Solar Energy Advances 4 (2024) 100076

Contents lists available at ScienceDirect

Solar Energy Advances

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

A minute-resolution downscaling algorithm for high-resolution global irradiance time series

Diamantis Almpantis [a,*], Henrik Davidsson [b], Martin Andersson [a]

[a] Department of Energy Sciences, Lund University, Lund, 221 00, Sweden
[b] Division of Energy and Building Design, Lund University, Lund, 221 00, Sweden

ARTICLE INFO

Keywords:
Global horizontal irradiance
Solar integration
Downscaling synthesis
Minute time resolution
Data validation
Solar-to-hydrogen coupling

ABSTRACT

Numerous studies have demonstrated the significant impact of the resolution of solar irradiation data on the outcomes of hourly production models. Accurate integration of photovoltaic (PV) systems sometimes demands a high-resolution global horizontal irradiance (GHI) time series to capture the rapid fluctuations in PV power output induced by swift irradiance changes. Most of the available databases provide data at hourly resolution, leading to a lack of accuracy in PV simulations. Those existing hourly averages of global horizontal irradiance in open sources fail to represent this volatility adequately, especially when PV systems are coupled with fast ramp rate technologies. In the present work, an easy-to-use algorithm is implemented to synthesize high-resolution GHI time series from hourly averaged and clear sky irradiance datasets. By employing Linear interpolation, a technique that helps to achieve the desired time resolution and afterward computing critical factors, the algorithm identifies periods characterized by short-term weather phenomena, thus creating a high-resolution time series that accurately represents these dynamics. Avoiding the probabilistic components and machine learning techniques conserves computational power and reduces calculation time, but this comes at the cost of reduced fidelity in reproducing the results. Improving accuracy in PV simulations is not always directly related to reproducing real phenomena, but enhancing the amount of information contained in the data is sufficient. This study's approach enhances user-friendliness and facilitates seamless integration into existing energy modeling frameworks, aiming for representation with sub-hourly time steps. The algorithm's effectiveness is demonstrated by applying the model to hourly averaged data to revert them to a one-minute time step, and finally comparing the synthetically produced one-minute GHI data to the original measured data. The comparative analysis between synthesized and measured data demonstrated a strong agreement, with normalized mean bias error (MBE) values ranging between 1.8% and 9.6% and normalized root mean square error (NRMSE) values between 2.7% and 16.1%, depending on weather conditions. Additionally, the coefficient of determination (R^2) consistently remained above 0.64. Successful algorithm validation makes our algorithm suitable for use in meteorological datasets and locations, with similar climatic characteristics.

1. Introduction

Global annual renewable capacity additions increased by nearly 50% to 510 gigawatts (GW) in 2023, marking the fastest growth rate in two decades. This is the 22nd consecutive year of record-setting renewable capacity additions. Europe, the United States, and Brazil reached all-time highs in renewable capacity, while China's acceleration was extraordinary. In 2023, China commissioned as much solar PV as the entire world did in 2022, and its wind power additions grew by 66% year-on-year. Solar PV alone accounted for three-quarters of global renewable capacity additions [1], underscoring its importance for advancements in renewable energy. Global module prices fell 22%, reaching a record low of $0.11/Wdc, while PV manufacturers remained profitable due to increased sales volumes.

The global transition to renewable energy emphasizes the crucial role of photovoltaic (PV) systems, especially as capital costs continue to fall. Effective deployment of solar PV systems depends on accurately assessing the solar energy potential at the target location, typically through solar irradiance measurements. However, in many regions, this data is unavailable due to the high costs and limited accessibility of measurement systems. Solar irradiance data can originate from

* Corresponding author.
E-mail addresses: diamantis.almpantis@energy.lth.se (D. Almpantis), henrik.davidsson@ebd.lth.se (H. Davidsson), martin.andersson@energy.lth.se (M. Andersson).
URL: https://www.linkedin.com/in/diamantisalmpantis/ (D. Almpantis).

https://doi.org/10.1016/j.seja.2024.100076
Received 5 July 2024; Received in revised form 16 October 2024; Accepted 4 November 2024
Available online 12 November 2024

satellite-based observations, ground-based measurement networks, numerical weather prediction models, and reanalysis datasets. While these sources provide accurate data, the first three are often costly and computationally intensive, making them impractical for many locations. As a result, solar irradiance must frequently be estimated using reanalysis techniques or synthetic models to overcome these limitations [2]. High-frequency solar radiation data is unavailable for many locations, limiting research to specific sites where such data exists. Alternatively, computer-simulated data can be used if the synthetic generation model produces data with statistical characteristics similar to typically measured data [3]. The requirements for temporal resolution (monthly, daily, hourly, or sub-hourly values) and accuracy depend strongly on the technology and the purpose of the modeling [4]. High-resolution data, captured with a high level of detail, plays a critical role in detecting short-term phenomena and preventing errors in diverse PV system applications.

Last decades, solar radiation data reporting from modern radiometric stations has shifted from hourly intervals to much shorter time steps, typically 1 to 10-minute intervals, and sometimes even less. Proper energy simulations for CSP projects also demand time steps shorter than the conventional hourly interval, due to non-linear and transient effects impacting system performance. An ideal simulation time step would be 10 min or less [5]. To meet this need, some commercial satellite-derived irradiance providers now offer 10–15-minute time series, alongside the standard hourly data. For PV systems, rapid cloud-induced irradiance fluctuations, or ramping effects, are analyzed with time steps as short as 3 s in [6]. These advancements highlight the increasing demand for GHI data at sub-hourly intervals.

Various applications of high-resolution data are explored in the literature, including computational simulations, solar integration in buildings, direct coupling in energy sectors, as well as grid and off-grid operations. For instance, Mayer [7] highlighted the reliability of meteorological datasets in the optimal design of photovoltaic power systems, noting that optimization based on averaged hourly data underestimates the levelized cost of electricity by up to 3%, overestimates the inverter sizing ratio by 0.05 on average, and the tilt angle by up to 5° compared to high-resolution datasets. Similarly, Cao and Sirén [8] examined the impact of time resolution on the matching of PV system production and household electric demand, finding that 1-hour resolution data can be a source of significant error, sometimes resulting in errors of 15% or higher under scattered cloud conditions compared to 1-minute resolution data. Furthermore, Klokov and Loktionov [9] compared grid-tied systems with off-grid PV system layouts and reliability, concluding that in off-grid systems, reliability was overestimated, recommending a 30-minute resolution. In more complex urban locations, Bottecchia et al. [10] proposed that high-resolution data is required to detect the impact of surrounding buildings, despite the higher costs. Martin Hofmann et al. [11] illustrated the importance of one-minute data for the simulation of PV systems, finding that an energy yield loss of 7% was calculated in hourly averaged data compared to 10% in one-minute power output when the 70% restriction is applied by the inverter.

Global Horizontal Irradiation is the link between PV electrical power output and time-resolution data. The efficiency of PV modules is primarily dependent on global irradiance, with secondary factors like module temperature also playing a role. Due to the nonlinear relationship between module efficiency and irradiance, high temporal resolution simulations are necessary. Vanicek et al. [12] supports that for understanding the dynamic interaction of PV generators, storage systems, loads, and grids on a global scale, high-quality one-minute data series are crucial. Simulating systems with hourly averaged values neglects significant behavior patterns like short-time power enhancements. Global irradiation time series are fundamental inputs for simulating electricity production by PV panels and essential for all energy modeling frameworks incorporating renewable energy sources. The time resolution requirements of irradiance time series vary depending on the specific application of the simulation models. Richard Meyer et al. [13] and Gall et al. [14] emphasized the need to use time series with a time step shorter than 1 h for detailed performance simulations, suggesting a time resolution of 5 min to 15 min as a reasonable compromise between computational cost and accuracy. The power output variability from PV systems is a concern for grid operators, as unanticipated changes in PV output can strain the grid. Therefore, it is important to statistically characterize the GHI fluctuations at high temporal resolution. Kleissl [15] mentioned that PV power variability can be counteracted by fast-ramping generation sources and storage systems, but these options are expensive and substantially increase plant costs.

Several commercial providers and free sources offer meteorological data at a resolution of one hour (e.g., Meteotest, SolarGIS, and TMY), covering nearly the entire earth. However, the availability of measured irradiance data with a resolution of less than an hour is very limited. This limited availability necessitates synthesizing high-resolution time series from hourly averaged data. The demand for high-resolution irradiance data is evident in the growing number of publications presenting models for generating synthetic solar irradiance data with high temporal resolutions [16,17]. Dean Laslett et al. [18] summarizes many algorithms used to generate synthetic solar irradiance. [19] were among the first to develop a stochastic procedure for generating synthetic sets of hourly solar irradiation values, suitable for use in solar simulation design work. Recent studies have focused on generating high temporal resolution pairs of GHI and Direct Normal Irradiation (DNI) values, preserving the relationship between them [3]. Many of these models rely primarily on autoregressive moving average (ARMA) or Markov transition matrix (MTM) techniques, which produce time series with enhanced temporal resolutions, aiding in modeling the dynamic behavior of solar radiation [20]. Ngoko et al. [21] introduced a second-order MTM model incorporating statistical characteristics related to atmospheric conditions (clear, cloudy, overcast), improving upon the first-order MTM model used by Richardson and Thomson [22] for generating 1-minute values. Additionally, the combination of wavelets and artificial neural network (ANN) techniques with Markov transition matrices for generating solar radiation values is mainly applied to forecast applications handling daily or hourly global irradiance data [23]. Physical spatiotemporal downscaling methods also exist for satellite-derived datasets. However, in complex multi-objective problems solved by evolutionary algorithms, the use of higher-resolution data is often impractical due to the long calculation times required [24]. The extended computation time associated with high-resolution datasets can be a significant limiting factor in many of these applications.

Deterministic studies provide general insights into the effects of increased integration of solar generators in power systems, especially in distribution systems, but they do not account for the inherent variability of solar power. Probabilistic studies, on the other hand, capture this variability, offering more detailed analysis by utilizing solar irradiation as input for simulation programs. While models based on Markov chain techniques achieve satisfactory results in reproducing the modeled data, they utilize complex approximate equations of the probability density functions, and the errors introduced by these approximations are not quantified. Markov chain techniques, as probabilistic models, are highly influenced by the randomness and stochastic nature of the analyzed environment. These models also often ignore the dependence of the clearness index on the solar elevation angle. To minimize computational requirements in downscaling synthesis methods, a non-probabilistic and artificial intelligence approach is essential. Therefore, this paper introduces a straightforward method, a simplified solar downscaling synthesis algorithm designed to produce high-resolution solar output data without depending on the stochastic nature of the Markov chain method and machine learning. This algorithm is invaluable for applications focused on understanding fluctuations in solar electricity supply rather than precise representation. The present work comprehensively examines the significance of high-resolution data in PV analysis, addressing its profound impact

on efficiency, energy matching, simulation accuracy, and the broader transition towards renewable energy sources, and suggests an algorithm for high-resolution solar data synthesis.

