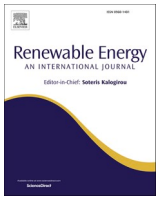
Renewable Energy 256 (2026) 124551

Contents lists available at ScienceDirect

Renewable Energy

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



A high-resolution downscaling approach for solar irradiance using statistical parameter matching

Olalekan Omoyele [a,b,*], Maximilian Hoffmann [a], Jann Michael Weinand [a], Miguel Larrañeta [c], Jochen Linßen [a], Detlef Stolten [a,b]

[a] Forschungszentrum Jülich GmbH, Institute of Climate and Energy Systems – Jülich Systems Analysis (ICE-2), 52425, Jülich, Germany
[b] RWTH Aachen University, Chair for Fuel Cells, Faculty of Mechanical Engineering, 52062, Aachen, Germany
[c] Departamento de Ingeniería Energética Universidad de Sevilla, 4 San Fernando Str, Seville, Spain

ARTICLE INFO

Keywords:
Solar irradiance
Downscaling
Clearness index
Variability index
Temporal resolution
Energy system modeling

ABSTRACT

The limited intra-hour variability of globally available hourly renewable energy system data leads to inaccuracies in the modeling of renewable energy systems. While sub-hourly data can improve model accuracy, such data are not globally available. The existing approaches to increase the temporal resolution of solar irradiance often rely on site specific measurements or complex models, limiting global scalability. This work, therefore, presents a methodology to increase the temporal resolution of the global horizontal irradiance from 1 h to 1 min using non-dimensional irradiance and parameters matching based on daily irradiance characteristics for arbitrary locations. The methodology is validated using statistical methods and energy system optimization. The hourly annual normalized root mean square error and Kolmogorov-Smirnov Integral range from 5 to 7 % and 0.1–0.7, respectively, for different locations consisting of varying weather conditions. The energy system optimization results of the synthetic data demonstrate superiority in terms of cost and feasibility relative to the average hourly resolution data. The use of synthetic minute resolution data significantly improves the design accuracy of dynamic components such as inverters and storage systems. The globally applicable method, based on Köppen-Geiger classification coverage, will enable more reliable energy systems modeling in the future.

1. Introduction

High-resolution solar irradiance data has gained importance in the last decades, primarily due to the growing incorporation of renewable energy sources into the energy mix [1]. The prevalence of hourly resolved measured, reanalyzed, and predicted data, along with the disproportionate increase in model complexity at higher resolution, has prompted numerous researchers to use it for modeling purposes. However, intra-hourly fluctuations, which are not captured in the prevailing hourly resolutions [2], may have severe impacts on future grid stability. This issue becomes even more critical as the balancing power provided by rotating masses from conventional power plants declines, increasing the risk of grid imbalances [3]. Therefore, modeling in high temporal resolution is imperative in applications such as photovoltaic system design, solar energy forecasting, and grid management [4].

Previous research [5] investigated the impact of sub-hourly resolution on energy system modeling, and showed that sub-hourly resolution is important for policy analysis, electric sector planning, and technology valuation [6]. Furthermore, it was demonstrated that coarser temporal resolution leads to underestimation of total annualized cost (TAC), with a discrepancy of up to 2 % between hourly and minutely resolutions [5,7,8], generator cycling and flexible generation [9], energy storage capacity and utilization [9–11], ramping [12], inverter clipping losses [13–15], and the levelized cost of electricity [16,17], among others. It is important to note that modeling at hourly resolution can also result in an infeasible system design. Particularly in the context of renewable energy systems, the undersizing of dynamically operating components such as inverters and batteries is an issue when relying on coarse temporal resolution data [5]. While the main drivers for the use of coarser resolutions have been data scarcity and computational complexity, the latter

* Corresponding author. Forschungszentrum Jülich GmbH, Institute of Climate and Energy Systems – Jülich Systems Analysis (ICE-2), 52425, Jülich, Germany.
E-mail address: o.omoyele@fz-juelich.de (O. Omoyele).

https://doi.org/10.1016/j.renene.2025.124551
Received 23 April 2025; Received in revised form 9 August 2025; Accepted 15 September 2025
Available online 1 October 2025

Table 1
Methods of increasing the temporal resolution of solar irradiance.

Downscaling Approach	Description	Characteristics	
		Strength	Limitation
Deterministic	Uses statistical interpolation methods.	- Computationally cheap	- Limited accuracy - Lacks comprehension
Stochastic	Applies randomness to generate high-frequency data	- Comprehensive - Good accuracy	- Computationally expensive - Can produce unrealistic patterns
Markov	Applies probability transitions between irradiance states.	- Comprehensive - Good accuracy	- Computationally expensive - Complicated transitions
Machine Learning	Learns patterns from historical data (e.g. neural networks)	- Comprehensive	- Requires large training datasets - Poor generalizability - Computationally expensive
Non-dimensional	Matches low resolution to normalized high resolution profiles using daily statistical indicators	- Comprehensive - Adaptable to arbitrary locations - Good accuracy - Computationally cheap	- Limited by database coverage

can be mitigated by averaging time series sub-samples [18], time series aggregation [19,20] and parallelization [20].

Several methods have been used so far to increase the temporal resolution of renewable energy data, particularly solar irradiance. The approaches for solar irradiance can be classified into Markov, deterministic, stochastic, non-dimensional, and machine learning methods [21]. As illustrated in Table 1, a comparison is presented of the various methods, with their description and characteristics. The deterministic approach utilizes statistical interpolation methods, which frequently underperform in accurately capturing the intra-hour fluctuations of the irradiation. The stochastic approaches frequently employ a deterministic approach as a foundation, yet prior to incorporating randomization into the interpolations, they often demonstrate superior accuracy in comparison to deterministic approaches. However, it should be noted that the accuracy of these stochastic approaches is significantly influenced by the employed deterministic approach. The stochastic approach has the capacity to engender a highly complex system through the introduction of the random variable. The Markov approach has demonstrated efficacy in capturing the sub-hourly fluctuations. However, this efficacy is contingent upon the complexity stemming from the order of the Markov chain that links the data dependencies. Consequently, several Markov approaches employ first-order Markov chains, which are incapable of fully capturing data dependencies, thereby resulting in limited accuracy. The machine learning approaches constitute a distinct class of approach. Although their applications in this context remain limited, they offer considerable potential and, with access to sufficient high-quality training data, could have improved performance. In comparison to alternative methods, the non-dimensional approach is characterized as explainable and comprehensive, uses real data, is adaptable to arbitrary locations with good accuracy, and has no modular error at a low computational expense. The non-dimensional approach downscales low-resolution data to high-resolution by normalizing solar irradiance and time (see details in Box 1 and section 2.2).

Box 1: The non-dimensional approach

For downscaling low-resolution data to high-resolution, the non-dimensional approach normalizes the time between the daily sunrise and the sunset [22], whereas the normalized irradiance is the ratio of the irradiance to its corresponding extraterrestrial or clear sky values [22]. The extraterrestrial irradiance is taken as the theoretical irradiance at the Earth's upper atmosphere [23]. The clear-sky value is defined as the amount of irradiance that reaches the Earth's surface when the effect of cloud cover is not taken into account [24]. Previous articles suggested to use clear sky irradiance when modeling or forecasting direct normal irradiance (DNI) and extraterrestrial irradiance when modeling or forecasting global horizontal irradiance (GHI) [25]. Several models have been developed to capture the clear sky solar irradiance data [26,27], such as curve fitting and the highly reviewed and benchmarked REST2V5 [28], MACC2 [29], and McClear models [30], among others. Given the inherent similarity between data from the same climate class, a minimum of one year of comparable data for a location intended for downscaling is necessary for the database of normalized, non-dimensional irradiance profiles [31]. Consequently, two databases are created, comprising the non-dimensional curves in minutes and the daily data. The daily parameters which represent the minutely non-dimensional curves in the database can be the clearness index (k_d), the variability index (VI), and the distribution (F_m). An exposition of these three parameters can be found in section 2.2.2.2. The hourly resolution to be downscaled produces the daily parameters (k_d, VI, and F_m) of the nature in the database. The most similar day from the database to each day of the hourly resolution to be downscaled is selected and processed back to irradiance to provide highly resolved downscaled data. A detailed explanation of the non-dimensional approach is provided in section 2.2.

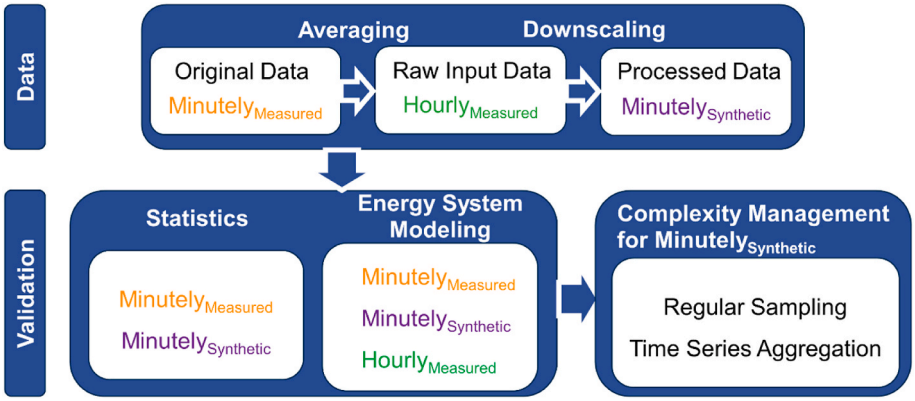


Fig. 1. The workflow of the methodology developed in this study.

