or the clearness index k_t, which is a measure for atmospheric transmission.

The contribution of Aguiar and Collares-Pereira was originally designed for the generation of hourly averaged time series with daily averages as input. It is based on the modeling of probability densities as Gaussian functions that depend on the clearness index k_t. Skartveit and Olseth focused on the modeling of frequency distributions of global and direct irradiance, depending on intrahour and interhour irradiance variability, while using first-order autocorrelation for the generation of the actual time series. Glasbey proposed nonlinear autoregressive time series generation with joint marginal distributions as multivariate Gaussian mixtures. The estimation of probability density distributions of the irradiance has recently been investigated by Voskrebenzev et al. [9].

Other important contributions to this topic were provided by Assunção et al. [10] with investigations on the dependency of k_t from the air mass and by Tovar et al. [11] with the analysis of the relation of hourly averaged *[clearness indices]* to one-minute clearness indices.

However, current algorithms only insufficiently withstand the validation against measurement values, since they underestimate irradiance enhancements caused by broken

clouds, overestimate mid irradiance values, and provide one-minute time series with a variability that is too high.

Therefore, we developed an improved algorithm capable of synthesizing one-minute global irradiance time series based on hourly averaged global irradiance. The algorithm takes three different weather conditions (cloudless, broken clouds and overcast) into consideration. We show that the improved algorithm exceeds the performance of the Aguiar and the Skartveit algorithm in terms of temporal variability and characteristic frequency distributions for the calculation of short-term global irradiance at two exemplary PV installation locations.

2. Measurement Data and Methodology

The new algorithm consists of two parts. The first part comprises a data preparation process that categorizes the input dataset and produces transition probability matrices (TPM) for three weather conditions: cloudless, broken clouds and overcast. The preparation process has to be executed only once.

The input dataset used for the initialization consists of global irradiance measurements conducted by the Baseline Surface Radiation Network (BSRN), featuring more than 50 locations all over the world with up to 20 years of measurements. The BSRN database is updated continuously with new measurement data; in this study we used a snapshot of May 2013. A subset of these data, one-minute global irradiance measurements performed in Lindenberg, Germany (2005), and Carpentras, France (2001), is used for the model validation.

The second part is the synthesis process for one-minute time series from hourly averaged time series. The required input of this process only consists of the prepared set of TPM and the hourly averaged time series of global irradiance that is to be disaggregated. The core of the process is based on Markov chains [12, 13], utilized in a similar way by McCracken [14].

The central idea in both parts of the new algorithm is the classification of weather situations by the temporal feature of the clearness index. In the first part, the preparation process, the BSRN dataset is split into three individual datasets corresponding to three weather conditions: cloudless, broken clouds and overcast. Each subset is then processed separately and transformed into a transition probability matrix. In the second part, the synthesis process, each daily dataset of the hourly averaged input values is categorized as well and processed according to their weather category. As a consequence, the main process steps of the new algorithm are only depending on those weather categories, in disregard of specific location information.

This leads to the advantage that the algorithm can be applied to hourly averaged datasets of arbitrary locations. Furthermore, the only required input is the hourly averaged datasets, once the TPM are created. Hence, the new algorithm combines aspects of existing work on this subject with a universally applicable method for the synthesis of one-minute time series from hourly averaged values.

2.1. Classification of Weather Condition by the Clearness Index. The determination of predominant weather conditions is needed in both steps of the presented algorithm. The weather conditions are determined by the calculation of the clearness index k_t. The clearness index is defined as the ratio of measured global irradiance E_{measured} at Earth's surface and the irradiance calculated for cloudless conditions at the particular measuring site, denoted by clear sky irradiance E_{clear}:

$$k_t = \frac{E_{\text{measured}}}{E_{\text{clear}}}. \tag{1}$$

The calculation of the clear sky irradiance has a significant influence on the k_t index. A modification of Bourges' calculation [15] is used in this work, since it provided the best results for all analyzed locations:

$$E_{\text{clear}} = 0.78 E_{\text{ext}} \sin \left(\gamma_S \right)^{1.15}, \tag{2}$$

where γ_S is the elevation of the sun and E_{ext} is the extraterrestrial irradiance. The extraterrestrial irradiance E_{ext} was calculated using Maxwell's approach [16], whereas the elevation of the sun γ_S was modelled by the algorithm of Reda and Andreas [17] from NREL.

The predominant weather condition on a particular day results in a characteristic temporal pattern of k_t that can be used to categorize the day into one of the three classes. The detection algorithm of the weather condition is based on the daily average of hourly averaged k_t values $\overline{k}_{t,\text{day}}$ and the variability during a day $\widetilde{k}_{t,\text{day}}$:

$$\overline{k}_{t,\text{day}} = \frac{1}{n} \sum_{i=1}^{24} k_{t,i}, \qquad \widetilde{k}_{t,\text{day}} = \frac{1}{n} \sum_{i=2}^{24} \left| k_{t,i} - k_{t,i-1} \right|, \tag{3}$$

where n is the number of hours where global and clear sky irradiance is above 0 W/m².