Our suggested algorithm is designed to enhance the coupling of photovoltaic (PV) systems with electrolyzers, enabling more accurate simulations of solar renewable electricity production, hydrogen generation, and overall system optimization. Integrated PV-electrolyzer systems offer a highly efficient solution for producing low-cost green hydrogen. This is facilitated by the significant reduction in solar panel prices and the abundance of sunlight, which supports the distributed production of hydrogen. In the absence of electrical storage, such as batteries, these systems typically rely on the utility grid to meet electricity demands and serve as a backup for the electrolyzer balance of the plant.

This energy configuration can generate profit through four key mechanisms. First, the system can participate in the hydrogen market by selling green hydrogen. Second, the generation of renewable electricity from PVs avoids reliance on fossil fuels and can lead to the acquisition of carbon credits. Third, surplus electricity can be sold in day-ahead electricity markets by oversizing PV capacity relative to the electrolyzers, generating additional revenue during peak hours. Fourth, the system can offer flexibility in auxiliary markets by adjusting the operation of electrolyzers and hydrogen storage, further enhancing profitability. High-ramp-rate electrolyzers, such as proton exchange membrane (PEM) electrolyzers, are especially suited for this.

Hydrogen production in such systems provides several forms of flexibility, including regulation, balancing, operational reserves, and seasonal energy arbitrage. Regulation addresses minute-to-minute uncertainties and generator responsiveness while balancing manages variability and uncertainty within and across hours, typically requiring sub-hourly or hourly data. Operating reserves offer backup for unexpected events, and seasonal energy arbitrage helps manage energy mismatches across seasons [25]. Sub-hourly solar irradiance datasets are crucial for maximizing this flexibility, especially in regulation and balancing operations. These datasets enable precise time-matching between solar generation and system demands, which is critical for optimizing performance and profit.

To support renewable hydrogen production, policies encouraging the sourcing of renewable energy from the grid must consider factors such as temporal correlation (simultaneity). This ensures that renewable electricity generation aligns with electrolysis use, ranging from 15-minute intervals to annual scales [26]. Post-2030, additional power systems will need to follow temporal resolutions of the same hour or less. This study focuses on simulating the co-location of solar panels and PEM electrolyzers with precise time matching, exploring its potential impact on future temporal correlation frameworks.

Compared to annual time matching, hourly time matching in H_2 production under additionality modeling frameworks requires higher capacities of contracted resources, typically resulting in low or even negative emissions [27]. Reducing the time resolution below one hour offers further advantages, particularly in solar integration, off-grid operation, and sector-coupling scenarios. With PEM electrolyzers capable of responding to power signals within seconds and startup times of less than a minute [28], relying on hourly averaged solar power output may cause operational errors and mismatches in systems without intermediate storage, such as batteries. Therefore, this study employs 1-minute synthesized global irradiance data, addressing the need for high-resolution datasets. While most studies focus on electricity generation when downscaling temporal resolution, [13] highlights the potential for using this approach in the planning and operation of green hydrogen systems.

To highlight the trend towards shorter time steps, this study focuses on 1-minute intervals, utilizing solar irradiance observations from a high-quality research station in Denmark. The data gathered from a relatively unstable climatic condition, are intended to ensure robust and general conclusions for this specific case study. This paper aims to introduce a novel, non-probabilistic downscaling method for synthesizing 1-minute global irradiance data, using hourly data as inputs. This approach overcomes the challenges of high computational demand and the complexity of probabilistic models, enabling efficient and accurate calculations of solar renewable electricity generated by PV systems. This method enhances the integration and optimization of renewable energy systems, supporting the continued advancement of renewable energy technologies.

The rest of the paper is organized as follows: In Section 2, an overview of the clear sky approaches and models utilized in this study are presented. This section aims to provide a comprehensive understanding of the theoretical and practical aspects of clear sky modeling, which is fundamental for accurate solar irradiance synthesis. Section 3 delves into the topology of real measured data with the high-frequency and the methodology of the proposed algorithm, detailing the step-by-step processes, techniques, and equations employed to synthesize high-resolution solar data. In Section 4, the results of the proposed algorithm are presented and discussed for specific days. Section 5 describes the validation techniques used to assess the algorithm's performance, followed by a thorough presentation of the validation results across the examined location of Lyngby, Denmark, and period. This section aims to demonstrate the robustness and reliability of the algorithm through extensive testing and analysis. Finally, Sections Section 6 & Section 7 include the discussion and conclusion, summarizing the key findings of the study, addressing potential implications, and suggesting directions for future research. This section encapsulates the overall contributions of the paper and reflects on the broader impact of the work within the context of renewable energy integration.

2. Clear sky global horizontal irradiation model

Assessing solar energy potential for technology deployment requires evaluating clear-sky solar irradiance, often estimated through models due to measurement limitations. Clear sky days are generally characterized by the absence of visible clouds across the entire sky dome. Clouds cause fluctuations in the GHI by affecting its direct and diffuse components. Clouds are a major unpredictable factor affecting the amount of irradiance reaching the Earth's surface. Bright clouds can increase the diffuse component of global irradiance, resulting in an overall increase in GHI, while clouds obstructing the sun reduce the GHI [24]. The degree of cloudiness is quantified through various methods, including human observation, ground-based measurements such as sky cameras and irradiance measurements, and satellite instruments, each with varying accuracy. The method used to determine a clear sky day depends on the available measurements at the site, and an inaccurate clear sky detection can lead to incorrect conclusions. Additionally, even without visible clouds, other atmospheric parameters like aerosols, precipitable water, or hazy conditions can affect atmospheric transmittance, influencing the amount of solar irradiance reaching the Earth's surface [29].

Selecting appropriate clear sky models depends on the models' accuracy and the availability of input parameters, which are primarily atmospheric or meteorological. In many developing countries and other regions, these parameters are often not readily available due to the high cost and maintenance requirements of necessary measuring equipment. Therefore, when the accuracy is not the focus of the paper, using simple models based on basic input parameters that provide sufficient accuracy to estimate localized clear sky global irradiance in areas where such measurements are not available is suggested [30]. These approaches rely on computing the solar zenith angle, a key parameter determining the magnitude of solar irradiance at any location and time on typical clear sky days, particularly in regions where atmospheric parameters such as water vapor show little or no fluctuation.

All clear sky models rely on extraterrestrial radiation. Extraterrestrial radiation (H_0) is defined as short-wave solar radiation in the absence of an atmosphere. It is a well-behaved function of the day of the

year, time of day, latitude, and longitude, providing a baseline measure of solar radiation unaffected by atmospheric conditions [24]. In many studies, extraterrestrial radiation is identified with the solar constant (G_{sc}), but this is not the case when an hourly or sub-hourly analysis is conducted. In addition, variation of the earth-sun distance does lead to a variation of extraterrestrial radiation flux in the range of ±3.3%. The dependence of extraterrestrial radiation on earth-sun distance is shown in Eq. (1) [31].

$$H_o = \frac{1}{\pi} \cdot G_{sc} \cdot d_r \left(\cos(\phi)\cos(\delta)\sin(\omega) + \frac{\pi}{180}\omega \sin(\phi)\sin(\delta) \right) \quad (1)$$

where H_o represents the extraterrestrial radiation during the hour (or shorter) period, measured in MJ m^{-2} $hour^{-1}$. It quantifies the solar radiation received from the sun under ideal conditions. G_{sc} denotes the solar constant, a fixed value of 1361 W/m^2. It represents the amount of solar radiation received at the top of Earth's atmosphere on a surface perpendicular to the sun's rays. d_r is the inverse relative distance factor for the earth-sun [unitless]. δ is the solar declination, measured in radians. It varies throughout the year due to Earth's axial tilt. ϕ represents the latitude of the location where solar radiation is being calculated, measured in radians. It influences the angle at which solar radiation reaches the Earth's surface. ω is the solar time angle, measured in radians. It is calculated based on local solar time and longitude. The declination angle can be approximated as [32]:

$$\delta = 23.45 \sin\left[\frac{360}{365} \cdot (n + 284) \right] \quad (2)$$

where n represents the day of the year. The hour angle can be approximated as [32]:

$$\omega = [(h - 12) + \frac{m}{60}] \times 15 \quad (3)$$

where h are hours and m are minutes in solar time.

Clear-sky solar radiation (G_{clear}) represents the amount of global solar horizontal irradiation (GHI) that would be received at a weather measurement site under clear-sky (cloud-free) conditions. The value of G_{clear} depends on the time of the year and latitude, and additionally on the time of day for hourly calculations. These factors influence the potential incoming solar radiation from the sun. Several other parameters also impact clear-sky solar radiation. Station elevation affects atmospheric thickness and transmissivity, while the amount of precipitable water in the atmosphere influences the absorption of some shortwave radiation. Moreover, the presence of dust or aerosols in the air can further modify the amount of clear-sky solar radiation reaching the surface. Despite there being many simplified clear sky models in literature [33], a more complex and accurate procedure is preferable, considering the effects of sun angle and water vapor on the absorption of short wave radiation and by separating the components of the beam and diffuse radiation [34]. Eq. (4) describes the clear-sky solar radiation in hours or sub-hours periods:

$$G_{clear} = (K_B + K_D)H_o \quad (4)$$

where K_B is the clearness index for direct beam radiation [unitless], K_D the transmissivity index for diffuse radiation [unitless] and H_o is the extraterrestrial radiation [MJ m^{-2} h^{-1}]. The following equation presents the clearness index for direct beam radiation K_B, [35]:

$$K_B = 0.98 \exp[\frac{-0.00146P}{K_t \sin\beta} - 0.075(\frac{W}{\sin\beta})^{0.4}] \quad (5)$$

where, K_t is the turbidity coefficient [unitless], and is recommended to be equal to 1, P is atmospheric pressure at the site elevation in kPa, β the angle of the sun above the horizon called solar altitude angle (radians) and W is the precipitable water in the atmosphere [mm]. Precipitable water is predicted as

$$W = 14.0 e_a P + 2.1 \quad (6)$$

where e_a is the actual vapor pressure of the air (at approximately 2 m) [kPa]. For hourly or shorter periods the sun angle ϕ is calculated as :

$$\sin\phi = \sin\phi \sin\delta + \cos\phi \cos\delta \cos\omega \quad (7)$$

Table 1
Intermediate variables for measured location, Lyngby, Denmark.