To the best of the authors' knowledge, the non-dimensional approach was first developed by Peruchena et al. [22], who obtained DNI data in minutely resolution for a location from its hourly resolution using curve fitting for a clear sky model. Meanwhile, the used training data required a 1-min resolution of a location which confirmed its locality. The authors further applied the methodology to different climatic zones [22,32–36]. Larrañeta et al. [31,37,38] improved the methodology of Peruchena et al. [22] by applying Köppen-Geiger weather classification[1] [39] to categorize the database into different climate zones and matching the parameters of a location to be downscaled to a similar location with the same Köppen-Geiger weather classification. Larrañeta et al. also improved the daily parameters from the k_d as utilized by Peruchena et al. [22] to include the VI and the F_m [40]. Furthermore, the matching between similar days to be downscaled and the used database was improved by utilizing the k-nearest neighbors instead of the Euclidean distance, cutting down the computational time. Larrañeta et al. [31] also developed the ND tool [41] which downscales DNI or coupled DNI + GHI from 1 h to 1 min resolution and was used for application studies in Refs. [36,42].

Although several approaches (as presented in Table 1) have been proposed by numerous studies to downscale GHI from hourly to sub-hourly resolutions [43–45], the non-dimensional approach is outstanding in terms of high accuracy, easy adaptability, and low complexity. Most existing approaches are either deficient in accurately capturing sub-hourly fluctuation, site-specific or lack generalizability across diverse climatic conditions. Furthermore, the current global tool, ND tool, which is based on the non-dimensional approach has been predominantly developed for the DNI or coupled DNI + GHI, with no robust standalone implementation for the GHI. Moreover, the ND tool relies on the clear-sky index, which is utilized for DNI without distinctly adapting the GHI dynamics. The objective of this study is to develop a more robust, parameter-based non-dimensional method to downscale hourly solar GHI data to a 1-min resolution using a comprehensive database of non-dimensional curves and defining parameters, to enable accurate globally applicable data. Furthermore, the downscaled synthetic dataset is validated using an energy system model. Methodologies that reduce the complexity of energy system modeling are employed to mitigate complexity arising from sub-hourly resolution data from the model. Fig. 1 illustrates the workflow of this study, which is divided into two sections:

1) **Data:** The data set encompasses the measured 1-min resolution data, which is converted to hourly resolution (by averaging). The non-dimensional methodology is then applied to downscale the measured hourly resolution data to 1-min synthetic data
2) **Validation:** The validation section is employed to assess the accuracy of the synthetic 1-min data in comparison to the 1-min measured data and 1-h measured data. The validation approaches employed encompass statistical methods and energy system modeling. As sub-hourly resolution can lead to increased complexity, approaches for complexity management in regular samplings and clustering-based time series aggregations are investigated.

The remainder of this work is structured as follows: Section 2 describes the used data, its preprocessing steps, and the proposed methodology. The methodology encompasses the database constructions to the downscaling procedures, as well as the validation and performance metrics of the profile's statistics and their application to the energy system model. Section 3 presents the result of both the statistical and energy system modeling validation of the methodology, which are discussed in section 4. Finally, section 5 concludes this study.

Table 2
High-resolution (1-min) irradiance data from ground-based measurements.

City	Latitude (°)	Longitude (°)	Height (m)	Köppen-Geiger Climate[2]	Years
Adelaide	−34.95	138.52	2	Csb	14
Alice Springs	−23.80	133.89	546	BWh	14
Broome	−17.95	122.24	7.42	BSh	14
Cape Grim	−40.68	144.69	95	Cfb	11
Cobar	−31.48	145.83	260	BSh	1
Cocos Island	−12.19	96.83	3	Ocean/Cfb	10
Darwin	−12.42	130.89	30.4	Aw	13
Geraldton Airport	−28.80	114.70	29.7	Csa	6
Geraldton Airport Comp.	−28.80	114.70	33	Csa	2
Kalgoorlie-Boulder	−30.78	121.45	365.3	BSh	4
Learmonth	−22.24	114.10	5	BWh	5
Melbourne	−37.66	144.83	113.4	Cfb	12
Mildura	−34.24	142.09	50	BSk	2
Mt Gambia	−37.75	140.77	63	Csb	2
Rockhampton	−23.38	150.48	10.4	Cfa	10
Tennant_Creek	−19.64	134.18	377.1	BSh	2
Townsville	−19.25	146.77	4.34	Aw	4
Wagga	−35.16	147.46	212	Cfa	13

[1] The Köppen-Geiger classification is divided into Major groups [Precipitation (Temperature)] as follows: *A* – *tropical* [f – Rainforest, m – Monsoon, w – Savanah], *B* – *Arid* [W – Desert (h – Hot, k – Cold), S – Steppe (h – Hot, k – Cold)], *C* – *Temperate* [s – Dry summer (a – Hot summer, b – Warm summer, c – Cold summer), w – Dry winter (a – Hot summer, b – Warm summer, c – Cold summer), f – Without dry season (a – Hot summer, b – Warm summer, c – Cold summer)], *D* – *Cold* [s – Dry summer (a – Hot summer, b – Warm summer, c – Cold summer), w – Dry winter (a – Hot summer, b – Warm summer, c – Cold summer), f – Without dry season (a – Hot summer, b – Warm summer, c – Cold summer)], *E* – *Polar* [T – Tundra, F – Frost].

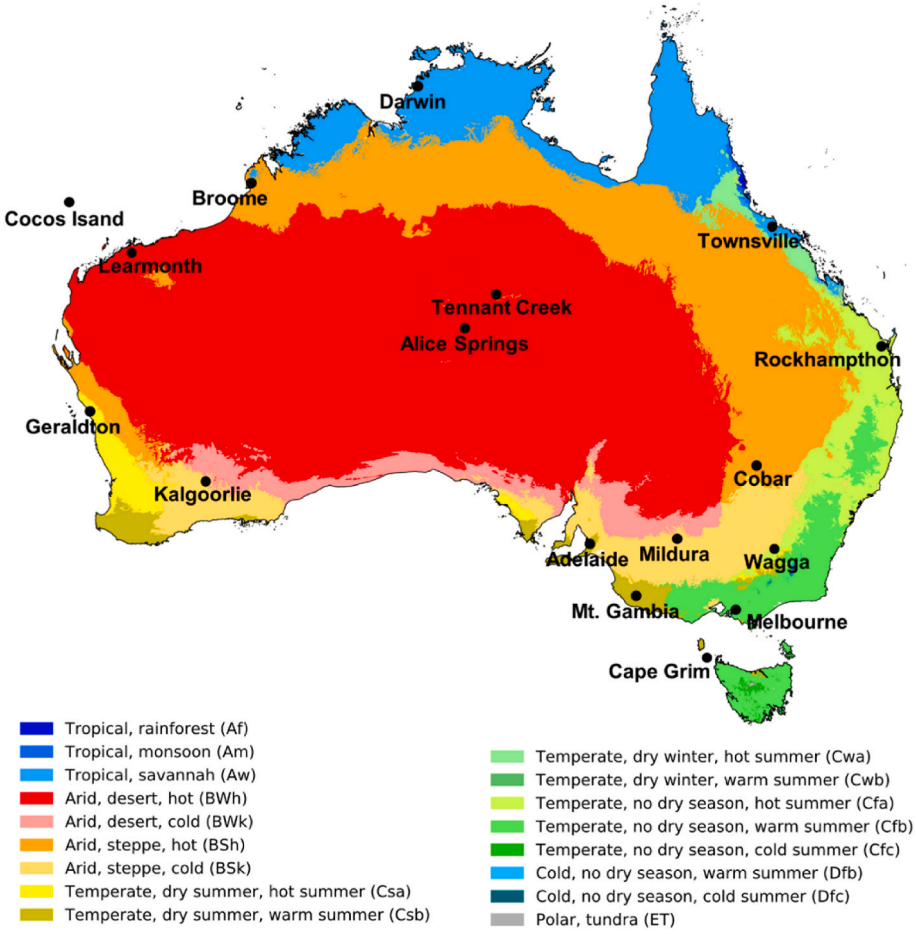


Fig. 2. Köppen-Geiger weather classification of Australia for the selected sites for the non-dimensional database construction [39].

Table 3
Correlated Köppen-Geiger climate.

Location	Number of Days	Correlated Köppen-Geiger Climate[2]	Köppen-Geiger Climate Represented[2]
Darwin, Townsville	4531	Aw	Aw, Am, Af, As
Broome, Cobar, Kalgoorlie Boulder, Tennant Creek, Mildura	6552	Bsh	Bsh, Bsk
Alice Springs, Learmonth	6942	Bwh	Bwh, Bwk
Rockhampton, Wagga	7123	Cfa	Cfa
Cape Grim, Melbourne,	6507	Cfb	Cfb, Cfc
Geraldton, Geraldton Airport	2453	Csa	Csa
Adelaide, Mt Gambia	5150	Csb	Csb, Csc, Cwa, Cwb, Cwc
Rockhampton, Wagga, Cape Grim, Melbourne, Cocos Island	15527	Cfa/Cfb/Ocean	Dfa, Dfb, Dfc, Dsa, Dsb, Dsc, Dwa, Dwb, Dwc, EF, ET

Table 4
Validation irradiance data measurements.

Location	Latitude (°)	Longitude (°)	Height (m)	Köppen-Geiger Climate	Year
Berlin, Germany	52.46	13.52	34	Cfb	2017 2018 2019 2020
Milan, Italy	45.50	9.16	120	Cfa	2017
Tamanrasset, Algeria	22.79	5.53	1385	BWh	2009
Tateno, Japan	36.06	140.13	25	Cfa	2013
Toravere, Estonia	58.24	26.46	70	Dfb	2008

2. Data and methodology

2.1. Data description

This section presents the data used for the developed methodology. Section 2.1.1 describes the data source for database construction, encompassing diverse climatic conditions classified according to the Köppen-Geiger system and multiple years of data. Section 2.1.2 details the data employed for the validation of the methodology, including various weather classifications and conditions.