Table 1 gives an overview about the three weather classes and their detection conditions. An example for the classification is shown in Figure 3 for some days in August 2005 in Lindenberg, Germany. For a better visualization we fall back on one-minute values here, whereas it is to be noted that the detection is based on hourly averaged values of the clearness index k_t, because these values form the input of the synthesis algorithm. The classification conditions are visualized as well in Figure 4 for an example dataset of Lindenberg, Germany.

2.2. Transition Probability Matrices. For each class that represents a specific weather situation, matrices of transition probabilities (TPM) are created. The TPM contain information on how probable the switch is from one specific k_t at time i to another value at the time $i + 1$. To create those matrices, diurnal courses of measured one-minute values of equal weather class, independent of their location, are analyzed and converted into a common matrix. The frequency of every possible transition in the measured data is registered and afterwards normalized to obtain the transition probabilities. Therefore, a TPM contains all probabilities of the change of

TABLE 1: Overview of the three weather classes and their detection conditions.

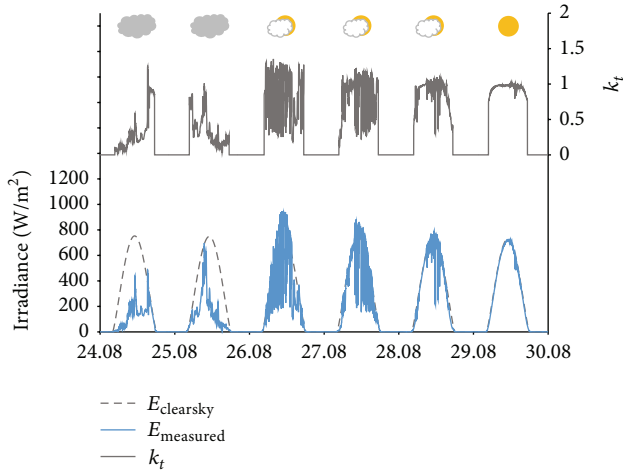| Weather class	Condition
Overcast	$0.6 - \overline{k}_{t,\text{day}} > \widetilde{k}_{t,\text{day}}$
Cloudless	$-0.72 + 0.8\overline{k}_{t,\text{day}} \geq \widetilde{k}_{t,\text{day}}$
Broken clouds	Otherwise

FIGURE 3: Example results of the weather category detection algorithm based on k_t patterns for 6 subsequent days from 24.08.2005 to 30.08.2005 in Lindenberg. Clear sky irradiance (dotted grey) and the clearness index (grey line on top) in comparison to the measured global irradiance (blue line, bottom). Visualization is done with one-minute values whereas the detection is based on hourly averaged values.
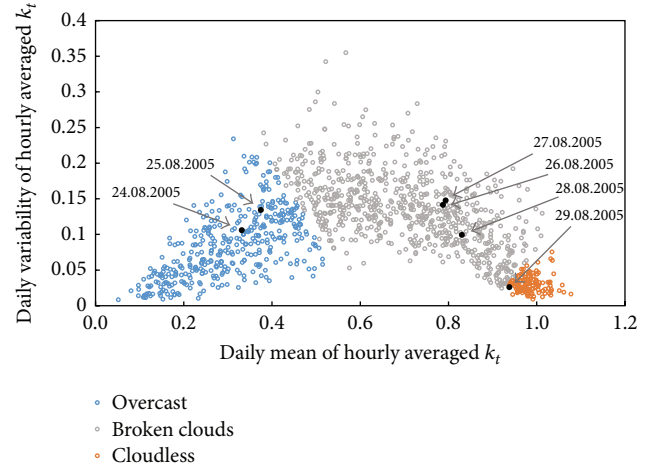

FIGURE 4: Visualization of the classification of weather conditions by k_t patterns using measurement values from Lindenberg, Germany. Cloudless days feature high daily means $\overline{k}_{t,\text{day}}$ with only little daily variability $\widetilde{k}_{t,\text{day}}$. Days with overcast sky can be characterized by low daily means and low to mid variability. Days with broken clouds feature a high daily variability and mid to high daily means. The black points refer to the example days of Figure 3.

a specific value of $k_{t,i}$ to $k_{t,i+1}$ from one minute to the next under a specific weather condition. An example for a TPM of broken clouds weather condition is given in Table 2. In this case, the probability of k_t to change from 0.1 to 0.09 during one minute is 17.6%, the probability to stay the same is 53.2%, and the probability to change from 0.1 to 0.01 is 0%.

The excerpt of a TPM shown in Table 2 is an example of how such transition probability matrices are structured. The actual values of the TPM however are subject to the underlying dataset that is used to create those matrices. In this study we will use different subsets of the BSRN databases for the creation process, depending on the dataset we use for validation. The validation dataset is omitted from the dataset for the TPM creation process to avoid self-reference. Hence, the resulting values in the matrices may vary, whereas the presented method to create the matrices is universal. For this reason we refrain from listing all 200×200 TPM in this study.