Variable	Symbol	Value set
Solar Constant	G_{sc}	1361 W m^{-2}
Longitudetz	L_{tz}	345°
Kt	K_t	1.0
Distance factor for the earth-sun	d_r	0.99
Declination angle	δ	12.89°
Solar N	N	13.13
Solar Z	Z	48.1°
Atmospheric pressure	P	100.9 kPa
Vapor pressure of the air	e_A	1.54 kPa
Precipitable water in the atmosphere	W	23.9 mm

After calculating the clearness index for direct beam radiation K_B, the diffuse radiation index is estimated as:

$$K_D = 0.35 - 0.36 K_B \quad (8)$$

These models, driven by extraterrestrial radiation and accounting for atmospheric conditions, provide crucial insights into potential energy generation under ideal conditions. However, their application across diverse geographical and meteorological settings presents complexities, as factors such as cloud cover and atmospheric variability pose challenges. The ongoing refinement of clear sky models, incorporating advanced techniques and precise input parameters, is essential for enhancing their reliability and applicability worldwide. For hourly periods, a clear sky calculator, available at [36], can be used to provide the intermediate variables for the previous equations for each location that needs to be examined and calculate the estimated shortwave. Table 1 presents the intermediate variables that be used for the measured location in Lyngby, Denmark for clear sky global horizontal irradiance hourly estimation, that be in line with the real measurements.

3. Data & methodology

3.1. Data used

The data used in this study were recorded from January 2023 to mid-December 2023 at a measurement site located in Lyngby, Denmark. This location was chosen due to its proximity to Lund, Sweden, and because Denmark was identified as a key site for technology installation within the framework of the European project Circular Fuels, to which this work is related. The station is on the Technical University of Denmark's Lyngby Campus, at Building 119 (55.79064° S, 12.52505° E, 50 m AMSL). The measurements included global horizontal irradiance (GHI), direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), longwave downward radiation (LWD), rain accumulation, air temperature, relative humidity, air pressure, average wind speed, and average wind direction. All data were sampled at 1-minute intervals [37]. The meteorological data were obtained using high-precision instruments. GHI and DHI were measured with Kipp and Zonen CMP11 pyranometers, while DNI was measured using a Kipp and Zonen CHP1 pyrheliometer. A rigorous quality control (QC) process was implemented, consisting of both automatic flagging of potentially erroneous values and manual inspection to remove invalid periods. While most QC tests were automated, some involved manual visual inspections by the evaluators. The applied QC tests are:

- Missing timestamps
- Missing values
- K-Tests [38,39]
- Baseline Surface Radiation Network (BSRN)'s closure tests [40]
- BSRN's extremely rare limits test [40]
- BSRN's physically possible limits test [40]
- Tracker-off test, improved from [41]

- Visual inspection, including shading assessment, closure test, AM/PM symmetry check for GHI, and calibration check using the clear-sky index.

These measures ensured the accuracy and reliability of the irradiance data used in the analysis. Further details on the quality control methodology can be found in Forstinger et al. [42]. After receiving the data, a comprehensive meta-analysis was performed to verify its integrity. This involved detailed visual inspections and the application of additional quality checks. These steps were critical in ensuring that only valid and reliable data were used for analysis.

3.2. Methodology

The suggested algorithm is based on real one-minute measured data, which are initially averaged hourly to simulate the hourly average irradiance available in open sources. This aids in the validation of the synthesized data by comparing them with the real measured data. The measured data are averaged to achieve the desired hourly resolution. It is important to note that while the time resolution differs between the one-minute and hourly datasets, the total daily GHI availability should remain consistent. This means that when integrated over an entire day, the total energy received by the surface, as measured by the global horizontal irradiance (GHI), will be the same regardless of whether the data is collected at one-minute or one-hour intervals. The finer resolution data (1 min) captures more detailed fluctuations throughout the day, but when summed over the full day, the total irradiance should equal that of the hourly averaged data, which does not capture much detail. This ensures that the daily solar energy input remains unchanged, regardless of the averaging method. For creating hourly averaged global horizontal irradiance data and other meteorological parameters, it is beneficial to use data points from half an hour before and half an hour after the desired hour. This method provides a more accurate representation of the hourly averaged data and aligns closely with clear sky models. Consequently, the developed hourly averaged global irradiance time series (G_i) is structured with 24 timesteps per day.

Having the hourly averaged time series of global irradiance on the surface (G_i) next step is the calculation of the clear sky irradiance $G_{\text{clear},i}$ according to the Environmental and Water Resources Institute of the American Society of Civil Engineers (Environmental and Water Resources Institute of the American Society of Civil Engineers, 2001) as described in Section 2. Afterward, hourly averaged time series and clear sky estimations are used in the calculation of clearness index values each hour k_i, the average clearness of the day $\bar{k}_{(i,\text{day})}$, the variability during the day $\tilde{k}_{(i,\text{day})}$, and the absolute ramp rate of the clearness index r_i. Clearness index (k_i) is a crucial parameter, representing the ratio of the observed measured global horizontal irradiance at the surface to the irradiance calculated for cloudless conditions at the measuring site, denoted by clear sky irradiance. It indicates the clarity of the atmosphere and the extent to which clouds and other atmospheric components attenuate solar radiation. The clearness index is calculated using the following Eq. (9), [33]:

$$\mathrm{k}_i = \frac{G_i}{G_{i,clear}} \tag{9}$$

where, G_i is the hourly averaged global irradiance (W m^{-2}) and $G_{i,clear}$ are the clear sky irradiance hourly data (W m^{-2}). The predominant weather condition on a particular day results in a characteristic temporal pattern of k_i that can be used to categorize the day into one of the three categories (cloudless, broken, and overcast). The detection algorithm of the weather condition is based on the daily average of hourly averaged k_i values $\bar{k}_{i,\text{day}}$ and the variability during a day $\tilde{k}_{(t,\text{day})}$. Eq. (10) describes the daily average and Eq. (11) the variability during the day, [33]:

$$\bar{k}_{(i,\text{day})} = \frac{1}{d}\sum_{d=1}^{24} k_i \tag{10}$$

Table 2
Overview of the three weather classes and their detection conditions [11].

Weather class	Conditions
Overcast	$0.6 - \bar{k}_{i,\text{day}} \le \tilde{k}_{i,\text{day}}$
Cloudless	$-0.72 + 0.8\bar{k}_{i,\text{day}} \ge \tilde{k}_{i,\text{day}}$
Broken clouds	Otherwise

$$\tilde{k}_{(i,\text{day})} = \frac{1}{d}\sum_{d=1}^{24} |k_i - k_{i-1}| \tag{11}$$

where d is the number of hours where global and clear sky irradiance is above 0 W m^{-2}.

The hourly clearness values and variability during the day are used to characterize the examined day as cloudless, overcast, or broken according to [11]. Table 2 gives an overview of the three weather classes and their detection conditions.

Having characterized the examined day, it is essential to calculate another crucial factor: the ramp rate of the clearness index throughout the day. The ramp rate refers to the rate of change in the clearness index (K_i) over time. This parameter is significant as it helps in understanding the variability and stability of solar irradiance, which is critical for the analysis. Analyzing the ramp rate can gain insights into the fluctuation patterns of solar radiation, which are influenced by atmospheric conditions. Eq. (12) describes the ramp rate of the clearness index each timestep, [33]:

$$r_i = \frac{k_i - k_{i-1}}{k_i} \tag{12}$$

Linear interpolation is the next step of the suggested algorithm and is conducted in three hourly time series (hourly averaged irradiance, hourly clear sky irradiance, and ramp rate) to achieve the desired time resolution. This analysis decomposes these functions into their oscillatory components, allowing for the attainment of the desired time resolution. Linear interpolation allows for even finer resolutions, such as seconds or intervals of 10 or 15 min. However, for this study, a 1-minute resolution is chosen to obtain more detailed results and to simulate an extreme scenario in anticipation of future EU regulations on time matching. Following this, the data is resampled to minute intervals and values are interpolated to further refine the time series. Assuming a one-minute time-resolution data decided to be synthesized, then from 24 size vector (i) time series, with linear interpolation a 1440 size vector (t) will be created. Linear interpolation is applied to three critical hourly time series: the hourly averaged global horizontal irradiance G_i, the hourly clear sky global horizontal irradiance estimation $G_{clear,i}$, and the ramp rate of the clearness index r_i. The absolute difference between one-minute resolution total irradiance and clear sky radiation is calculated $|G_{t,\text{clear}} - G_t|$. Finally, a proposed linear function $f\left(|G_{t,\text{clear}} - G_t|, |r_i|\right)$ is used to modify the linear interpolation data and synthesize the high resolution global horizontal irradiance data. Fig. 1 is a graphical representation of the methodology described in this section. The primary purpose of this algorithm is to evaluate its ability to downscale it back to a higher time resolution while maintaining minimal errors.