2.1.1. Data for database construction

The 1-min resolution data in Table 2, which was utilized for database construction, was obtained from the Australian Bureau of Meteorology.[2] Fig. 2 provides the geographic distribution and climate diversity of the selected cities of Australia, encompassing diverse Köppen-Geiger climate zones. As the McClear clear sky model [30] includes data starting from 2004, this is the earliest year represented in the database and hence the earliest downscaling year of the methodology.

[2] https://reg.bom.gov.au/climate/reg/oneminsolar/index.shtml.

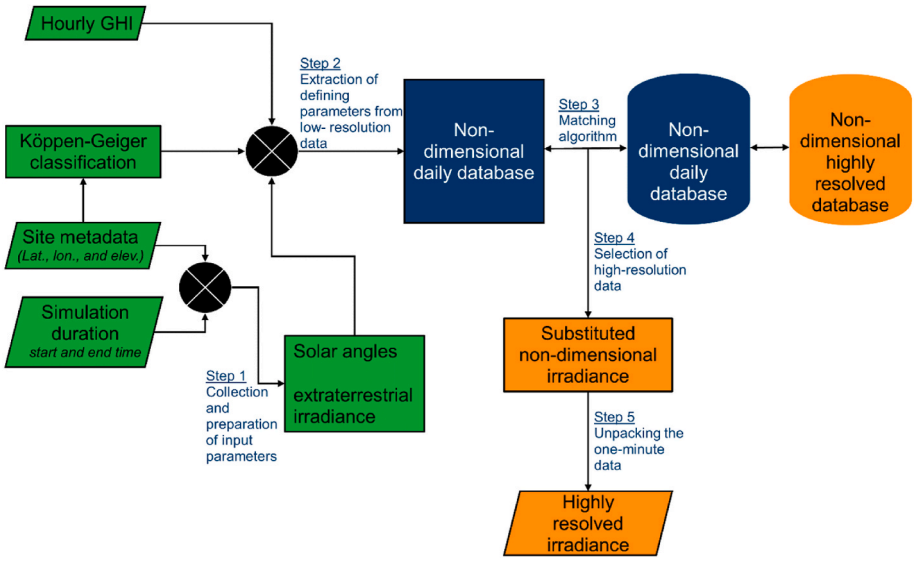


Fig. 3. Flow chart of the non-dimensional methodology of this study.

Consequently, the non-dimensional database spans from 2004 to 2021. The GHI measurements obtained from the Australian Bureau of Meteorology were derived from CM-11 pyranometers manufactured by the company Kipp&Zonen. The Australian Bureau of Meteorology adheres to the World Meteorological Organization-approved standard of the 'Alternate Method' [46] for reporting and calibration. The GHI data undergoes quality control procedures by pvanalytics [47], and any days meeting the following criteria are excluded from the analysis:

- Measurements with night values
- GHI data with values less than 0 (negative data), or greater than the solar constant of 1361.1W/m^2 [48]
- Data with gaps or stale data
- Data with any measurement errors or unavailable values

Since Table 2 does not contain all the Köppen-Geiger climate zones, a correlated Köppen-Geiger climate is provided as shown in Table 3 to represent all the Köppen-Geiger climates [41] in the database. The final database encompasses 50,749 days from the 18 different locations in Table 2, with each day having 1000 non-dimensional data points, yielding 41,155 days after the quality control (Table 3), retaining 81.1 % of the original data in Table 2.

2.1.2. Data for validation

The validation data (see Table 4) is obtained from open sources, including Berlin [49], Milan [50], and Baseline Surface Radiation Network (BSRN) locations of Tamanrasset, Tateno, and Toravere [51]. The Berlin data was measured with the combination of SP-Lite2, CMP11, and SMP21 pyranometers of Kipp&Zonen using weather data of HTW Berlin. Data preprocessing, including the handling of missing values, was carried out by HTW Berlin. The missing values were imputed or replaced depending on the data gap. For gaps shorter than 1 h, linear interpolation was used for imputation. For gaps exceeding 1 h, missing values were replaced using data from the previous hour or the same hour of the previous day. The Milan data of 2017 has about 1 % of missing data points. Of these missing data points, two complete days (days 115 and 116) are replaced by their previous days and other days with missing data are imputed using k-nearest neighbors as suggested by Mantuano et al. [52]. The BSRN locations of Tamanrasset, Tateno, and Toravere were measured with Pyranometer, Eppley, PSP, SN 30123 F3, WRMC No. 42001; Pyranometer, Kipp & Zonen, CMP21, SN 090229, WRMC No. 16035; and Pyranometer, Kipp & Zonen, CM11, SN 903301, WRMC No. 9005, respectively. The Tamanrasset and Toravere have 8 and 3 missing days, respectively, with the missing days being replaced by their previous days. The locations are selected to represent diverse Köppen-Geiger climates, demonstrating the methodology's applicability across a wide range of global conditions.

2.2. Methods

Fig. 3 below shows the flow chart of the developed methodology. Two databases are used: Firstly, the non-dimensional minutely resolved solar GHI against the non-dimensional time (*non-dimensional highly resolved database* or *database 1*). Secondly, the non-dimensional daily matching parameters (*non-dimensional daily database* or *database 2*) containing the k_d, VI, normalized variability index (NVI), F_m, and integrated complementary cumulative distribution function (ICCDF) for each day of the minutely resolved solar GHI in database 1. The exposition of database 1 and database 2 is found in sections 2.2.2.1 and 2.2.2.2, respectively. The 1-h resolution (green color) data to be downscaled requires the input parameters of the hourly resolved GHI, simulation duration (start and end time), and the site metadata (latitude, longitude, and elevation of the location). From these, the daily matching parameters (blue color) are calculated and matched with database 2, and the closest day to each day using the k-nearest neighbor algorithm is selected. The corresponding high-resolution data (orange color) to the selected day is substituted to the hour resolution.

The methodology section comprises the downscaling procedure (section 2.2.1), the database construction (section 2.2.2), as well as the validation and performance metrics split into statistics and energy system modeling (section 2.2.3).

2.2.1. Downscaling procedure

The developed downscaling procedure includes five steps, which are explained in the following.

- Step 1: Collection and Preparation of Input Parameters.

The input data to be downscaled is collected, extracted, and processed. This includes the simulation duration, which states the start and end time of the downscaling process, and the site metadata (latitude, longitude, and altitude) through which the solar angles from pvlib [53] and Köppen-Geiger weather class are obtained. The combination of the simulation duration and the site metadata produces the extraterrestrial irradiance from the McClear clear sky model.

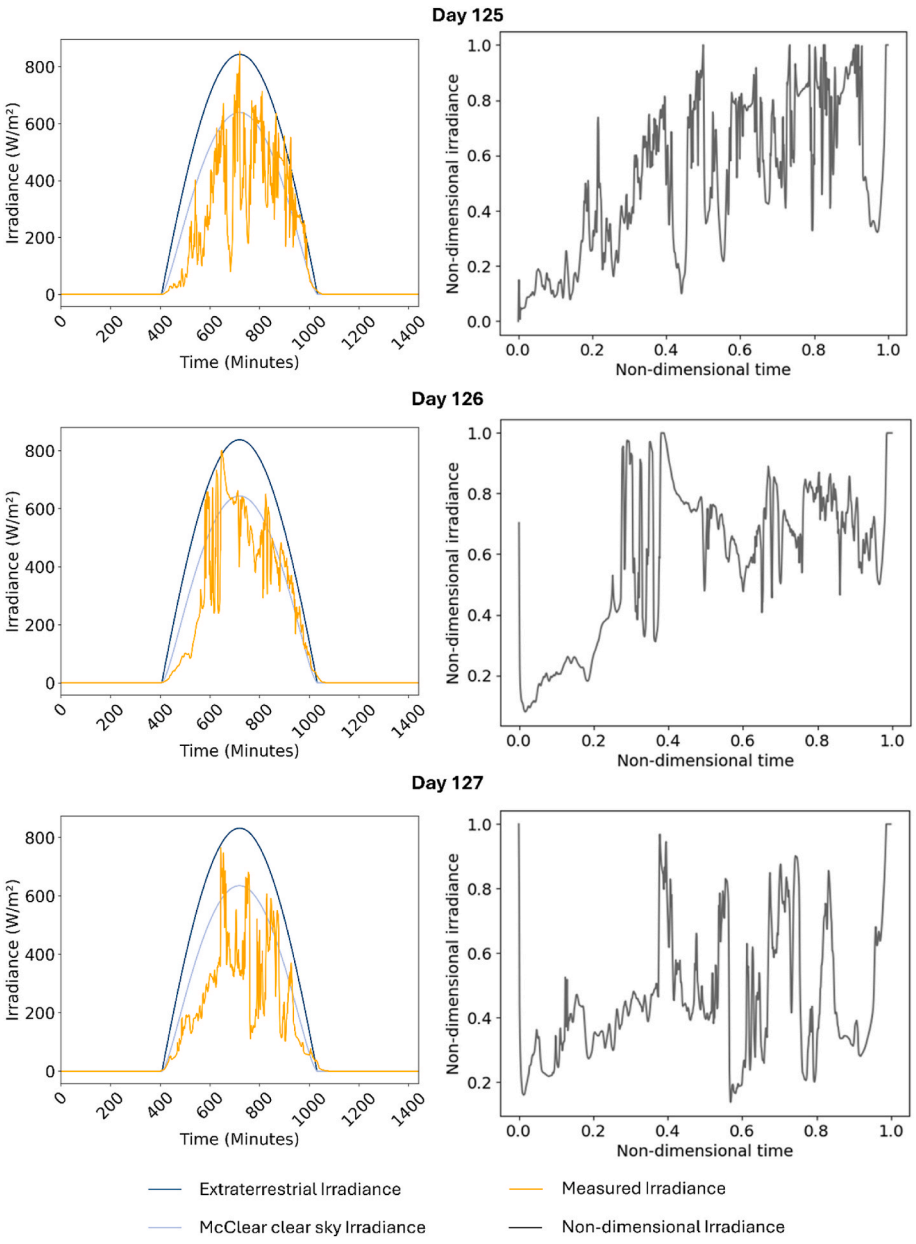


Fig. 4. Non-dimensional database construction consisting of daily non-dimensional irradiance over non-dimensional time.