Since the TPM are created using real weather data in one-minute resolution, each measured irradiance within a given time interval leaves a fingerprint in a TPM. Hence, the spatial and temporal validity of the algorithm is increasing with the number of input datasets. As of May 2013, the BSRN comprises more than 6900 irradiance measurement months distributed globally, which is equal to more than 200 000 measurement days in one-minute resolution that leave their fingerprint in the TPM. The influence of the number of input

data on the synthesis quality is referred to in Results section as well.

2.3. Generation of k_t Sequences with Markov Chains.
To generate one-minute values from hourly averaged sequences of the global irradiance, the weather condition of the day in question is detected at first. Depending on the weather condition the correspondent k_t-TPM is chosen.

The actual generation of the one-minute values is conducted with the help of the so-called discrete-time Markov chains (DTMC). DTMC is a state-based process for the modelling of real-world events. In the first order, the process is memory-less, so that the next state only depends on the current state [12, 13].

To determine the successor $k_{t,i+1}$ of a specific k_t value $k_{t,i}$ at a given point in time i, the probabilities belonging to $k_{t,i}$ are cumulated. Then, a Markov number between 0 and 1 is generated and inserted as a threshold value into the cumulated probability function. The point at which the probability function is bigger than the Markov number for the first time is defined as $k_{t,i+1}$. The process continues in the same manner and generates a chain of 60 k_t values per hour. From these k_t sequences, the global irradiance for every point in time can be calculated with the help of the clear sky irradiance:

$$E_{\text{generated},i} = k_{\text{t},i} \cdot E_{\text{clear sky},i}. \tag{4}$$

This process is repeated until the mean value of the generated one-minute values equals the hourly averaged input value

TABLE 2: Excerpt of an example TPM for broken clouds weather condition. For each k_t value at time i (rows), the probability of a switch to another k_t at time $i + 1$ (columns) is given. The k_t values range from 0 to 2.

$k_{t,i}$	$k_{t,i+1}$											
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	$\cdots$
0	0	0	0	0	0	0	0	0	0	0	0	
0.01	0	0.82927	0.17073	0	0	0	0	0	0	0	0	
0.02	0	0.10345	0.72414	0.17241	0	0	0	0	0	0	0	
0.03	0	0	0.06897	0.76724	0.15517	0	0.00862	0	0	0	0	
0.04	0	0	0.00709	0.12057	0.70922	0.14894	0.01418	0	0	0	0	
0.05	0	0	0	0	0.07004	0.75875	0.15564	0.01167	0.00389	0	0	
0.06	0	0	0	0	0.01136	0.14773	0.64394	0.14773	0.03788	0.00758	0.00379	
0.07	0	0	0	0	0	0.0084	0.19328	0.53361	0.2395	0.02521	0	
0.08	0	0	0	0	0.0059	0	0.0236	0.17109	0.57817	0.17404	0.03245	
0.09	0	0	0	0	0	0	0	0.03378	0.19932	0.46959	0.21284	
0.1	0	0	0	0	0	0	0	0	0.02067	0.17571	0.5323	
$\cdots$												

E_{hour} with desired accuracy δ. If necessary, the values are scaled as well:

$$\left| \frac{E_{\text{hour}}}{\sum_{i=1}^{60} E_{\text{generated},i}} - 1 \right| < \delta,$$

$$E_{\text{gen., scaled},i} = \frac{E_{\text{hour}}}{\sum_{m=1}^{60} E_{\text{generated},m}} E_{\text{generated},i}.$$

(5)

3. Results

In the following section the new algorithm is validated with measurement data and compared to the algorithms by Aguiar and Skartveit. The result comparison is conducted for two exemplary datasets of one year at two different locations: Lindenberg, Germany, 2005, and Carpentras, France, 2001. Both datasets are taken from the BSRN database. To avoid self-reference in the presented results, the creation process of the TPM excludes all measurement data of the respective location.

First, the results are presented on the basis of diurnal courses to assess the temporal variability, afterwards in the form of frequency distributions. In addition we provide a table with comparative uncertainties.

When assessing the temporal variability of synthetized one-minute values, the results for days with broken clouds and overcast skies are more important, since the simulation of sunny days is not difficult. In Figure 5 the measured course of the global irradiance (black (a)) is displayed in comparison to the temporal course of the values generated with the new algorithm (blue (b)) and the algorithms by Aguiar (c) and Skartveit (d) for an overcast day.