In minute-level irradiance datasets, variations in solar intensity throughout the day lead to higher irradiance levels during periods of clear skies or midday, contrasting with lower levels during partial cloud cover, early morning, or late afternoon. To minimize errors in the time series during these transitional periods, a small threshold is implemented. Specifically, an irradiance threshold of 35 W m^{-2} is set to manage significant differences at the start and end of each day. Additionally, a ramp rate threshold of 2.5% is applied to moderate extreme changes in irradiance at the beginning or end of the day, for the same reasons. The introduction of those thresholds would be avoided in case more constraints in data cleaning were applied in the raw data series of the high or low-frequency GHI. When the absolute difference or ramp

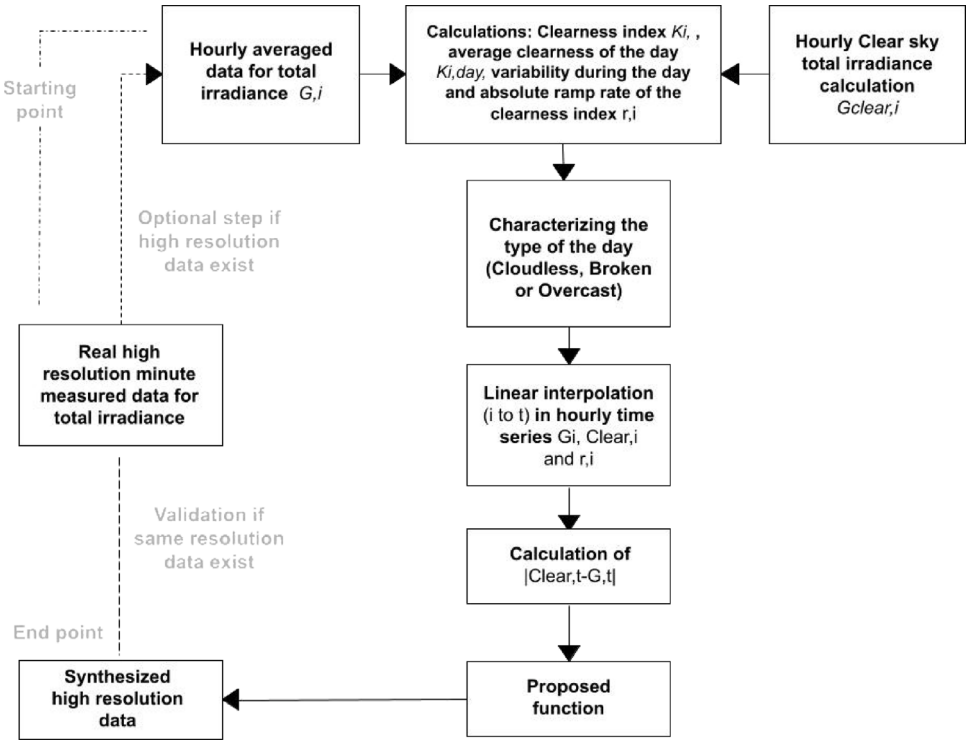


Fig. 1. Methodology overview.

$$f(t)=\begin{cases}\text{Linear interpolation}(t) & \text{if } |G_{t,\text{clear}}-G_t|\le 35 \text{ or } |r_t|\le 0.25\\ \text{Linear interpolation}(t)+\alpha\cdot\text{random}(0,1)\cdot\text{Linear interpolation}(t) & \\ -\frac{\alpha}{3}\cdot\text{Linear interpolation}(t) & \text{if } |G_{t,\text{clear}}-G_t|>35 \text{ or } |r_t|>0.25\end{cases} \tag{13}$$

Box I.

rate exceeds these thresholds, a linear function adjusts the synthesized data through a random addition or subtraction based on a factor α. This adjustment is performed at each time step during linear interpolation to emulate short-term weather phenomena captured by variations in the clearness index. Optimizing the dimensionless factor α is critical, particularly considering the specific weather conditions of the day. The proposed function is formulated as shown in Box I.

In Box I, linear interpolation (t) constructs a vector with 60 parameters from a vector with 1 parameter (ranging from 1 to 60-time steps). The function random$(0,1)$ utilizes a random number between -1 and 1 to introduce stochastic variability into the synthesized data and is mathematically presented as random $\sim \mathcal{U}(-1,1)$.

3.2.1. Step-wise method description

In this subsection, an analytical description of the methodology is presented as a guidance to the reader. Beginning with the data set of real measured global horizontal irradiance some preliminary statistics are presented, in Table 3, to help choose the most representative month-of-the-year data to implement the suggested algorithm. As a sample, the 1-min measured global horizontal irradiance data are used in the Lyngby, Denmark location.

According to Table 3, the average GHI per month and the standard deviation follow a similar pattern throughout the year at this location, whereas the maximum GHI shows slight variations. Since the focus is on detecting short-term phenomena, the primary factor to consider is the standard deviation of global horizontal irradiance each month. This metric allows for an understanding of the fluctuation or variability within each month. Additionally, a month that encompasses all three categories of days (cloudless, broken, and overcast) is aimed to be selected, which means avoiding months with the highest or lowest average GHI. Based on these criteria, July is proposed as a good representation of the year, and the algorithm will be implemented on its days. To obtain a holistic view and a representative sample of the month, six consecutive days that encompass the three categories of weather conditions and exhibit short-term phenomena were examined. The days examined were from July 5th, 2023 to July 10th, 2023. Following this, the 1-minute measured data was hourly averaged to imitate free open sources of global horizontal irradiance, which are accessible for download. Fig. 2 shows the real measured GHI data for the selected days in Lyngby, Denmark.

The second step after finding the sample is to hourly averaged the 1-minute time series and to calculate the clear sky global horizontal irradiance of each day according to Section 2. Both results are presented in Fig. 3. It is important to note that while this step in presenting the work is the second stage in our methodology, it may serve as the first step in other contexts. High-frequency data is often not available, which is why our methodology is designed to be flexible, allowing users to start the downscaling algorithm from various points. In cases where high-frequency time series are available, there are more effective ways to reproduce them by training on hourly datasets after applying linear interpolation to capture trends, using Fourier analysis, and calibrating the model effectively.

The third step, after obtaining the hourly averaged time series, involves calculating the hourly clearness index, the daily average of them, the variability of the day, and the ramp rate of clearness indexes for each day using Eqs. (9)–(12), as shown in Figs. 4 and 5. Regarding

Table 3
Sorted Monthly Maximum Global Horizontal Irradiance for Lyngby, Denmark (2023).

Months	Average GHI ($W\,m^{-2}$)	Standard deviation of GHI ($W\,m^{-2}$)	Maximum GHI ($W\,m^{-2}$)
June	279.77	299.64	1164.00
May	263.82	287.49	1123.00
July	216.43	242.77	1151.00
April	190.89	238.83	1099.00
August	154.13	195.48	1111.00
September	137.08	190.88	829.00
March	91.42	147.78	975.00
October	50.89	94.31	644.80
February	49.55	93.80	563.30
November	22.05	48.02	470.70
January	15.74	37.41	411.10
December	8.22	20.25	163.30

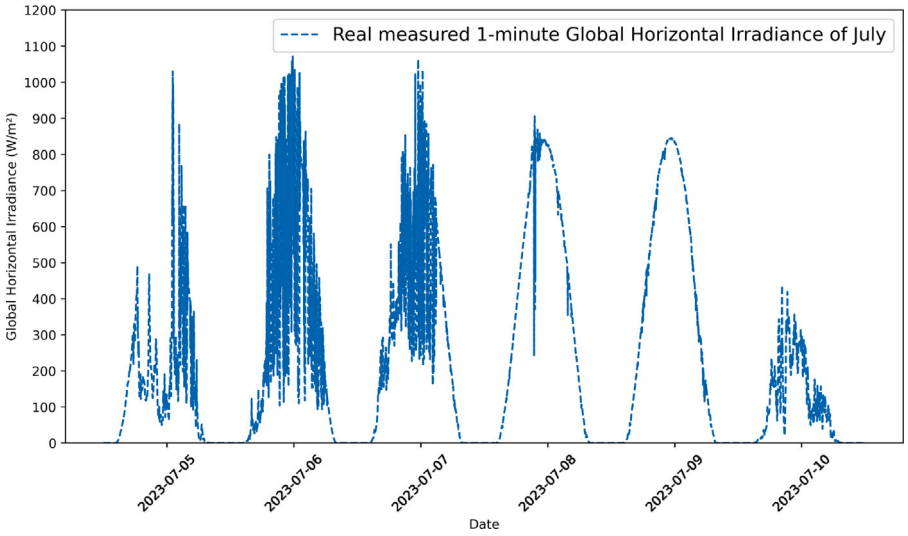


Fig. 2. 1-minute measured global horizontal irradiance in some days in July.

the clearness index, it is observed that, as expected, high values appear in the early morning and late afternoon on almost every day. These high values are not connected to short-term phenomena but rather to sunrise and sunset. On cloudless days, the clearness index remains close to 1 throughout the day, indicating that the global horizontal irradiance follows the trend of clear sky day irradiance. Conversely, on broken and overcast days, the clearness index fluctuates significantly, indicating the presence of short-term phenomena. The ramp rate provides crucial insights into the variability and dynamics of solar irradiance within each examined hour. By highlighting rapid and significant fluctuations, the ramp rate is an essential factor for understanding the underlying short-term phenomena affecting solar energy production. As expected in cloudless days the ramp rate is close to 0, but not in broken and overcast with short-term phenomena days.

The fourth step involves the days' characterization in one of the three clusters as cloudless, broken, and overcast. The focus is on broken and overcast (with short-term phenomena) days. Cloudless days generally follow clear sky estimations and usually do not present short-term phenomena, although exceptions can occur, as seen on July 8th. A broken day represents the most complex type of weather condition and tends to produce the largest errors in electricity production simulations for PV systems in similar locations like the examined one in Lyngby, Denmark, often bigger than 10%. Overcast days typically exhibit behavior similar to cloudless days. Still, when they occur in summer, such as in July, they can also involve short-term phenomena, like the 5th and 10th of July. It is emphasized that an exact correspondence between artificial data and minute-level measured data is not crucial. Instead, the algorithm's ability to accurately capture behavior patterns is of paramount importance.

The categorization of each day and the calculation of the hourly clearness index and its ramp rate completed, is following the increase of the time resolution from hours to minutes. This is achieved by implementing linear interpolation on the hourly time series, splitting each hour into 60 points, corresponding to 60 min. Consequently, the new step changes from 24 (hours) to $24 \times 60 = 1440$ (minutes).

The final step before applying the linear function to detect short-term phenomena is to calculate another important factor: the difference between the clear sky global horizontal irradiance and the hourly averaged global horizontal irradiance values after linear interpolation. This difference is calculated for all 1440 timesteps as $|G_{t,\text{clear}} - G_t|$. This variable, combined with the ramp rate of the clearness index already calculated at the desired resolution with linear interpolation, serves as the independent variable for the final proposed function. Box I incorporates 'noise' into the linear interpolation of global horizontal irradiance data to ensure the final time series closely resembles the original 1-minute measured data. There are two critical factors in this function. The first is the thresholds that imitate the sharp changes occurring early in the morning and late in the afternoon, corresponding to sunrise and sunset. The second is the factor α, which varies based on the category of the day. An optimization technique is employed to find the optimal value of α for each type of day between the three categories for July. Although this method may not provide the best results for every single day within each category due to the broad conditions described by Martin Hofmann et al. [11], it achieves a good match for the examined days and the entire month of July. The constraints used in the optimization of the factor α were:

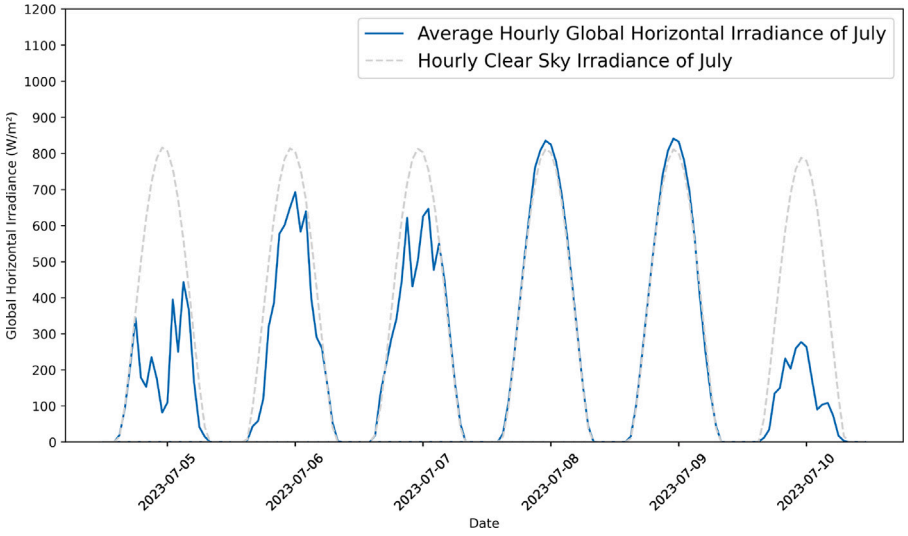


Fig. 3. Hourly averaged global horizontal irradiance in July with the estimated clear sky irradiance based on Methodology presented in Section 2.