- Step 2: Extraction of Defining Parameters from Low-Resolution Data

The defining parameters (k_d, VI, NVI, F_m, and ICCDF) are calculated for each day from hourly resolution data to be downscaled. These parameters are detailed in section 2.2.2.2. The Köppen-Geiger weather class is also obtained to determine the closest weather class to the defining parameters.

- Step 3: Matching Algorithm

The daily defining parameters obtained in step 2 are compared against database 2, which contains the corresponding daily defined minutely resolved profiles in the same correlated Köppen-Geiger weather class (see Table 3). The comparison is performed using the k-nearest neighbor machine learning algorithm with one neighbor, enabling the selection of the single most similar day matching. The matching is performed using the five daily indicators elaborated in section 2.2.2.2. The Euclidean distance is used to determine similarity in the five-dimensional feature space. The day in the database with the smallest distance to the current day's indicators is selected as the best match.

- Step 4: Selection of High-Resolution Data

The best match from step 3 and the corresponding highly resolved data from database 1 is selected. The selected highly resolved profile is non-dimensional and needs to be converted to a dimensional solar irradiance in W/m^2.

- Step 5: Unpacking the 1 Minute Data

The specific time of the day between sunrise and sunset is obtained

for each day. The highly resolved data from step 4 is then distributed over the daytime to create the non-dimensional irradiance. This, in turn, is multiplied by the 1-min extraterrestrial irradiance and a constant daily value *k*, to create a 1-min resolution synthetic GHI. The value of *k* is optimized for each day such that the difference between the daily sum of measured data and the daily sum of the synthetic data is minimized.

2.2.1.1. Optimization of k factor for daily Cummulative

$$min\left(GHI_{measured} - GHI_{syn}/60\right) \tag{1}$$

$$GHI_{syn} = k \times k_{nd} \times I_o \tag{2}$$

where, $GHI_{measured}$ is the measured irradiance in hourly resolution, GHI_{syn} is the synthetic irradiance in minutely resolution, k is the daily scaling factor, k_{nd} is the non-dimensional irradiance, I_o is the extraterrestrial irradiance.

A summary of the downscaling procedure, consisting of the steps, descriptions, inputs, and outputs, is presented in Table 6 of Appendix A.

2.2.2. Database construction

The two databases explained in the following sections 2.2.2.1 and 2.2.2.2 are constructed and categorized based on their Köppen-Geiger weather conditions.

2.2.2.1. Database 1 with one minute resolution data. The non-dimensional highly resolved database comprises non-dimensional GHI and time. The database 1 comprises several daily profiles of minutely non-dimensional irradiance against the non-dimensional time as shown in Fig. 4, which presents three consecutive days in 2004 for Adelaide as shown in Table 2. As evident in Fig. 4, the extraterrestrial irradiance provides a better GHI envelope as compared to the clear sky irradiance, hence its adoption for GHI, irrespective of the cloudiness of the location. For further illustrations, Fig. 14 in Appendix B presents the plots of measured, clear sky and extraterrestrial irradiance in a polluted climate of Delhi, India (latitude = 28.58°, longitude = 77.45°, altitude = 207m). It is evident that even though this location is prone to high pollution, the clear sky irradiance envelope does not accurately capture the variability of the measured irradiance as compared to the extraterrestrial irradiance. Therefore, the extraterrestrial irradiance which is determined solely by astronomical factors such as the Earth's distance from the sun and the solar zenith angle, is considered instead of the clear sky irradiance for GHI [25]. The clear sky irradiance instead depends on the local atmospheric conditions such as water vapor, aerosol levels, and surface albedo, and therefore limits the GHI due to the air mass passage of the irradiance [25]. The non-dimensional solar irradiance is therefore the ratio of the measured irradiation to the extraterrestrial irradiance, while the non-dimensional time is the normalization of the daily time between sunrise and sunset. The normalization workflow involving the creation of this database is presented in Fig. 15 of Appendix C.

2.2.2.2. Database 2 with daily parameter aggregates. Database 2 contains the daily profiles of the defining parameters for the minutely resolved non-dimensional database in section 2.2.2.1. The defining parameters are clearness index k_d, VI, NVI, F_m, and ICCDF. These parameters, which are elaborated in the following, are both purely statistical and geographically influenced:

- Clearness Index:

The clearness index, k_d, is the ratio of the measured irradiance of a location to its corresponding extraterrestrial irradiance [25]. Mathematically, k_d is expressed as shown in Equation (3).

$$k_d = \frac{H}{H_o} \tag{3}$$

Here, k_d is the daily clear sky index, H is the daily measured irradiance, and H_o is the daily extraterrestrial irradiance from the McClear model [30]. The k_d indicates the cloudiness of the day, and ranges from 0 (cloudy or overcast atmospheric conditions) to 1 (clear atmospheric conditions).

- Variability Index:

The variability index, VI, is the ratio of the length of the variations of the daily measured irradiance of locations to its corresponding extraterrestrial irradiance [54]. The VI provides information about the variability of the day and is quantified by Equation (4).

$$VI = \frac{\sum_{k=2}^{n}\sqrt{\left(H_k - H_{k-1}\right)^2 + \Delta t^2}}{\sum_{k=2}^{n}\sqrt{\left(H_{o,k} - H_{o,k-1}\right)^2 + \Delta t^2}} \tag{4}$$

Here, VI is the variability index, H_k and H_{k-1} are measured irradiance at time steps k and k-1, respectively, $H_{o,k}$ and $H_{o,k-1}$ are measured extraterrestrial irradiance at k and k-1, respectively.

- Normalized Variability Index:

The concept of the normalized variability index, NVI, was proposed by Moreno-Tejera et al. [55]. The VI is not purely statistical as it depends on the time of the year and geographical location, but the NVI is a statistical approach which normalizes the irradiance without considering its corresponding atmospheric conditions. With the NVI, the statistical variability of a location can be assessed without atmospheric influence. The NVI (see Equation (5)) is the ratio of the length of the variations of the daily irradiance of a location to the length of the maximum variability of the daily profile.

$$NVI = \frac{\sum_{k=2}^{n}\sqrt{\left(H_k - H_{k-1}\right)^2 + \Delta t^2}}{\sum_{k=2}^{n}\sqrt{\left(H_{max,k} - H_{max,k-1}\right)^2 + \Delta t^2}} \tag{5}$$

Here, NVI is the normalized variability index, H_k and H_{k-1} are measured irradiance at time steps k and k-1, respectively, $H_{max,k}$ and $H_{max,k-1}$ are maximum measured irradiance at k and k-1, respectively.

- Distribution:

The distribution, F_m (see Equation (6)), is the ratio of the total morning fraction of the irradiance to the total daily irradiance [31].

$$F_m = \frac{H_{mn}}{H_T} \tag{6}$$

Here, F_m is the distribution, H_{mn} is the total morning fraction of the irradiance, and H_T is the total daily irradiance. The total morning fraction is obtained by extracting the irradiance value when the hour angle[3] is below zero. Analysis of the distribution fraction allows for the quantification of the diurnal radiation profile, specifically assessing the concentration of irradiance between the pre-and post-solar noon intervals.

- Integrated Complementary Cumulative Distribution Function:

The integrated complementary cumulative distribution function, ICCDF (see Equation (7)), is another statistical parameter which provides the overall picture of the solar irradiance variability. The ICCDF is

[3] The hour angle is the angular displacement of the Sun from its position at solar noon, measured relative to the local solar time.