Although the exact occurrence of irradiance peaks in the modelled time series may differ from the measured time series, the variability of the values modelled with the new algorithm agrees with measured values to a very high degree. The mean variability of irradiance changes from one minute to the next is 7.0 W/m² for measured time series, whereas it is 8.2 W/m² for the data modelled with the new algorithm

in the example dataset of Figure 5. With n being the number of minutes of a day (1440), the mean variability of the global irradiance is calculated as follows:

$$\overline{E_{\text{var}}} = \frac{1}{n} \sum_{i=2}^{n} |E_i - E_{i-1}|.$$

(6)

The methods of Aguiar and Skartveit lead to higher mean variability values of 13.1 W/m² and 16.6 W/m², respectively. Scientists of the Sandia National Laboratories as well refrain from using these algorithms for this reason:

> *Without an adequate method to account for autocorrelations (of relatively high order) in the one-minute time series of clearness index, simulations using these distribution forms would likely prove too variable, as we found for simulations using Glasbey's model, and as we suspect would have resulted using the model of Skartveit and Olseth [18].*

A more complete picture of the variability of solar irradiance can be obtained by analyzing the frequency of its gradients over a whole year. The gradients, in this case the absolute difference of the irradiance values of one minute to the next for the measured data and model data, are calculated and transferred into frequency plots. Figure 6 shows the frequency of irradiance gradients for Lindenberg, 2005, whereas the data for Carpentras, 2001, is displayed in Figure 7.

In both cases, the frequency distribution of the data modelled by the algorithms of Aguiar and Skartveit, respectively, shows significant overestimations for the gradient range from 10 to 100 W/m², while the new algorithm is able to produce irradiance values that feature a similar frequency distribution in this range. For gradients of less than 10 W/m² the data modelled by all algorithms show similar deviations from the measured data. For gradients of more than 100 W/m², the new algorithm and the approach of Skartveit display similar
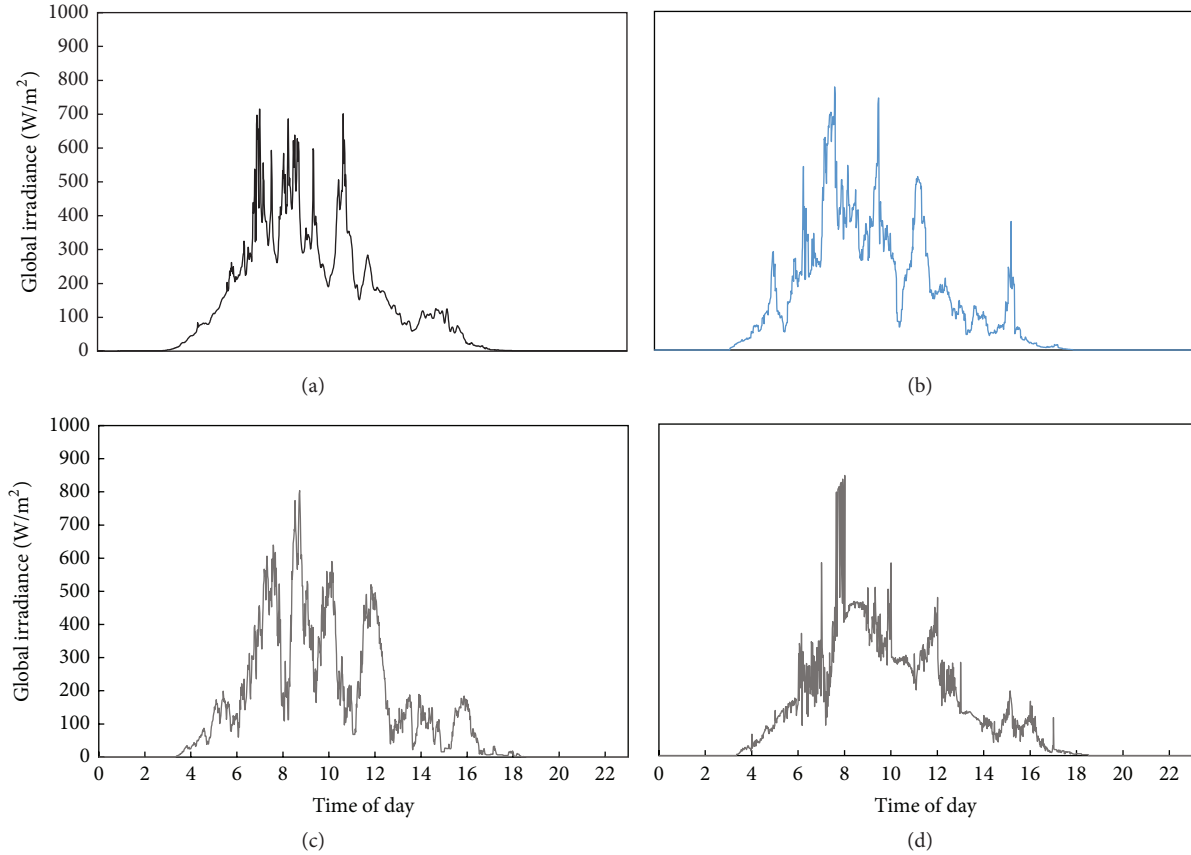
FIGURE 5: Lindenberg, May 14, 2005. The temporal course of the measured global irradiance (a) on a day that was rated as a day with overcast sky is compared to values generated by the new algorithm (b), the algorithm by Aguiar (c), and the algorithm by Skartveit (d). The mean variability, that is, the mean irradiance change from one minute to the next, of the measured irradiance of 7.0 W/m^2 shows good congruence with the new algorithm (8.2 W/m^2), while the usage of the algorithms by Aguiar and Skartveit leads to higher variability values of 13.1 W/m^2 and 16.6 W/m^2.

quality, whereas the algorithm of Aguiar shows significant underestimations for both locations.