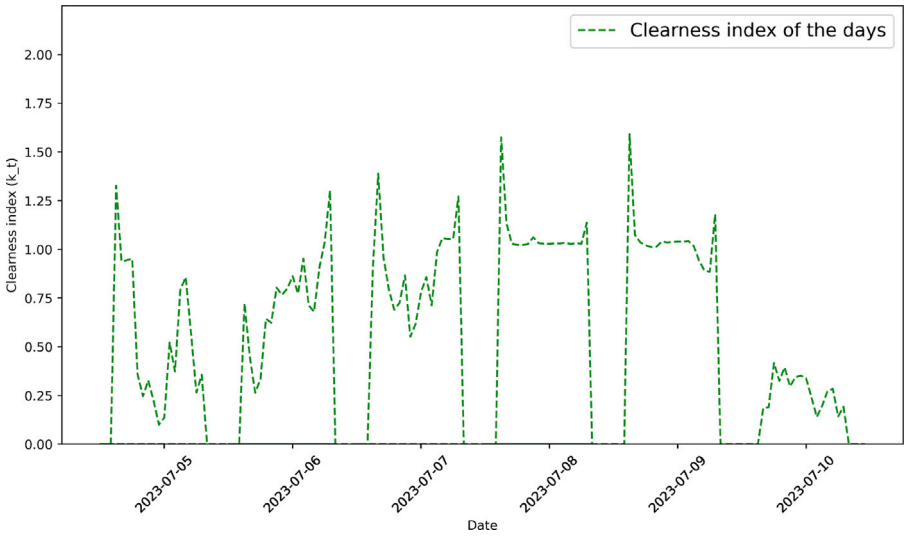


Fig. 4. Clearness index of the example days on July.

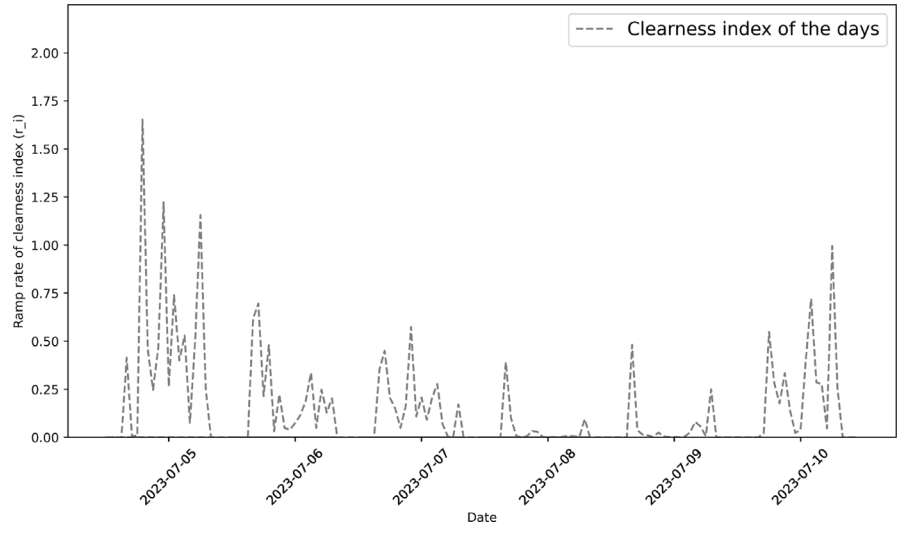
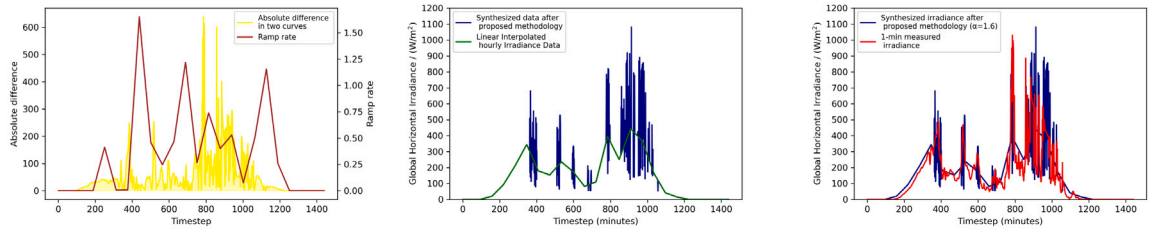


Fig. 5. Ramp rate of clearness index of the days.



(a) Difference and ramp rate between clear sky and linear interpolation data

(b) Comparison between suggested algorithm and linear interpolation

(c) Synthetic irradiance values compared to measured ones

Fig. 6. Results for 5th July.

Table 4
Factor α optimization.

Day category	Factor α
Cloudless	−0.4
Broken	1.2
Overcast	1.5

Table 5
Performance metrics.

Day	NMBE (%)	NRMSE (%)	R^2
5 July (Overcast)	5.0	9.4	0.64
6 July (Broken)	9.6	16.1	0.67
7 July (Broken)	7.2	12.1	0.75
8 July (Cloudless)	2.2	4.0	0.99
9 July (Cloudless)	1.8	2.7	1.00
10 July (Overcast)	6.5	12.0	0.75
5-10 July	4.6	9.4	0.87

(a) Maximum f(t) value is equal to maximum measured 1-min global horizontal irradiance G_m

(b) Mean f(t) value is equal to the mean 1-min global horizontal irradiance G_m

The average factor α is calculated for each type of day as presented in Table 4.

4. Results

In this section, the are presented results to demonstrate the implementation of the algorithm. The initial step involves handling the measured 1-minute global horizontal irradiance data. The dataset spans an entire year. Incomplete and missing data, which constituted approximately 4%, were excluded from the analysis. These missing data points were primarily concentrated in March and December. Figs. 6–11 display the results of the algorithm applied to synthesized data of the six days selected in Section 3.2.1 in comparison with real high-resolution measured global horizontal irradiance data for the same six days in July.

In each dataset, the first diagram presents the difference between hourly averaged GHI and clear sky data after linear interpolation, alongside the ramp rate at the same resolution. The second diagram illustrates the implementation of the function $f(t)$ in the linear interpolation data series. Finally, the third diagram compares the high-time-resolution synthesized data with the 1-minute real measured GHI data in Lyngby. The initial assessment of the algorithm's accuracy focuses on its ability to detect short-term phenomena, a fact that is proved by the Figs. 6–11. However, to draw more reliable conclusions, a comprehensive performance evaluation is necessary to gauge the fidelity of the synthesized data.

5. Performance evaluation

The performance of the models was evaluated by computing various statistical error parameters such as the root mean square error (RMSE), the mean bias error (MBE), and the coefficient of determination R^2 values. The RMSE is defined by the following equation [43]:

$$\mathrm{RMSE} = \left[\frac{1}{p}\sum_{p=1}^{p}\left(\mathrm{G}_t - \mathrm{G}_{t,m}\right)^2\right]^{\frac{1}{2}} \tag{14}$$

where G_t and $\mathrm{G}_{t,m}$ are the synthesized and measured global solar irradiance values, respectively, while p is the number of measurements considered. The RMSE values, in general, demonstrate the overall accuracy of the model, whereas smaller RMSE values signify a more accurate model. The MBE, which describes the trend of the models in overestimating or underestimating the measured irradiance, is given as follows in Eq. (15), [43]:

$$\mathrm{MBE} = \frac{1}{p}\sum_{p=1}^{p}\left(\mathrm{G}_t - \mathrm{G}_{t,m}\right) \tag{15}$$

Negative values of MBE signify an underestimation of the measured values, while positive MBE values indicate an overestimation. The R^2 value, which takes on values between 0 and 1, measures how well the models fit the measured data and is computed as follows in Eq. (16), [43]:

$$R^2 = 1 - \frac{\sum_{p=1}^{p}\left(\mathrm{G}_{t,m} - \mathrm{G}_t\right)^2}{\sum_{p=1}^{p}\left(\mathrm{G}_{t,m} - \frac{1}{p}\sum_{p=1}^{p}\mathrm{G}_{t,m}\right)^2} \tag{16}$$

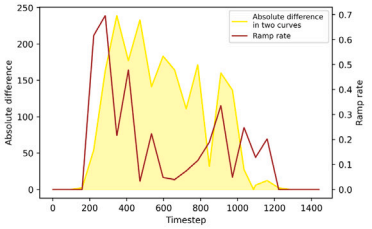
Generally, the higher the R^2 value is, the better the model is at predicting the measured data. For a better and quicker understanding, NRMSE and NMBE were calculated as normalized percentages of their corresponding mean values using the following equations, [43]:

$$\mathrm{NRMSE} = \frac{\left[\frac{1}{p}\sum_{p=1}^{p}\left(\mathrm{G}_t - \mathrm{G}_{t,m}\right)^2\right]^{\frac{1}{2}}}{\left(\frac{1}{p}\sum_{p=1}^{p}\mathrm{G}_{t,m}\right)} \tag{17}$$

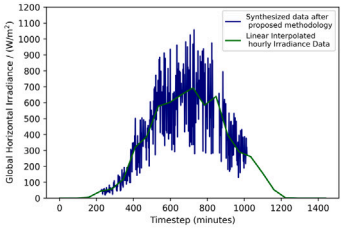
$$\mathrm{NMBE} = \frac{\sum_{p=1}^{p}\left(\mathrm{G}_t - \mathrm{G}_{t,m}\right)}{\sum_{p=1}^{p}\mathrm{G}_{t,m}} \tag{18}$$

Eqs. (17) and (18) give dimensionless NRMSE and MBE, respectively, which are normally expressed as percentages. These quantities should be as low as possible for better performing models. According to [44], in evaluating clear sky GHI, NMBE within ±10% and NRMSE < 20% would signify good fitting between modeled results and measured data and is quite often used as a quantitative measure for model performance. Table 5 compares the error and coefficient of determination values between the synthesized and measured 1-minute global horizontal irradiance data for the selected days, and the entire six days period.