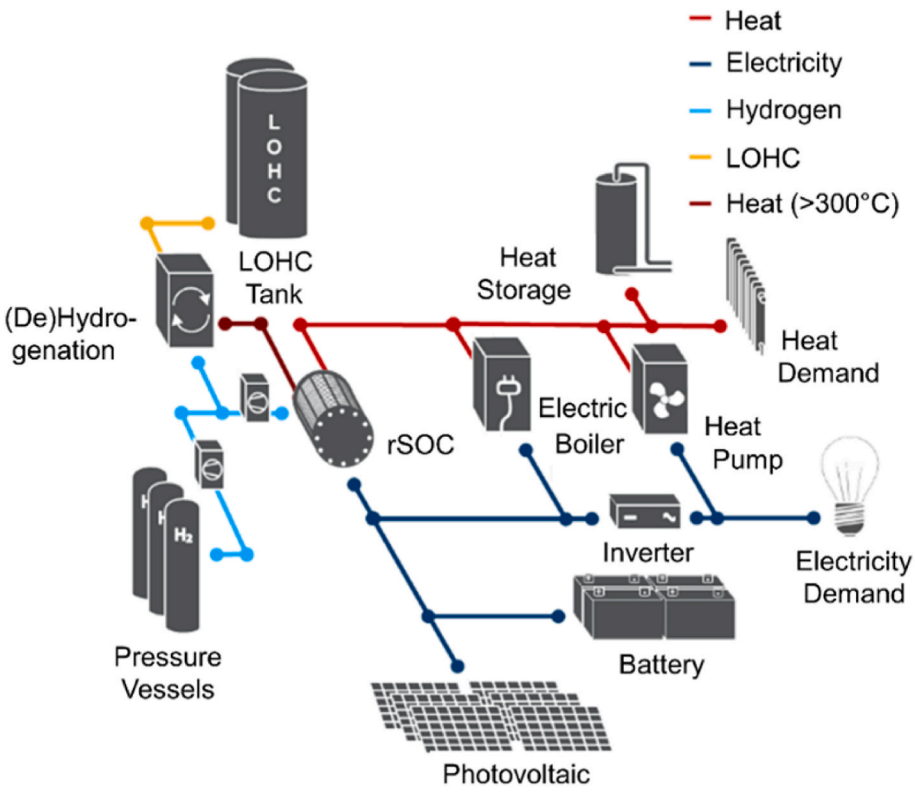


Fig. 5. The self-sufficient building model as presented by Knosala et al. [60]. The rSOC is a reversible solid oxide cell and the LOHC is a liquid organic hydrogen carrier.

calculated as the area under the complementary cumulative distribution function daily curve [56]. Mathematically, the ICCDF is calculated below:

$$ICCDF = \int_{H_{min}}^{H_{max}} CCDF(H)\, dH \tag{7}$$

Here, ICCDF is the integrated complementary cumulative distribution function, and CCDF is the complementary cumulative distribution function. The CCDF is 1 – CDF, where CDF is the cumulative distribution function. The H_{min} and H_{max} represent the minimum and the maximum values of the irradiance, respectively.

2.2.3. Validation and performance metrics

To validate the developed model for downscaling solar irradiance, three different validation metrics are considered, namely the normalized root mean square error (NRMSE), the Kolmogorov-Smirnov Integral test (KSI), and an energy system model, to which the synthesized minutely resolved data is applied.

- Normalized Root Mean Square Error (NRMSE)

The root mean square error is the measure of the root of the squared errors between synthetic data and the original measured data [21]. The NRMSE is the root mean square error that is normalized between the minimum and the maximum values of the measured data. The lower the value of the NRMSE, the more accurate the synthetic data is, relative to the measured data. Equation (8) below gives the mathematical representation of the NRMSE.

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{y_{max} - y_{min}} \tag{8}$$

Here, y_i is the synthetic data, $\hat{y}_i$ is the measured data, n is the number of data points, and y_{min} and y_{max} are the minimum and maximum values of measured data, respectively.

- Kolmogorov-Smirnov Integral Test

The Kolmogorov-Smirnov Integral (KSI) Test is used for comparing the statistical distribution of downscaled data with actual measurements (see Equations (9)–(11)) (KSI) [31].

$$KSI = \frac{\int_{x_{min}}^{x_{max}} |F(x) - G(x)|dx}{a_{critical}} \tag{9}$$

$$a_{critical} = V_c(x_{max} - x_{min}) \tag{10}$$

$$V_c = \frac{1.63}{\sqrt{n}}; \quad n \geq 35 \tag{11}$$

Here, F(x) is the synthetic data, G(x) is the measured data, x_{min} and x_{max} are the minimum and the maximum values of the measured data, and n is the number of data points.

- Energy System Modeling

As discussed by Mantuano et al. [52], highly resolved renewable time series data is increasingly applied to energy system models. Therefore, synthesized minutely resolved time series may also be evaluated by determining whether they yield similar results to measured data when applied to energy system models. For that sake, we consider a self-sufficient building energy system model in this study (see Fig. 5). Energy self-sufficiency means that the building is off-grid and powered by its own renewable energy system [57]. The model optimizes the

Table 5
Validations of different locations. Due to a lack of data availability, the inter-annual uncertainty analysis could only be performed for Berlin.

Location	Köppen-Geiger Climate	Year	NRMSE Minute (%)	NRMSE Hour (%)	KSI Hour (%)
Berlin, Germany	Cfb	2017	9.11	7.45	0.12
		2018	8.32	6.95	0.11
		2019	8.33	6.97	0.11
		2020	8.80	6.71	0.10
Milan, Italy	Cfa	2017	8.42	6.93	0.74
Tamanrasset, Algeria	BWh	2009	7.85	5.58	0.14
Tateno, Japan	Cfa	2013	9.32	7.09	0.21
Toravere, Estonia	Dfb	2008	8.26	7.34	0.24

building energy by integrating renewable energy with efficient energy storage systems, including battery, thermal storage, hydrogen storage, and liquid organic hydrogen carriers to achieve the optimal utilization of solar energy for meeting electrical and thermal demands. The underlying equations of the capacity expansion optimization problems of the models are described by Omoyele et al. [5]. The model is optimized using the ETHOS.FINE [58] optimization framework with the objective of optimizing the TAC of the system (see Equation (12)). Table 7 in Appendix D presents the techno-economic parameters of the self-sufficient building model. The demand for the self-sufficient building energy system modeling is in 1-min resolution and simulated using SynPro [59], which is presented in Table 8 of Appendix E for Milan 2017 and Berlin 2019.

$$TAC = CAPEX \times \left(\frac{i}{1-(1+i)^{-n}} + OPEX_{rel} \right) \tag{12}$$

Where the TAC is the total annualized cost, CAPEX is the capital expenditure, $OPEX_{rel}$ is the operational expenditure relative to the capital expenditure, i is the interest rate, and n is the components' lifetime.

Validation: For validation purposes, the model is solved with minutely resolved measured data. Then it is also solved with the minutely resolved synthetic data. The percentage of variation in terms of the TAC and the capacities of the installed technologies is quantified. Since the 1-h resolution is the most common resolution for energy system modeling, the model is also solved for a 1-h resolution. The percentage of accuracy loss between the measured minutely resolution to hourly resolution that can be recovered by the synthetic minutely resolution data is also quantified.

Reducing Complexity: Apart from the lack of sub-hourly resolution data for sub-hourly resolved modeling, another deterrent is the computational complexity and associated long runtimes that arise owing to the significantly larger number of constraints and variables in minutely resolved optimization models as compared to hourly resolved ones. To ameliorate this, different methods ranging from non-exact heuristics to exact methods based on brute-force computational power and parallelization can be leveraged to overcome this problem [18]. These methods are the mean of the regular hourly samples of the 1-min resolution and the time series aggregation using a hierarchical clustering algorithm. The mean of the regular samples is obtained by solving the 60 different samples of the 1-min resolution (00:00, 01:00, 02:00 … 23:00 for sample 1; 00:01, 01:01, 02:01 … 23:01 for sample 2 up to 00:59, 01:59, 02:59 … 23:59 for sample 60). In this case, 60 different hourly resolution optimizations are obtained, and the average is taken [5,15].

The time series aggregation utilizes clustering algorithms to reduce the number of data points which can be in typical periods or segments [19,61]. The typical period is the representative periods or days the time series data can be reduced to, while the segment is the number of data points that each typical period contains. Both typical periods and segments rely on feature-based data selection using hierarchical clustering [62,63]. The time series aggregation package used is an open-access Python package, tsam[4] [19,63]. The aim is to achieve accurate and reliable designs while taking significantly less computing time than the 1-min resolution modeling. As recommended by Omoyele et al. [18] for a highly resolved self-sufficient building model, 160 or 365 typical days with 24 segments are employed (other configurations can be used with tsam). The 160 typical days with 24 segments can be denoted by TD160, while the 365 typical days with 24 segments can be denoted by TD365. In terms of the computational runtime, the 60 different hourly optimizations from the regular sampling can be solved with a computing cluster array in parallel. Hence, making the whole process run at the computational speed of the traditional average hourly resolution of the original time series. Tsam technically reduces the data points to 160 or 365 typical periods and 24 segments which is still typically reducing the data points around average hourly resolution (365 x 24 = 8760, which is the annual average hourly resolution time series). This makes the optimization solve at a significantly reduced computational runtime while maintaining accuracy [64].

3. Results

The results are split into two parts, which are the statistical validation of the methodology and the energy system modeling, comparing the measured data, synthetic data, average hourly resolution of the measured data, as well as additional methods to accelerate computational time, concretely, by means of regular samples and clustering-based time series aggregation methods.

3.1. Statistical validation

Table 5 below presents the statistical results of the validation in the predefined locations comprising different climatic conditions. The validation metrics used are the NRMSE and the KSI. The minutely and hourly NRMSE are both determined. The hourly resolution data of these locations is determined by taking the hourly average of their respective minutely resolution data.

For the minutely resolved measurement data of Milan, Berlin 2019, Tamanrasset, Tateno, and Toravere, extended validation results are presented in Figs. 6–10, respectively. Figs. 6–10 delve deeper into the statistical validation by presenting line and CDF plots of the locations for measured and synthetic profiles. Each of the figures is divided into 4 parts based on its daily NRMSE values. Part (a) represents three selected random days. Part (b) represents the days with the lowest NRMSE values. Part (c) represents days with the highest NRMSE values, and Part (d) is the plot of the annual daily NRMSE values. The lower the NRMSE, the better the result of the synthetic irradiance profile. Therefore, Part (b) and Part (c) of the profiles correspond to the best and worst days captured, respectively. A clearer exposition of Figs. 6 and 7 showing the comparison of measured and synthetic GHI in Milan 2017 and Berlin 2019 for three different days are presented in Figs. 16 and 17 of Appendix F. Furthermore, box plots showing the statistical properties of the measured and synthetic data for each location are presented in Fig. 18 in Appendix G.