For the transfer into deviation indicators, the deviations of the modelled data from the measured ones for each irradiance value are squared, weighed by its frequency, and summed up. The frequency weight f_i is added in order to obtain information about the energetic relevance of each irradiance gradient. For the calculation of root mean square errors, these sums are then divided by the number of gradient steps and the square root is applied. Table 3 shows the results for all three analyzed models. In accordance with the visual impression of the frequency plots in Figures 6 and 7, the RMSE values for the new model presented in this study are significantly smaller than the RMSE values of the other two models by Aguiar and Skartveit:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f_i\left(x_{\text{Model},i} - x_{\text{Measurement},i}\right)\right)^2}. \quad (7)$$

Since the frequency distributions of the global irradiance and the k_t values are more reliable indicators for the applicability to simulations of photovoltaic systems, they are displayed

TABLE 3: Root mean square errors (RMSE) of the frequency distributions of irradiance gradients of modelled data versus measurement in W/m^2. The new model is able to produce significantly smaller values of RMSE than the models of Aguiar and Skartveit for both locations, Lindenberg, 2005, and Carpentras, 2001.

Model	Lindenberg	Carpentras
Aguiar	8131	8541
Skartveit	4758	5112
New	2787	3218

in Figures 8, 9, 10, and 11. Measured values (black) are compared to values calculated by the conventional algorithms by Aguiar and Skartveit (grey dotted and solid), as well as to the new algorithm presented in this study (blue). Each of these distributions is calculated from values of one whole year.

For the generation of those figures, measured one-minute values were averaged to hourly means, which were then fractionized again using the new improved algorithm as
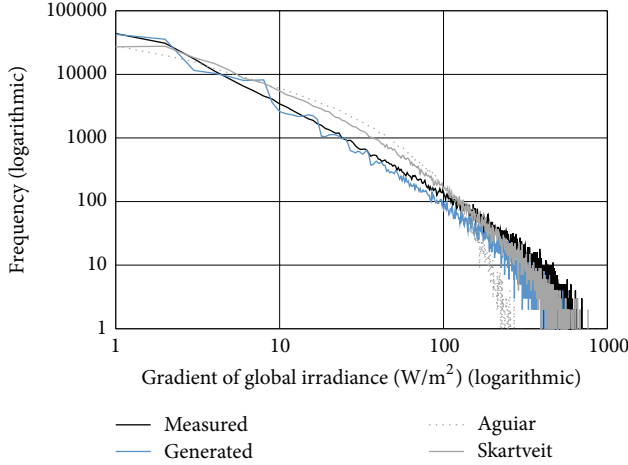
FIGURE 6: Frequency of gradients of the global irradiance in Lindenberg, Germany (2005). The model quality in lower gradient ranges of up to 10 W/m² is similar in all models. In the range of 10 to 100 W/m², significant deviations can be detected for the models of Aguiar and Skartveit (grey), whereas the new method (blue) shows good congruence. For gradient values of more than 100 W/m², the model of Aguiar underestimates the frequency significantly, while the new method and the method of Skartveit feature similar frequency values compared to the measurement data (black).
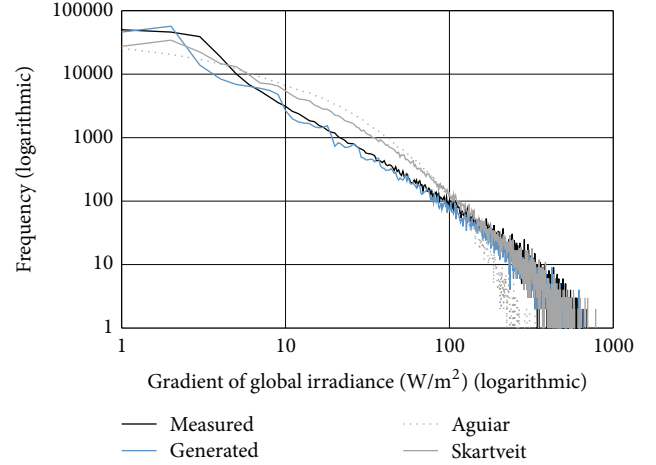

FIGURE 7: Frequency of gradients of the global irradiance in Carpentras, France (2001). As in Figure 6, the differences of the models for gradient values of less than 10 W/m² are comparable. In the range of 10 to 100 W/m², the new model (blue) shows a better congruence with the measured data (black) than the models of Aguiar and Skartveit (grey), while only the model of Aguiar fails to produce good congruence for gradient values of more than 100 W/m².

well as the approaches of Aguiar and Skartveit. The figures show how often a specific irradiance value occurs in a year. The maximum at high irradiance values represents clear sky situations, while the second maximum at lower values is evoked by skies covered by clouds. Hence, the maximum at high irradiance values is considerably more pronounced at sunnier locations than at locations with very variable weather.