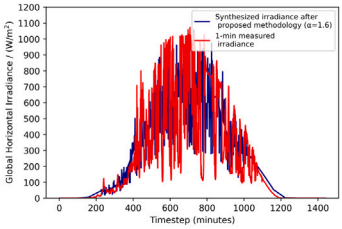
To assess the performance of the proposed algorithm over an extended period, performance metrics were calculated for the entire period from the 5th to the 10th of July, demonstrating positive results



(a) Difference and ramp rate between clear sky and linear interpolation data

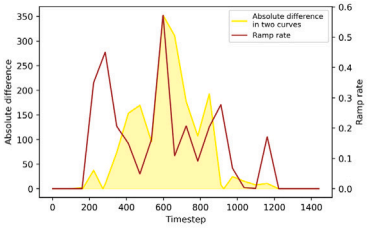


(b) Comparison between suggested algorithm and linear interpolation

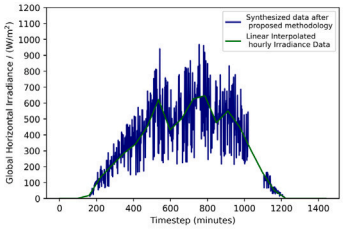


(c) Synthetic irradiance values compared to measured ones

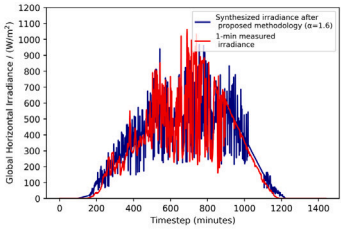
Fig. 7. Results for 6th July.



(a) Difference and ramp rate between clear sky and linear interpolation data

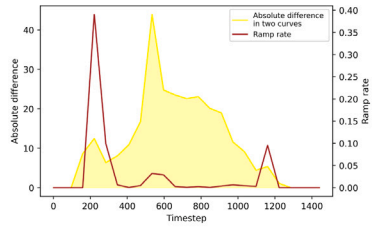


(b) Comparison between suggested algorithm and linear interpolation

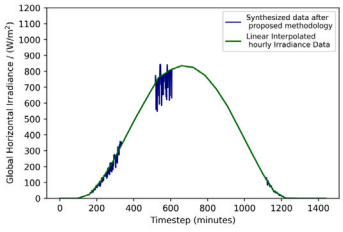


(c) Synthetic irradiance values compared to measured ones

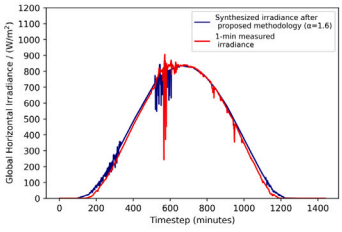
Fig. 8. Results for 7th July.



(a) Difference and ramp rate between clear sky and linear interpolation data

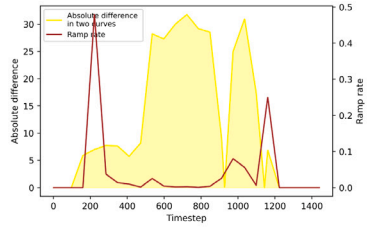


(b) Comparison between suggested algorithm and linear interpolation

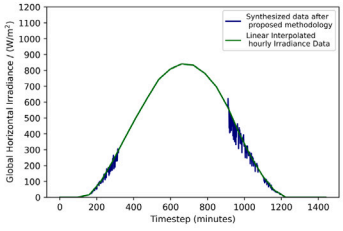


(c) Synthetic irradiance values compared to measured ones

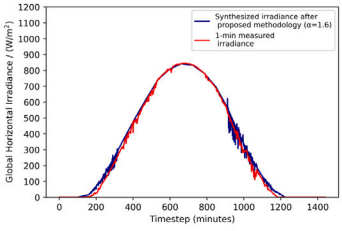
Fig. 9. Results for 8th July.



(a) Difference and ramp rate between clear sky and linear interpolation data



(b) Comparison between suggested algorithm and linear interpolation

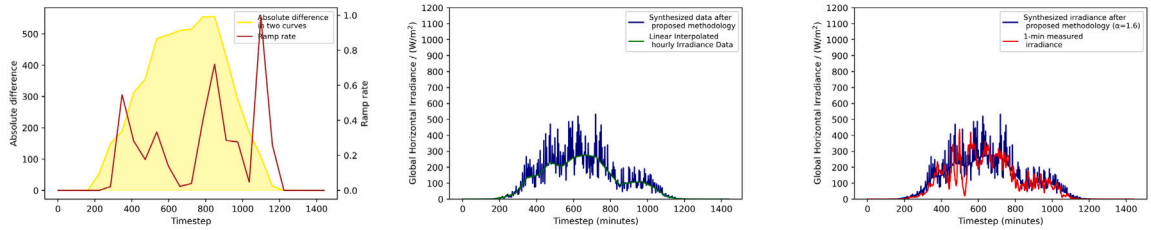


(c) Synthetic irradiance values compared to measured ones

Fig. 10. Results for 9th July.

for algorithmic accuracy. Fig. 12 presents the aggregated NRMSE, NMBE, and R^2 metrics for all days combined. The error margin is produced based on the factor α in the suggested function (Box I) after linear interpolation of hourly averaged data. This factor introduces the fluctuation of the synthesized data. The factor α is a result of averaging an optimization solution on each day and climate condition of the examined period and the only way to mitigate the error is by introducing more constraints and decreasing or increasing the frequency of adding the randomness in interpolated data series. Testing more days and different climate conditions can road to the best-optimized factor α but again a universal solution can have better or worse results depending on the climatic condition that is tested each time.

To evaluate the statistical representation of synthetic datasets, Probability Density Function (PDF) and Cumulative Distribution Function (CDF) analyses were conducted at a 1-minute resolution, crucial for assessing PV performance. The PDF analysis generates frequency diagrams illustrating the probability distribution of synthesized and measured data across irradiance clusters. This analysis reveals where the synthesized data either matches or deviates from the measured data within each cluster. Additionally, the CDF analysis portrays the cumulative probability distribution of GHI, offering insights into the



(a) Difference and ramp rate between clear sky and linear interpolation data

(b) Comparison between suggested algorithm and linear interpolation

(c) Synthetic irradiance values compared to measured ones

Fig. 11. Results for 10th July.

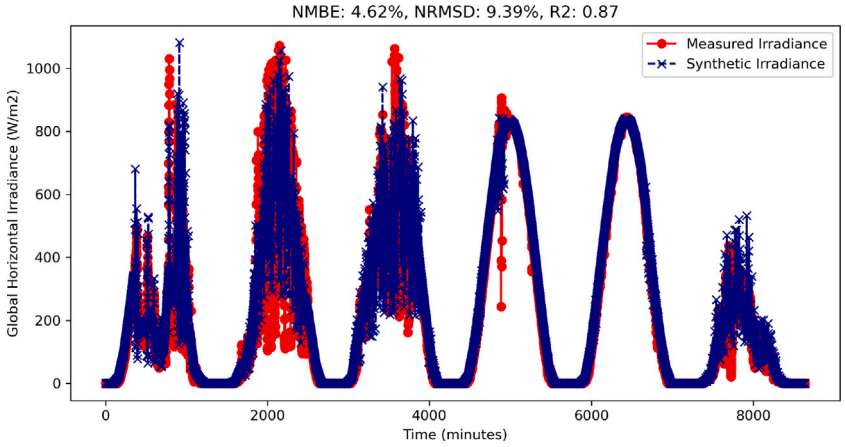
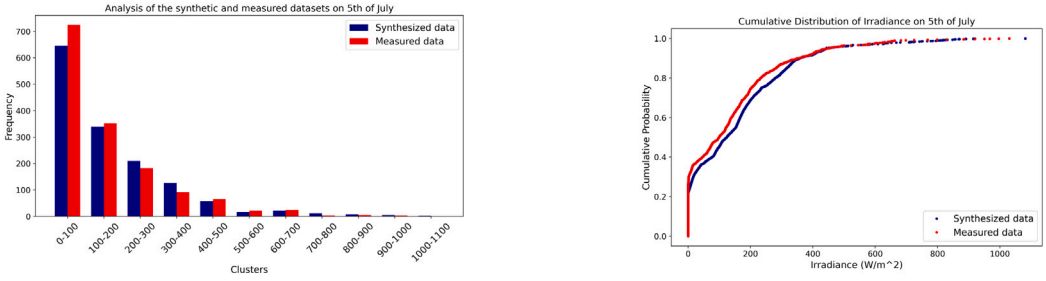


Fig. 12. Performance metrics for the sum of the six examined days.



(a) Frequency plot for 5th July

(b) CDF plot for 5th July

Fig. 13. Analysis for 5th July.

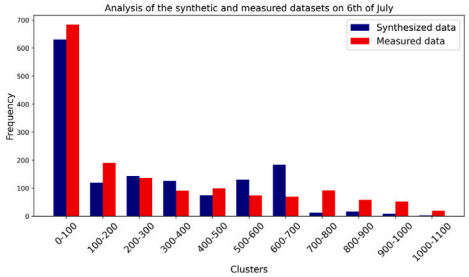
consistency and reliability of solar irradiance synthesis. Comparisons with the original measured data show that cubic interpolation of hourly values to a 1-minute scale results in slightly altered frequency distributions, particularly noticeable during partly cloudy conditions (broken and overcast days) where significant fluctuations occur. In contrast, CDF plots for cloudless days demonstrate closer alignment between synthesized and measured data. PDF and CDF analysis results are presented in Figs. 13–18.