The validation data spans across different locations, years, and weather classes. This helps to perform some sensitivity analyses on the methodology. For example, to assess the database diversity, the Milan validation site is tested across the entire database to see which database matches the location, as presented in Table 9 of Appendix H. Furthermore, the validation site of Milan is tested for database size across different years of the corresponding weather class database, which is presented in Table 10 of Appendix H.

[4] https://github.com/FZJ-IEK3-VSA/tsam.

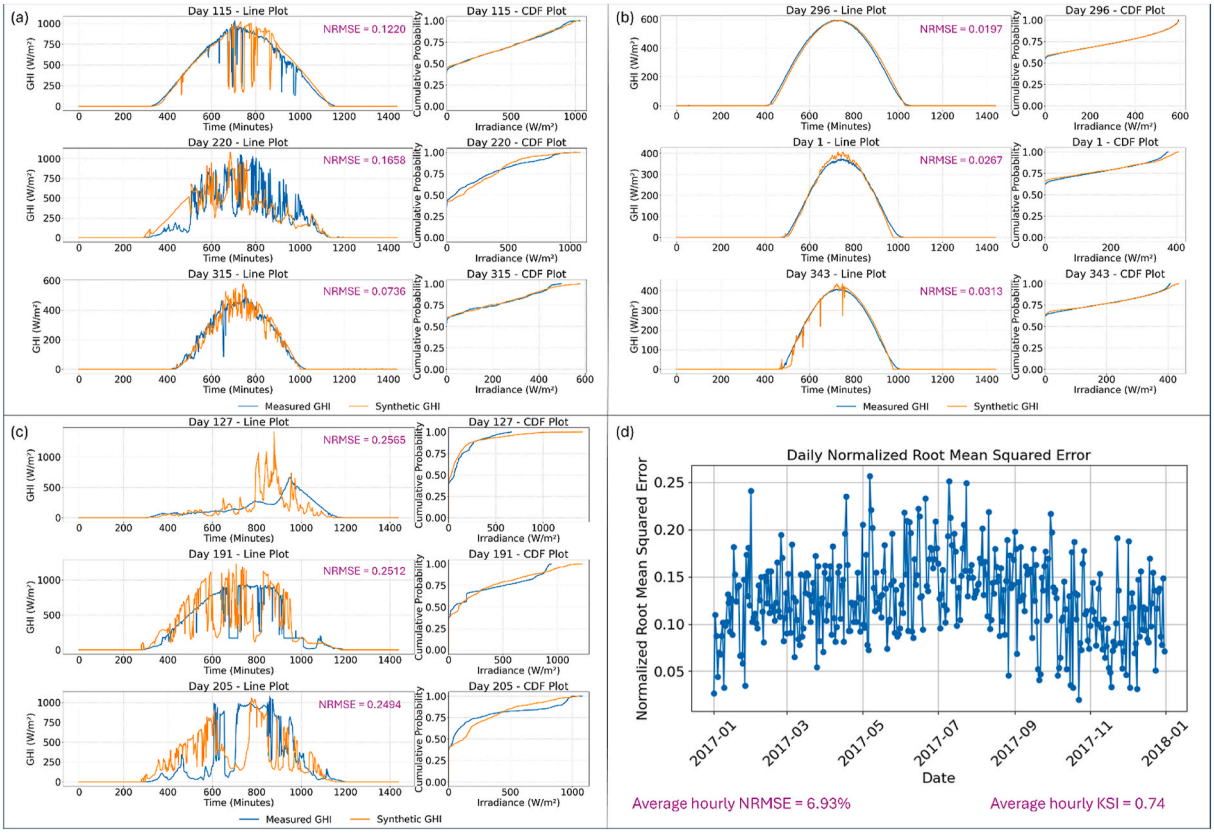


Fig. 6. Comparison of measured and synthetic GHI in Milan 2017 for (a) three random days, (b) days with lowest normalized root mean squared error, (c) days with highest normalized root mean squared error, (d) annual daily normalized root mean squared error. The CDF is the cumulative distribution function.

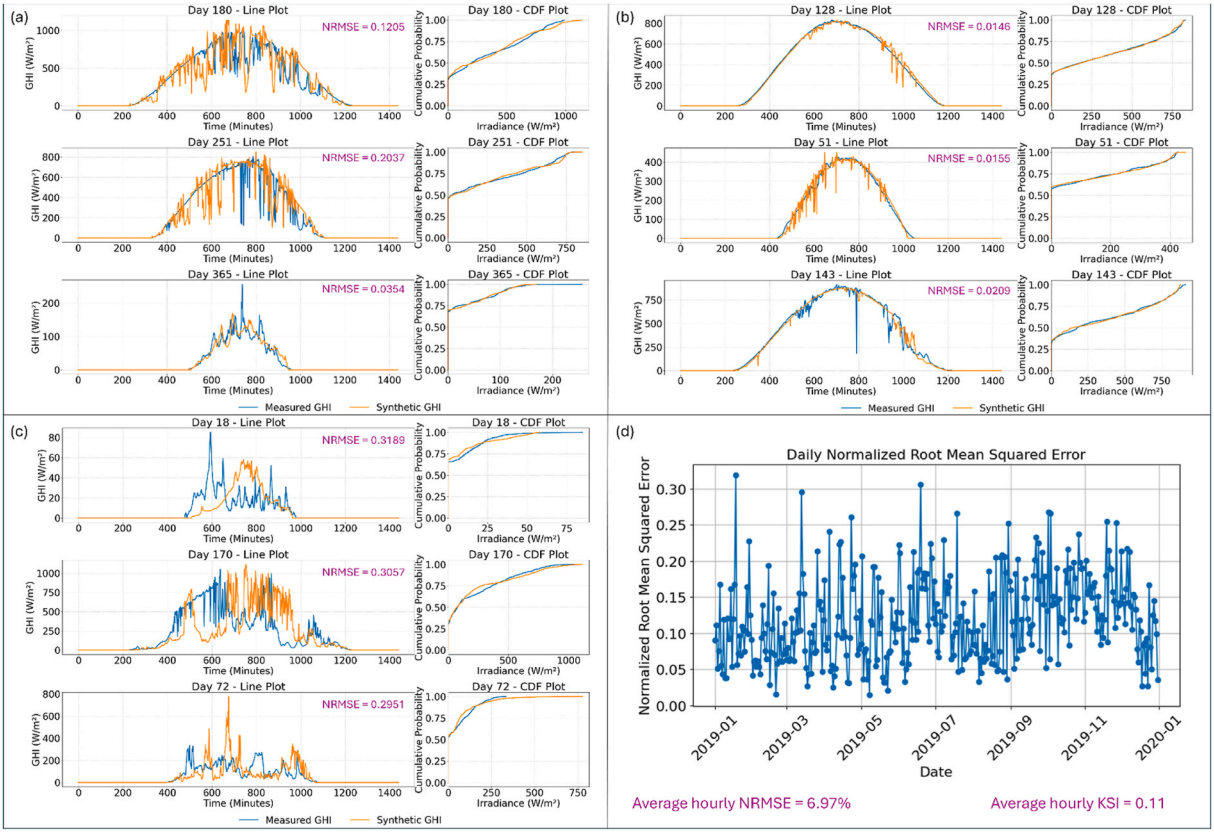


Fig. 7. Comparison of measured and synthetic GHI in Berlin 2019 for (a) three random days, (b) days with lowest normalized root mean squared error, (c) days with highest normalized root mean squared error, (d) annual daily normalized root mean squared error. The CDF is the cumulative distribution function.

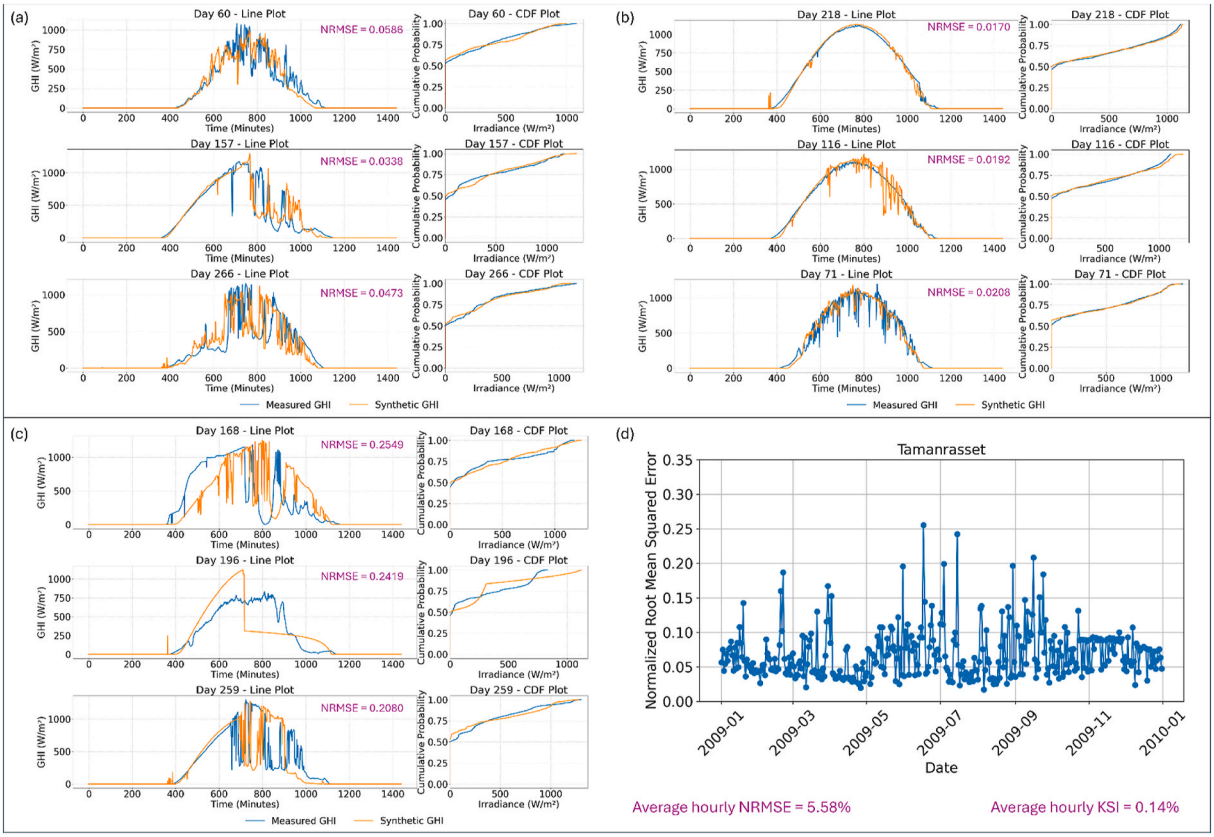


Fig. 8. Comparison of measured and synthetic GHI in Tamanrasset for (a) three random days, (b) days with lowest normalized root mean squared error, (c) days with highest normalized root mean squared error, (d) annual daily normalized root mean squared error. The CDF is the cumulative distribution function.