It can be seen that the new algorithm is reproducing the frequency distributions of the global irradiance much better than the conventional approaches. Mid irradiance values are not overestimated, and a good modelling quality is present at high irradiance values. However, very high irradiances above 1100 W/m² are slightly overestimated.

If those frequency distributions are looked at in the form of the clearness index k_t, the problems of the existing algorithms become equally apparent (see Figures 10 and 11). With the improved algorithm the k_t distributions can be reproduced very well, and the typical bimodal character of the distribution is modelled very precisely for cloudy locations (Lindenberg) as well as for sunnier locations (Carpentras) with a pronounced clear sky peak of a k_t value near 1. The practical relevance of these effects was demonstrated with the help of the introductory example of the maximum power clipping at 70%.

These visual impressions give an indication, but an analysis of the uncertainty can be used for quantitative assessment. Table 4 lists root mean square errors (RMSE) for all distribution diagrams shown in Figures 8–11. For each irradiance or clearness index class i the modelled distributions are compared to the measured data, and the deviations are squared and summed up for the whole range

TABLE 4: Root mean square error (RMSE) values comparing the frequency distributions of the existing and the new algorithms with measured data. Smaller values of RMSE denote better congruence of the frequency distributions of modelled one-minute values with measured values.

Model	RMSE of irradiance in %		RMSE of k_t in counts	
	Lindenberg	Carpentras	Lindenberg	Carpentras
Aguiar	0.530	0.549	596	801
Skartveit	0.684	0.575	862	962
New	0.210	0.237	207	248

and then divided by the number of classes n. The square root of this value gives the RMSE listed:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{\text{Model},i} - x_{\text{Measurement},i}\right)^2}. \quad (8)$$

The RMSE of both the irradiance and the clearness index distributions can be considerably decreased with the new algorithm compared to the conventional ones. In the case of Carpentras both distribution RMSE can be reduced between 24 % and 35 %, in case of Lindenberg between 31 % and 43 %.

To analyze the influence of the amount of input data for the TPM on the synthesis quality of the algorithm, the creation process of the TPM is varied as follows.

First, the algorithm is processed three times with its original setup, which includes all TPM except the ones from the respective location, to estimate the influence of the random Markov number generator on the RMSE range. Second, only TPM of the respective location are used. In a third iteration, the only measurement values included in the creation process are taken from BSRN stations that are located in the same
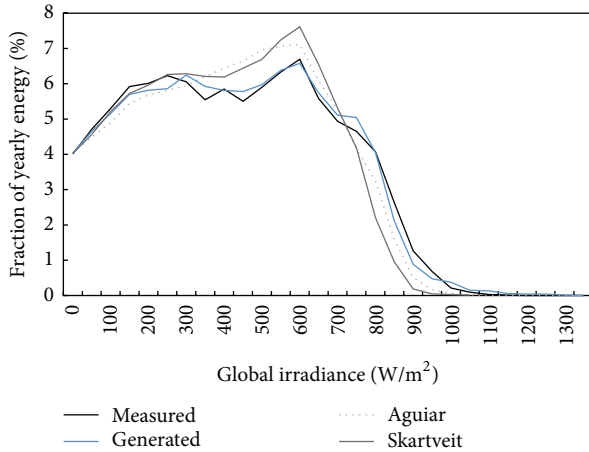
FIGURE 8: Frequency distributions of measured global irradiance for Lindenberg (Germany, 2005) against values generated by different algorithms. Mid values are slightly overestimated, and high values are underestimated by the existing models (grey dotted), resulting in RMSE of 0.530% for Aguiar and 0.684% for Skartveit. The modelling quality of the new method (blue) does not overestimate mid irradiance values and shows only little underestimation of high irradiance values. The new RMSE can be reduced to 0.210%.
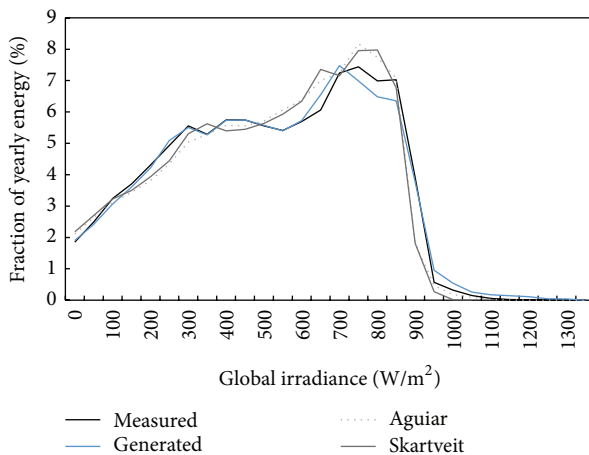