6. Discussion

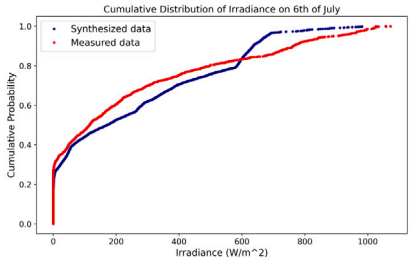
6.1. Practical applications

The primary application of the methodology presented is to improve understanding of rapid fluctuations in solar power output throughout hours, which directly affects PV usage. The focus is on detecting short-term phenomena across different types of days and visualizing them, prioritizing the identification of variability rather than striving for perfect accuracy in representation. The significance of this work lies in introducing a simple yet effective algorithm for reanalyzing existing hourly averaged GHI data. Its key advantage is that it does not require additional financial investment or computational resources, making it accessible for use in locations worldwide that lack high-frequency GHI data, thus eliminating the need for costly research efforts.

This approach can significantly reduce errors in solar output calculations for computational simulations, solar integration in buildings, and direct coupling of PV systems with electrical loads. Typically, the error between hourly and minute resolution modeling of PV output is around 10%, and this algorithm offers a practical solution to dramatically decrease that margin. Applications that demand quick ramp rates, such as coupling PVs with PEMs, stand to benefit significantly. While the focus of this study is on daily performance, the algorithm has demonstrated its effectiveness over longer periods such as weeks and months. Though not delivering the highest precision, it achieves accurate results without unnecessary complexity.

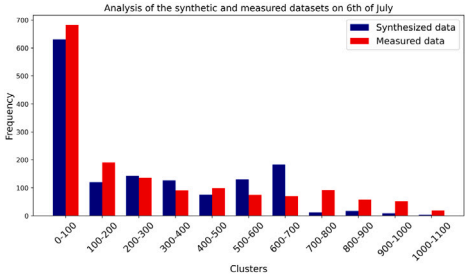


(a) Frequency plot for 6th July

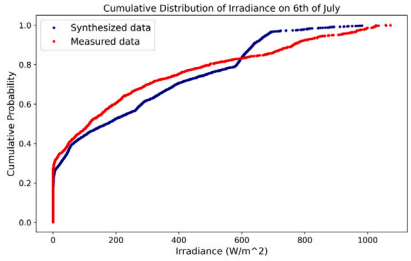


(b) CDF plot for 6th July

Fig. 14. Analysis for 6th July.

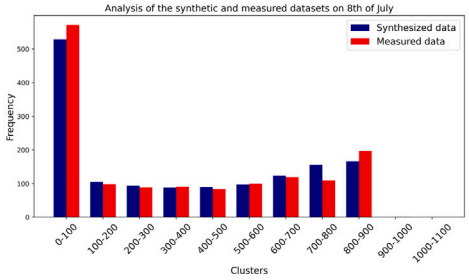


(a) Frequency plot for 6th July

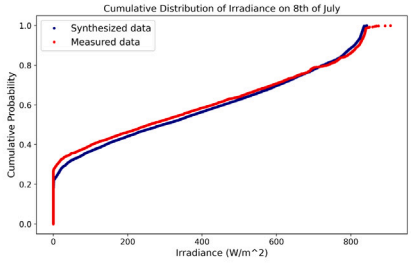


(b) CDF plot for 6th July

Fig. 15. Analysis for 7th July.

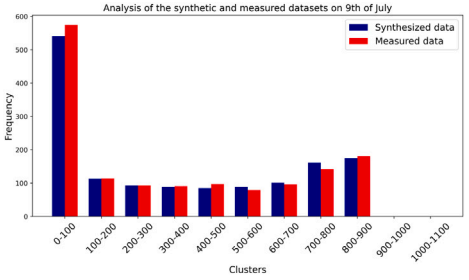


(a) Frequency plot for 8th July

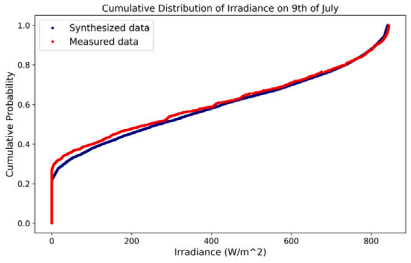


(b) CDF plot for 8th July

Fig. 16. Analysis for 8th July.



(a) Frequency plot for 9th July



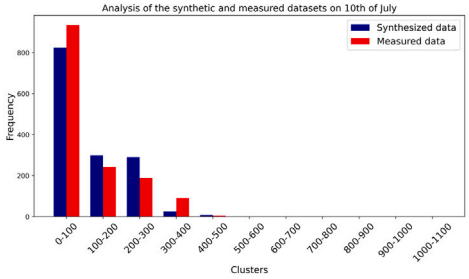
(b) CDF plot for 9th July

Fig. 17. Analysis for 9th July.

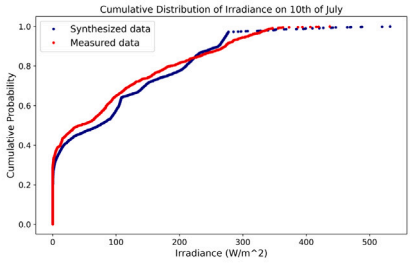
6.2. Future work

Future work will focus on improving the algorithm's robustness and applicability. First, the algorithm will be applied to yearly datasets to assess its performance across extended periods. Validation will also be conducted using a broader range of meteorological data, considering different solar and climatic conditions and locations. This will help determine how well the algorithm performs under diverse environmental factors, especially where solar irradiance patterns vary.

Additionally, testing the algorithm with different renewable energy sources, such as wind, and various measurement instruments will further improve its versatility and robustness. The effects of altering the time resolution within the algorithm will also be examined,



(a) Frequency plot for 10th July



(b) CDF plot for 10th July

Fig. 18. Analysis for 10th July.

comparing system-level outcomes like hydrogen production and overall profitability when using synthesized minute-resolution data versus hourly averaged data.

Finally, while the algorithm shows promise in addressing the limitations of current downscaling methods, particularly its non-probabilistic approach and efficient computation of minute-level data, further optimization is necessary. More research is required to assess the spatial coverage to which this algorithm can be applied, especially across different geographic locations with varying solar irradiance patterns. By refining its parameters and validating its performance across a wide range of datasets, we aim to ensure the algorithm's effectiveness in diverse contexts and its ability to provide reliable, high-resolution solar data in different geographical locations and climates.

7. Conclusion

This work presents a simple but efficient procedure to downscale to minute resolution hourly GHI data. The algorithm suggested in this study effectively synthesizes high-resolution global irradiance time series from hourly averaged datasets, which are freely available in meteorological sources worldwide. It successfully captures short-term fluctuations in solar irradiance in every type of sky coverage by clouds, crucial for accurate simulation and integration of photovoltaic systems in various applications. It combines the advantages of conventional algorithms and adds new elements while using an accurate clear sky hourly estimation, the difference between hourly averaged and clear sky irradiance data, the ramp rate of clearness indexes, and the optimization of factor α that moves the data points from linear interpolation line. It could be demonstrated that with the proposed algorithm, it is possible to synthesize minute values of meteorological data with high statistical quality and realistic temporal variability. For the location Lyngby, Denmark results in the weekly period analysis show promising accuracy, with a normalized mean bias error of 4.6%, a normalized mean square error of 9.4%, and a coefficient value of 0.87, demonstrating strong agreement with real measured data. The algorithm has been tested under conditions with significant fluctuations in GHI, showing its potential for even greater accuracy in locations or periods with more stable irradiance patterns. The primary application of this simple methodology is to improve the modeling of rapid changes in solar power output, without increasing the complexity and computational time, but improving the accuracy of the simulations. Future work will focus on increasing the temporal frequency of GHI series in various locations lacking high-resolution satellite or ground-based measurements.

Nomenclature

Abbreviations

$BSRN$	Baseline Surface Radiation Network
DHI	Diffuse Horizontal Irradiation ($\mathrm{W\,m^{-2}}$)
DNI	Direct Normal Irradiation ($\mathrm{W\,m^{-2}}$)
GHI	Global Horizontal Irradiance ($\mathrm{W\,m^{-2}}$)
MBE	Mean Bias Error
NMBE	Normalized Mean Bias Error (%)
NRMSE	Normalized Root Mean Square Error (%)
PV	Photovoltaic
QC	Quality Control
RMSE	Root Mean Square Error

Symbols

$\bar{k}_{(i,\mathrm{day})}$	Daily average of hourly averaged k_i
d_r	Distance factor for the earth-sun
e_A	Vapor pressure of the air (Pa)
G	Global Horizontal Irradiance ($\mathrm{W\,m^{-2}}$)
G_{clear}	Global clear sky horizontal irradiance ($\mathrm{W\,m^{-2}}$)
G_{i}	Hourly averaged global horizontal irradiance ($\mathrm{W\,m^{-2}}$)
G_{m}	1-minute measured global horizontal irradiance ($\mathrm{W\,m^{-2}}$)
H_{o}	Extraterrestrial solar horizontal irradiance ($\mathrm{W\,m^{-2}}$)
$H_{\mathrm{o,d}}$	Daily extraterrestrial solar horizontal irradiance ($\mathrm{W\,m^{-2}}$)
i	Hours
K_B	Clearness index for direct beam radiation
k_i	Clearness index
K_D	Transmissivity index for diffuse radiation
k_{day}	Daily clearness
n	Day of the year
P	Atmospheric pressure (Pa)
r_i	Ramp rate of the clearness index during the day
R^2	Coefficient of determination
t	Time resolution (minutes or group of minutes)
$\tilde{k}_{(i,\mathrm{day})}$	Variability of clearness index k_i during the day
W	Precipitable water in the atmosphere (mm)

Greek Characters

α	Optimization factor (dimensionless)
β	Solar Altitude Angle (°)
δ	Declination angle (°)
ϕ	Latitude (°)
ω	Hour angle (°)
z	Zenith angle (°)

CRediT authorship contribution statement

Diamantis Almpantis: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Henrik Davidsson:** Writing – review & editing, Supervision, Resources, Data curation. **Martin Andersson:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Diamantis Almpantis reports financial support was provided by LTH. Diamantis Almpantis reports a relationship with Lund University that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors wish to gratefully acknowledge the support of the Horizon Europe Circular Fuel project 101118239, funded by the European Climate, Infrastructure and Environment Executive Agency (CINEA), which funds this research.

References

[1] International Energy Agency (IEA), Renewables 2023- analysis and forecast to 2028, 2024, URL www.iea.org.

[2] G. Bel, M.M. Bandi, Geographic dependence of the solar irradiance spectrum at intermediate to high frequencies, Phys. Rev. Applied 12 (2019) 024032, http://dx.doi.org/10.1103/PhysRevApplied.12.024032, URL http://dx.doi.org/10.1103/PhysRevApplied.12.024032.