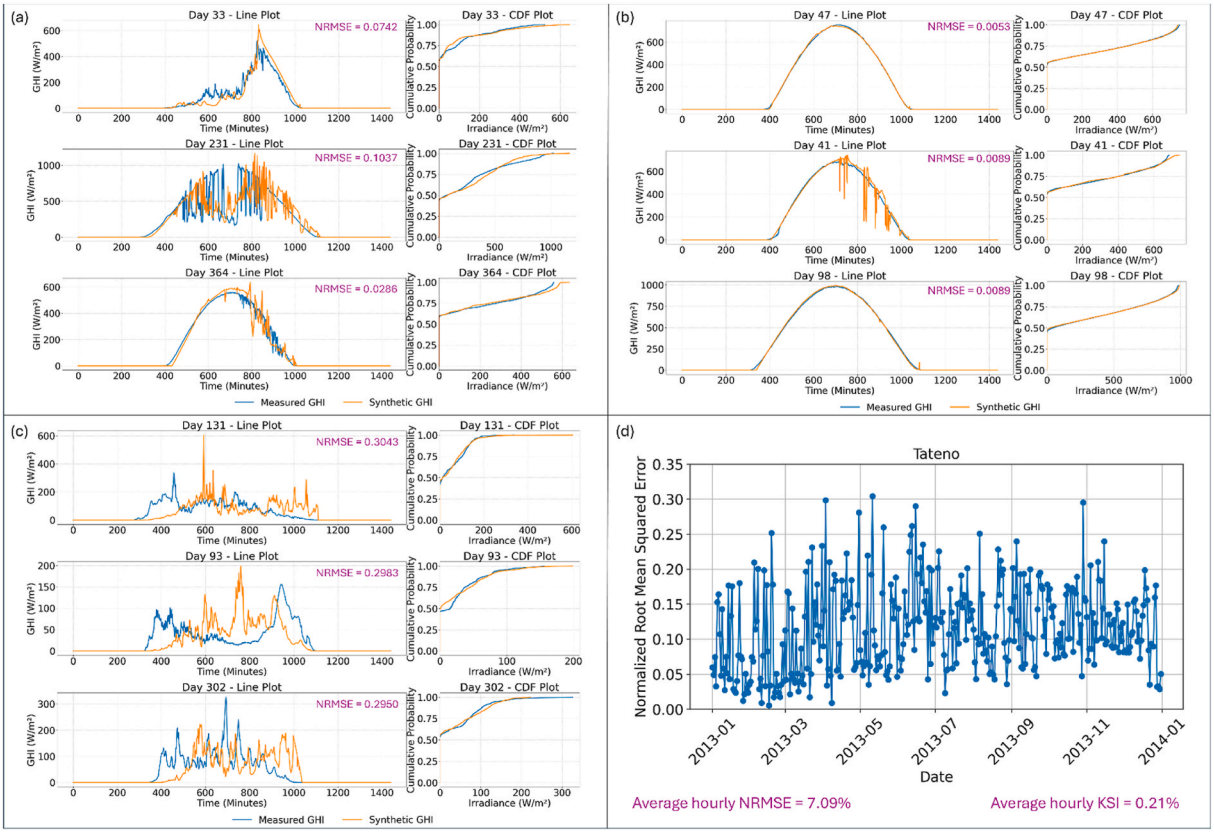


Fig. 9. Comparison of measured and synthetic GHI in Tateno for (a) three random days, (b) days with lowest normalized root mean squared error, (c) days with highest normalized root mean squared error, (d) annual daily normalized root mean squared error. The CDF is the cumulative distribution function.

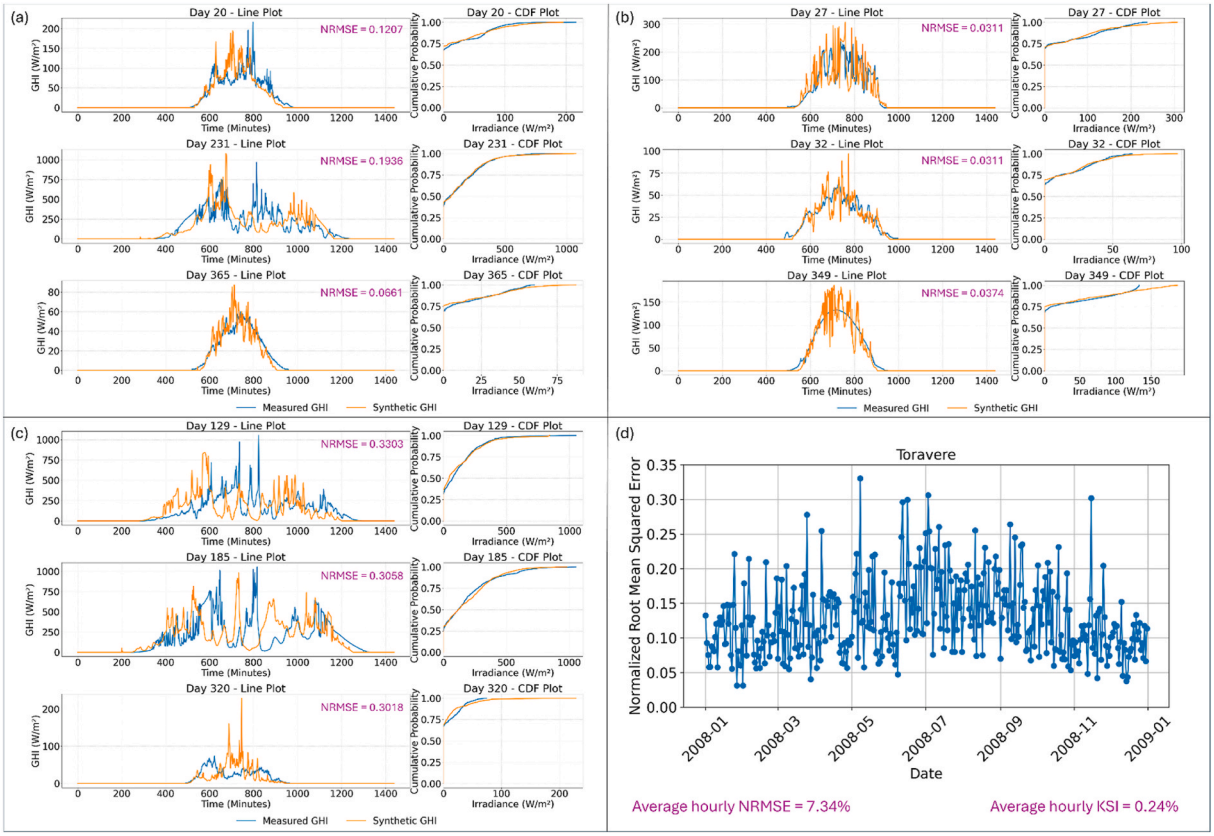


Fig. 10. Comparison of measured and synthetic GHI in Toravere for (a) three random days, (b) days with lowest normalized root mean squared error, (c) days with highest normalized root mean squared error, (d) annual daily normalized root mean squared error. The CDF is the cumulative distribution function.

3.2. Energy system modeling validation

Different data are tested in the energy system optimization model to quantify the accuracy relative to the measured original data of the Milan 2017 and Berlin 2019 locations. The results are presented in Figs. 11–13, with comprehensive data provided in of Tables 11 and 12 of Appendix I. While Fig. 11 presents the bar plot of the TAC, the corresponding computational time of each input data method (in purple), and the associated cost deviation from the original measured result (in green), Figs. 12 and 13 present the capacity deviation of the components of the self-sufficient building model.

4. Discussion

The discussion is divided into two parts: Sections 4.1 and 4.2 present statistics and energy system modeling, respectively.