FIGURE 9: Frequency distributions of measured global irradiance for Carpentras (France, 2001) against values generated by different algorithms. For locations with higher yearly global irradiation, the high irradiance peak grows. Mid values are slightly overestimated, and high values are underestimated by the existing models (grey dotted), resulting in RMSE of 0.549% for Aguiar and 0.575% for Skartveit. The modelling quality of the new method (blue) does not overestimate mid irradiance values and shows only little underestimation of high irradiance values. The new RMSE can be reduced to 0.237%.

climate zone as per the definition of Köppen [19]. Current data published by Rubel and Kottek [20] is taken to assign the locations to climate zones. Lindenberg is located in the climate zone *Cfb,* which mainly comprises Western Europe. In the BSRN dataset there are another seven locations in this climate zone: Cabauw (the Netherlands), Camborne and Lerwick (Great Britain), Cener (Spain), Lauder (New Zealand),
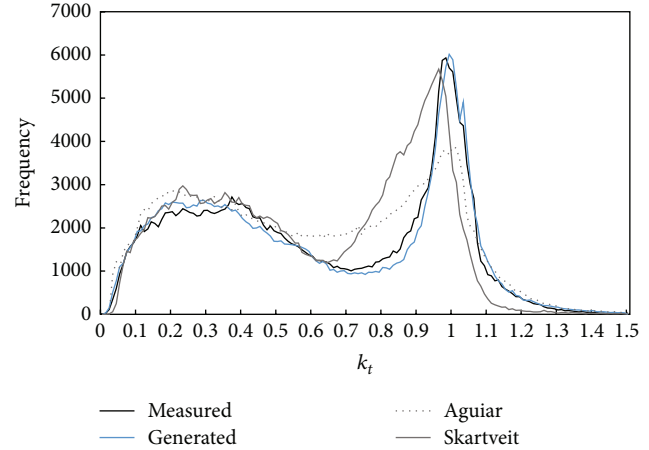

FIGURE 10: Frequency distributions of the clearness index k_t for Lindenberg (Germany, 2005). With the new method, the RMSE between measurement and synthesis can be reduced significantly (RMSE = 596 and 862 counts for Aguiar and Skartveit and RMSE = 207 counts for new method).
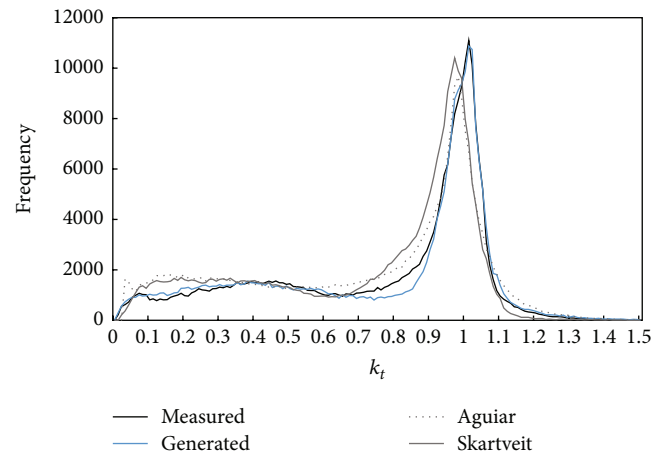

FIGURE 11: Frequency distributions of the clearness index k_t for Carpentras (France, 2001). For locations with higher yearly global irradiation, the second peak (clear sky) of the distribution grows. With the new method, the RMSE between measurement and synthesis can be reduced significantly (RMSE = 801 and 962 counts for Aguiar and Skartveit and RMSE = 248 counts for new method).

Palaiseau (France), and Payerne (Switzerland). According to Rubel and Kottek, Carpentras lies in climate zone *Csa*, but unfortunately there is no other location of this climate zone in the BSRN dataset. So this third iteration is conducted for Lindenberg only. The fourth iteration comprises the usage of all available TPM, this time including the TPM of Lindenberg and Carpentras.

The synthesis of one-minute irradiance values is now repeated with all varied TPM. The RMSE values are determined according to the previous chapter. Table 5 compares the error values of the variations with the original version of the process.

TABLE 5: Comparison of synthesis quality of the new algorithm as a function of input data for the locations of Lindenberg, Germany, 2005, and Carpentras, France, 2001.

Variation	RMSE of irradiance in %		RMSE of k_t in counts	
	Lindenberg	Carpentras	Lindenberg	Carpentras
All except own (1)	0.210	0.237	207	248
All except own (2)	0.244	0.174	210	273
All except own (3)	0.235	0.217	204	239
Own TPM only	0.232	0.199	202	279
Cfb only	0.193	—	315	—
All TPM	0.253	0.186	204	254

The repetition of the synthesis process with the original setup (all TPM except own 1–3) demonstrates the RMSE range that can be expected due to the random nature of the Markov number generator. The interesting aspect of the various TPM modifications (own TPM only, *Cfb* TPM only, and all TPM) is that the resulting RMSE mostly lie well within the natural RMSE range of the original algorithm. In other words, the synthesis quality remains approximately the same, whether the algorithm uses only data of the respective location or all globally available data except those from the respective location. By classifying the weather situation on a daily level, the influence of location specific weather phenomena is reduced at the best. This implies that the presented algorithm is location-independent and can be applied to every location worldwide.