[3] A.P. Grantham, et al., Generating synthetic five-minute solar irradiance values from hourly observations, Sol. Energy 147 (2017) 209–221, http://dx.doi.org/10.1016/j.solener.2017.03.026.

[4] J. Polo, et al., A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution, Sol. Energy 85 (2011) 1164–1170, http://dx.doi.org/10.1016/j.solener.2011.03.011.

[5] T. Hirsch, et al., Dynamics of oil-based parabolic trough plants—Impact of transient behaviour on energy yields, in: Proc. SolarPACES Conf., Perpignan, France, 2010, URL https://api.semanticscholar.org/CorpusID:108186430.

[6] A.D. Cronin, et al., Comparing ramp rates from large and small PV systems, and selection of batteries for ramp rate control, 2013, http://dx.doi.org/10.1109/PVSC.2013.6744493.

[7] M.J. Mayer, Effects of the meteorological data resolution and aggregation on the optimal design of photovoltaic power plants, Energy Convers. Manage. 241 (2021) http://dx.doi.org/10.1016/j.enconman.2021.114313.

[8] S. Cao, K. Sirén, Impact of simulation time-resolution on the matching of PV production and household electric demand, Appl. Energy 128 (2014) 192–208, http://dx.doi.org/10.1016/j.apenergy.2014.04.075.

[9] A.V. Klokov, E.Y. Loktionov, Temporal resolution of input weather data strongly affects an off-grid PV system layout and reliability, Solar 3 (2023) 49–61, http://dx.doi.org/10.3390/solar3010004.

[10] L. Bottecchia, et al., Discussing the needs of high resolution data: their impact in evaluating solar potential considering the horizon height, J. Building Perform Simul 16 (2023) 705–716, http://dx.doi.org/10.1080/19401493.2023.2195838.

[11] Martin Hofmann, et al., Improved synthesis of global irradiance with one-minute resolution for PV system simulations, Int. J. Photoenergy 2014 (2014) http://dx.doi.org/10.1155/2014/808509.

[12] P. Vanicek, et al., Quality of the calculation of the inverter losses with standard simulation programs, in: 26yh European Photovoltaic Solar Energy Conference and Exhibition, 2011, pp. 3772–3775, URL https://api.semanticscholar.org/CorpusID:53928469.

[13] Richard Meyer, et al., Towards standardization of CSP yield assessments, 2009, URL https://www.researchgate.net/publication/228412585.

[14] J. Gall, et al., Simulation and control of thermal power plants, Int. Conf. Renew. Energ. Power Quality (2010) 294–298, http://dx.doi.org/10.24084/repqj08.294.

[15] J. Kleissl, Solar Energy Forecasting and Resource Assessment, Academic Press, 2013, URL https://www.sciencedirect.com/book/9780123971777/solar-energy-forecasting-and-resource-assessment.

[16] J.M. Bright, et al., Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data, Sol. Energy 115 (2015) 229–242, http://dx.doi.org/10.1016/j.solener.2015.02.032.

[17] A.P. Grantham, et al., Synthetically interpolated five-minute direct normal irradiance, 2013, pp. 1–6, URL www.mssanz.org.au/modsim2013.

[18] Dean Laslett, et al., A method for generating synthetic hourly solar radiation data for any location in the south west of western Australia, in a world wide web page, Renew. Energy 68 (2014) 87–102, http://dx.doi.org/10.1016/j.renene.2014.01.015.

[19] V.A. Graham, K.G.T. Hollands, A method to generate synthetic hourly solar radiation globally, 44, 1990, pp. 333–341, http://dx.doi.org/10.1016/0038-092X(90)90137-2.

[20] A. Hafzullah, D. Ahmad, Technology data – renewable fuels, J. Hydrol. 562 (2018) 758–779, URL https://ens.dk/en/our-services/technology-catalogues/technology-data-renewable-fuels.

[21] B.O. Ngoko, et al., Synthetic generation of high temporal resolution solar radiation data using Markov models, Sol. Energy 103 (2014) 160–170, http://dx.doi.org/10.1016/j.solener.2014.02.026.

[22] I. Richardson, M. Thomson, Integrated simulation of photovoltaic micro-generation and domestic electricity demand: A one-minute resolution open-source model, 2011, URL https://dspace.lboro.ac.uk/.

[23] A. Mellit, et al., A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach, Sol. Energy 79 (2005) 469–482, http://dx.doi.org/10.1016/j.solener.2004.12.006.

[24] N. Kwarikunda, Z. Chiguvare, Performance analysis of clear sky global horizontal irradiance models: Simple models adapted for local conditions, J. Renew. Energy 2021 (2021) 1–12, http://dx.doi.org/10.1155/2021/4369959.

[25] Energy Systems Integration Group, Assessing the flexibility of green hydrogen in power system models, Energy Systems Integration Group, Reston, VA, 2024, URL https://www.esig.energy/green-hydrogen-in-power-system-models.

[26] Gregor Erbach, Sara Svensson, BRIEFING towards climate neutrality, 2023, URL https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/747085/EPRS_BRI(2023)747085_EN.pdf.

[27] M.A. Giovanniello, et al., The influence of additionality and time-matching requirements on the emissions from grid-connected hydrogen production, Nature Energy 9 (2024) 197–207, http://dx.doi.org/10.1038/s41560-023-01435-0.

[28] Danish Energy Agency, Technology Data - Renewable fuels, 2023, URL https://ens.dk/technologydata.

[29] V. Badescu, A method for determining the solar global and defining the diffuse and beam irradiation on a clear sky, 2018, http://dx.doi.org/10.1007/978-3-540-77455-6_4.

[30] R. Chauvin, A New Approach for Assessing the Clear-Sky Direct Normal Irradiance in Real Time, vol. 40, IEEE, 2020, http://dx.doi.org/10.1109/EEEIC.2016.7555435.

[31] M. Iqbal, An introduction to solar radiation, Department of Mechanical Engineering, The University of British Columbia, Academic Press, Vancouver, British Columbia, Canada, 1983, http://dx.doi.org/10.1016/B978-0-12-373750-2.X5001-0.

[32] J.A. Duffie, W.A. Beckman, Solar Engineering of Thermal Processes, fourth ed., Wiley, 2013, http://dx.doi.org/10.1002/9781118671603, URL https://onlinelibrary.wiley.com/doi/epdf/10.1002/9781118671603.

[33] V. Badescu, Verification of some very simple clear and cloudy sky models to evaluate global solar irradiance, 61, 1997, pp. 251–264, http://dx.doi.org/10.1016/S0038-092X(97)00057-1.

[34] R.G. Allen, et al., The ASCE Standardized Reference Evapotranspiration Equation, American Society of Civil Engineers, 2018, p. 214, http://dx.doi.org/10.1007/978-3-540-77455-6_4.

[35] R.G. Allen, R. Trezza, M. Tasumi, Analytical integrated functions for daily solar radiation on slopes, Agricult. Forest. Meterol. 139 (2006) 55–73, http://dx.doi.org/10.1016/j.agrformet.2006.05.012.

[36] Apogee Instruments, Inc., Clear sky calculator, 2024, https://www.clearskycalculator.com/, (Accessed: 2024-06-13).

[37] Adam R. Jensen, Solar irradiance and weather data, 2024, https://weatherdata.construct.dtu.dk/, (Accessed: 2024-02-20).

[38] N. Geuder, et al., Screening and flagging of solar irradiation and ancillary meteorological data, Energy Procedia 69 (2015) 1989–1998, http://dx.doi.org/10.1016/j.egypro.2015.03.205, URL http://dx.doi.org/10.1016/j.egypro.2015.03.205.

[39] C. Gueymard, Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode radiometers. Part 1: Impacts on global horizontal irradiance, Sol. Energy 153 (2017) 755–765, http://dx.doi.org/10.1016/j.solener.2017.05.004, URL http://dx.doi.org/10.1016/j.solener.2017.05.004.

[40] C. Long, E. Dutton, BSRN global network recommended QC tests, V2.0, Baseline Surface Radiation Network, 2002, URL https://bsrn.awi.de/fileadmin/user_upload/bsrn.awi.de/Publications/BSRN_recommended_QC_tests_V2.pdf.

[41] C. Long, Y. Shi, An automated quality assessment and control algorithm for surface radiation measurements, The Open Atmosp. Sci. J 2 (2008) 23–37.
[42] A. Forstinger, et al., Expert quality control of solar radiation ground data sets, in: Proceedings of SWC 2021: ISES Solar World Congress, International Solar Energy Society, 2021, pp. 1037–1048, http://dx.doi.org/10.18086/swc.2021.38.02.
[43] D. Mayer, D. Butler, Statistical validation, Ecol. Model. 68 (1–2) (1993) 21–32, http://dx.doi.org/10.1016/0304-3800(93)90105-2, URL https://www.sciencedirect.com/science/article/pii/0304380093901052.
[44] Viorel Badescu, et al., Computing global and diffuse solar hourly irradiation on clear sky. Review and testing of 54 models, Renew. Sustain. Energy Rev. 16 (2012) 1636–1656, http://dx.doi.org/10.1016/j.rser.2011.12.010.

Diamantis Almpantis is a Ph.D. candidate at the Department of Energy Sciences at Lund University, specializing in the development of comprehensive energy system models and sector coupling and also actively involved in a European project, Circular Fuels, aimed at optimizing the coupling efficiency between photovoltaic systems and electrolyzers. He holds an MSc in Sustainable Energy with a concentration on energy system analysis and an MSc in Chemical Engineering. His research interests include the modeling and simulation of large-scale energy systems, with a strong focus on hydrogen clusters.

Henrik Davidsson is a lecturer at the department of Building and Environmental Technology. His main research is regarding energy simulations for buildings, ventilation, solar energy, and solar drying. Davidsson has a PhD in solar energy and building technology from Lund University (2014). Henrik has participated in numerous Erasmus+ Capacity Building projects assisting in starting new master programs. Today Henrik teach solar energy, building simulations at Lund University at the master's program in Energy-efficient and Environmental Building Design.

Martin Andersson is a Professor at the department of Energy Sciences, where his main research is on modeling of fuel cells and electrolyzer at various length scales. Andersson has a M.Sc. in Environmental Engineering from Lund University (2007) and a PhD in Heat Transfer from Lund University (2011). Martin was granted a Marie Curie Fellowship in 2015, which enabled him to work as a Guest Professor at the research center Jülich in Germany between 2015 and 2019. Martin has served as Adjunct Professor at University of Electronic Science and Engineering of China (UESTC) in Chengdu, China since 2017. Martin is also the founder and director of the Lund University master's programme in Sustainable Energy Engineering.