4. Conclusions

An improved method for synthesizing one-minute time series of global irradiance has been presented that was developed on the basis of a large worldwide measurement dataset. It combines the advantages of conventional algorithms and adds new elements like the differentiation of weather conditions. It could be demonstrated that with the new approach it is possible to synthesize one-minute values of high statistical quality and realistic temporal variability. The independence on the location has been shown for selected cases. Such an independence would allow synthesizing one-minute time series for any location.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] R. Haselhuhn, U. Hartmann, and P. Vanicek, "Uncertainty in yield prediction—what are the causes, how can they be reduced?" in *Proceedings of the 25th European Photovoltaic Solar Energy Conference and Exhibition*, pp. 4722–4725, 2010.

[2] B. Herteleer, J. Cappelle, and J. Driesen, "Quantifying low-light behaviour of photovoltaic modules by identifying their irradiance-dependent efficiency from data sheets," in *Proceedings of the 27th European Photovoltaic Solar Energy Conference and Exhibition*, pp. 3714–3719, 2012.

[3] P. Vanicek, G. Karg, and R. Haselhuhn, "Quality of the calculation of inverter losses with standard simulation programs," in *Proceedings of the 26th European Photovoltaic Solar Energy Conference and Exhibition*, pp. 3772–3775, 2011.

[4] EEG, *Act on Granting Priority to Renewable Energy Sources*, Renewable Energy Sources Act, EEG, 2012.

[5] R. J. Aguiar, M. Collares-Pereira, and J. P. Conde, "Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices," *Solar Energy*, vol. 40, no. 3, pp. 269–279, 1988.

[6] R. Aguiar and M. Collares-Pereira, "TAG: a time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation," *Solar Energy*, vol. 49, no. 3, pp. 167–174, 1992.

[7] A. Skartveit and J. A. Olseth, "The probability density and autocorrelation of short-term global and beam irradiance," *Solar Energy*, vol. 49, no. 6, pp. 477–487, 1992.

[8] C. A. Glasbey, "Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions," *Journal of the Royal Statistical Society, Series C: Applied Statistics*, vol. 50, no. 2, pp. 143–154, 2001.

[9] A. Voskrebenzev, S. Riechelmann, A. Bais, H. Slaper, and G. Seckmeyer, "Estimating probability distributions of solar irradiance," *Theoretical and Applied Climatology*, 2014.

[10] H. F. Assunção, J. F. Escobedo, and A. P. Oliveira, "Modelling frequency distributions of 5 minute-averaged solar radiation indexes using Beta probability functions," *Theoretical and Applied Climatology*, vol. 75, no. 3-4, pp. 213–224, 2003.

[11] J. Tovar, F. J. Olmo, F. J. Batlles, and L. Alados-Arboledas, "Dependence of one-minute global irradiance probability density distributions on hourly irradiation," *Energy*, vol. 26, no. 7, pp. 659–668, 2001.

[12] Markov, "Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga," *Izvestiya Fiziko-Matematicheskogo Obschestva pri Kazanskom Universitete, Seriya 2*, vol. 15, pp. 135–156, 1906.

[13] R. A. Howard, *Dynamic Probabilistic Systems*, vol. 1, John Wiley & Sons, New York, NY, USA, 1971.

[14] D. A. McCracken, *Synthetic high resolution solar data [M.S. thesis]*, Department of Mechanical Engineering, University of Strathclyde, 2011.

[15] G. Bourges, "Reconstitution des courbes de fréquence cumulées de l'irradiation solaire globale horaire reçue par une surface plane," Report CEE 295-77, ESF of Centre d'Energétique de l'Ecole Nationale Supérieure des Mines de Paris, tome II, Paris, France, 1979.

[16] E. L. Maxwell, "A quasi-physical model for converting hourly global horizontal to direct normal insolation," Tech. Rep. SERI/TR-215-3087, Solar Energy Institute, Golden, Colo, USA, 1987.

[17] I. Reda and A. Andreas, "Solar position algorithm for solar radiation applications," *Solar Energy*, vol. 76, no. 5, pp. 577–589, 2004.

[18] J. S. Stein, C. W. Hansen, A. Ellis, and V. Chadliev, "Estimating annual synchronized 1-min power output profiles from utility-scale PV plants at 10 locations in Nevada for a solar grid integration study," in *Proceedings of the 26th European Photovoltaic Solar Energy Conference and Exhibition*, pp. 3874–3880, 2011.

[19] W. Köppen, "Klassifikation der Klimate nach Temperatur, Niederschlag und jahreslauf," *Petermanns Geographische Mitteilungen*, vol. 64, pp. 193–203, 243–248, 1918.

[20] F. Rubel and M. Kottek, "Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification," *Meteorologische Zeitschrift*, vol. 19, no. 2, pp. 135–141, 2010.