4.1. Statistics

It can be inferred from the daily line plot that the synthetic data captures the measured data effectively. The selected days in Fig. 6 (a) to Fig. 10 (a) are all cloudy (days without smooth curves) days, and it can be shown that even though the daily NRMSE varies between 0 % and 35 %, the days with large NRMSE still have similar profiles for both measured and synthetic data, and their large NRMSE values arise from the high irradiation values of such days. While the cloudy days are subject to high fluctuations and may therefore have medium to high NRMSE, the clear days (days with smooth curves) have the best NRMSE, as shown in Fig. 6 (b)–10 (b). This means that the clear days are statistically well captured. Fig. 6 (c) to Fig. 10 (c) present the daily highest NRMSE corresponding to the worst-represented days, where the model does not find perfect non-dimensional data matching the days. This problem is primarily due to the limitation of the database arising from the lack of sufficient 1-min resolution data in the database. It may also be as a result of the possible inaccuracies of the measured data of such days, which are not subject to quality control. For example, in Fig. 6 (c), the worst selected days are affected by problems with the measured data. These problems come from interpolations, large jumps in the irradiance data, and stale data. However, some of these days have low irradiance values, and therefore, the deviations have less impact on the energy system modeling and optimization. Fig. 6 (d)–10 (d), which depict the annual daily NRMSE, show that there exists a strong agreement between the synthetic and measured irradiance, with an hourly average NRMSE of approximately 7 %. These data are recorded at a minute resolution and checked for adverse measurement error and unavailable data. For instance, in the case of Milan, two days in 2017 were completely replaced by their previous days and other missing data are imputed with k-nearest neighbors (as discussed in section 2.2.1).

The corresponding CDF plot of the line plots of Figs. 6–10 is plotted to show the distribution of the measured and the synthetic CDFs, indicating that the model accurately captures the overall distribution of the irradiance. While the CDF plots demonstrate strong alignment between the measured and the synthetic GHI in Fig. 6 (a and b) to Fig. 10 (a and b), a discrepancy is observed in the distribution of GHI, particularly in the high irradiance ranges in Fig. 6 (c)–10 (c). The average hourly KSI test is also presented with values less than 1 % indicating a good match in the distribution of the measured and the synthetic values. While both Milan and Tateno are classified under the temperate climate of Cfa, their KSI values differ markedly from 0.74 % for Milan to 0.21 % for Tateno, despite the similar NRMSE around 7.0 %. The elevated KSI at Milan may be due to the specific dynamics that are not prevalent in Tateno. For example, Milan is situated in the Po Valley, a region prone to low-level haze, fog, and aerosol accumulation [65]. These factors introduce complex irradiance patterns which are more difficult to model, as

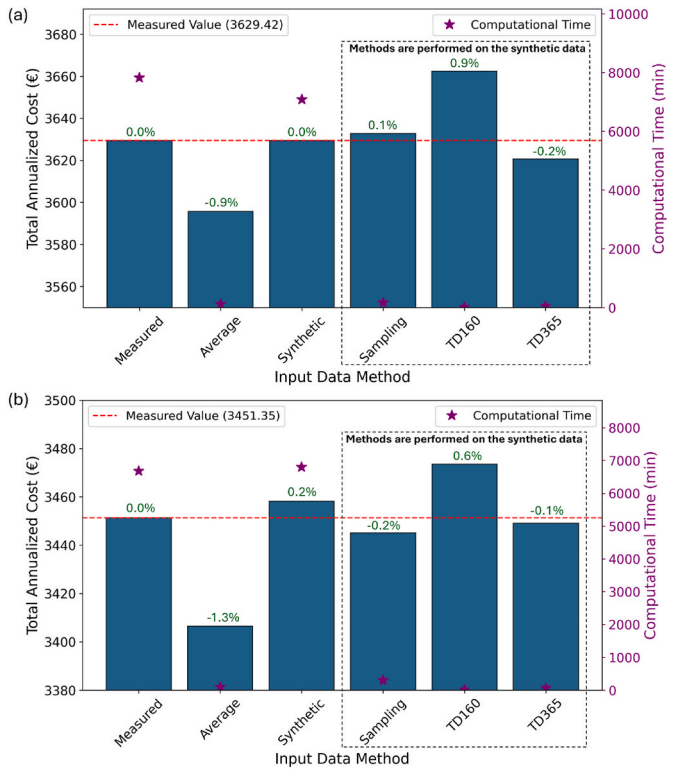


Fig. 11. Results of the total annualized cost for (a) Milan in 2017 (b) Berlin in 2019, using different data in measured, average hourly, synthetic data, average of 60 hourly samples in 1-min, typical days of 160 (TD160), and typical days of 365 (TD365). The corresponding computational time and the deviation from the measured value of each input data method is on the bar plot in purple and green colors, respectively.

compared to other locations with lower KSI values. This phenomenon illustrates the influence of the regional meteorological factors on the performance of the downscaling model. The analysis of interannual uncertainty revealed a NRMSE range between 6.71 % and 7.45 % for the Berlin location, with an average NRMSE of 7.02 %. The KSI is averaged at 0.11 % with minor standard variation, suggesting excellent consistency in capturing irradiance distribution and relatively stable model performance across the years. Furthermore, the box plots across multiple locations (see Figure 5) illustrate the annual distribution of GHI values, with both measured and synthetic data closely matching in median, interquartile range, and variability. The plotted mean and standard deviation markers further confirm the consistency of the synthetic data in capturing the central tendency and spread of measured GHI values. Minor variations in extreme values are visible through the whiskers and outliers, which are expected given natural variability and model approximations. Overall, these box plots reinforce the reliability of the synthetic irradiance data in representing the observed measurements across diverse climatic conditions.

For the various climates considered during the validation process, the location of Tamanrasset (see Fig. 8), which corresponds to the arid hot desert location (BWh), demonstrated remarkable performance. This location is characterized by minimal seasonal variability and predominantly clear skies. The Tamanrasset achieved an exceptional result, with an hourly NRMSE of 5.58 and a KSI of 0.14. Conversely, the temperate climate of Cfa and Cfb, analogous to the climates of Tateno (see Fig. 9), Milan (see Fig. 6), and Berlin (see Fig. 7), exhibits intermediate characteristics. Despite the decline in model accuracy in these climates compared to Tamanrasset, which exhibits less variability, the daily CDF plots demonstrate a high degree of correspondence with the observed data. This suggests that the models possess notable strengths in preserving distributions under varying dynamics. These arid (B climate class)

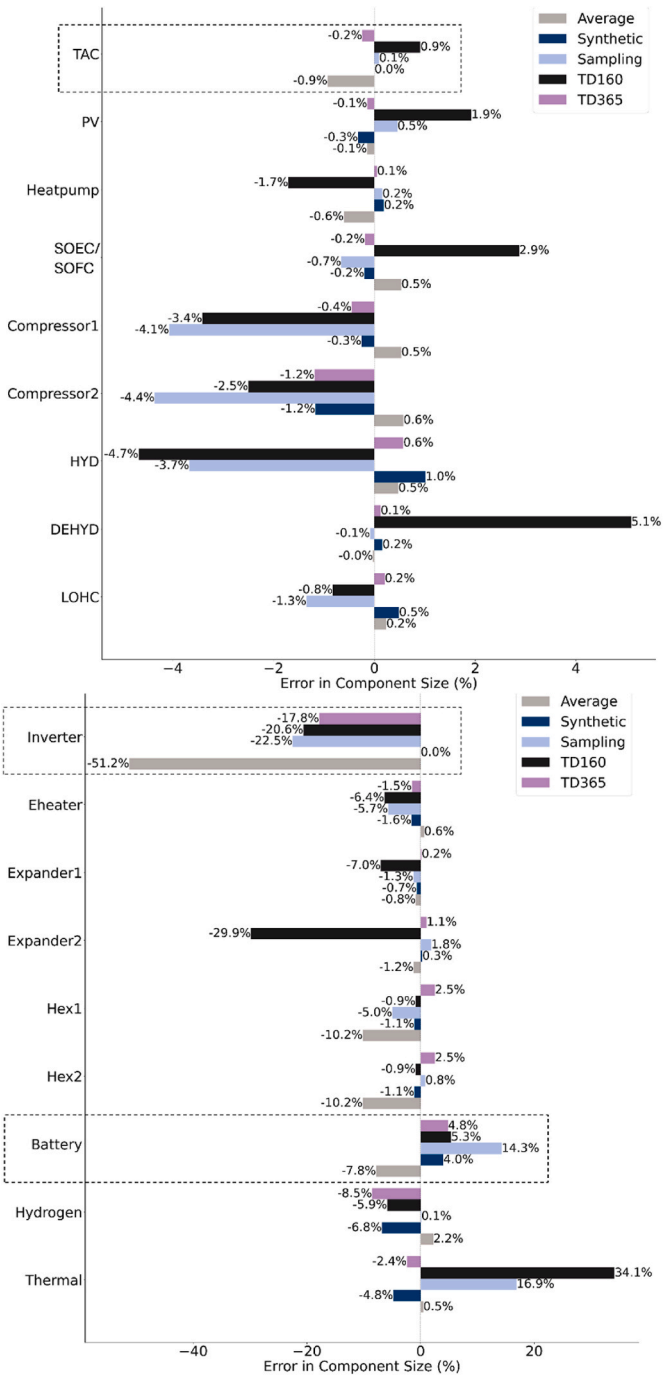


Fig. 12. Result of the components' capacity deviation of different optimization and acceleration approaches from the measured data in percentage for the self-sufficient building for Milan in 2017. TAC – Total Annualized Cost; PV – Photovoltaic; SOEC – Solid Oxide Electrolyzer Cell; SOFC – Solid Oxide Fuel Cell; HYD – Hydrogenation; DEHYD – Dehydrogenation; Hex – Heat Exchanger; LOHC – Liquid Organic Hydrogen Carrier; TD – Typical Days.

and temperate (C climate class) climates corroborate the effectiveness of the model performance in clear climates with less variability over cloudy climates with high variability. The Toravere region, characterized by a cold, no dry season and warm summer, exhibits pronounced seasonal variability. This phenomenon is characterized by low irradiance values in winter and high ones in summer, transitioning to low NRMSE values in winter and high NRMSE in summer.

As indicated in Appendix H, Tables 9 and 10 show sensitivity analyses on the impact of location diversity and database size on the

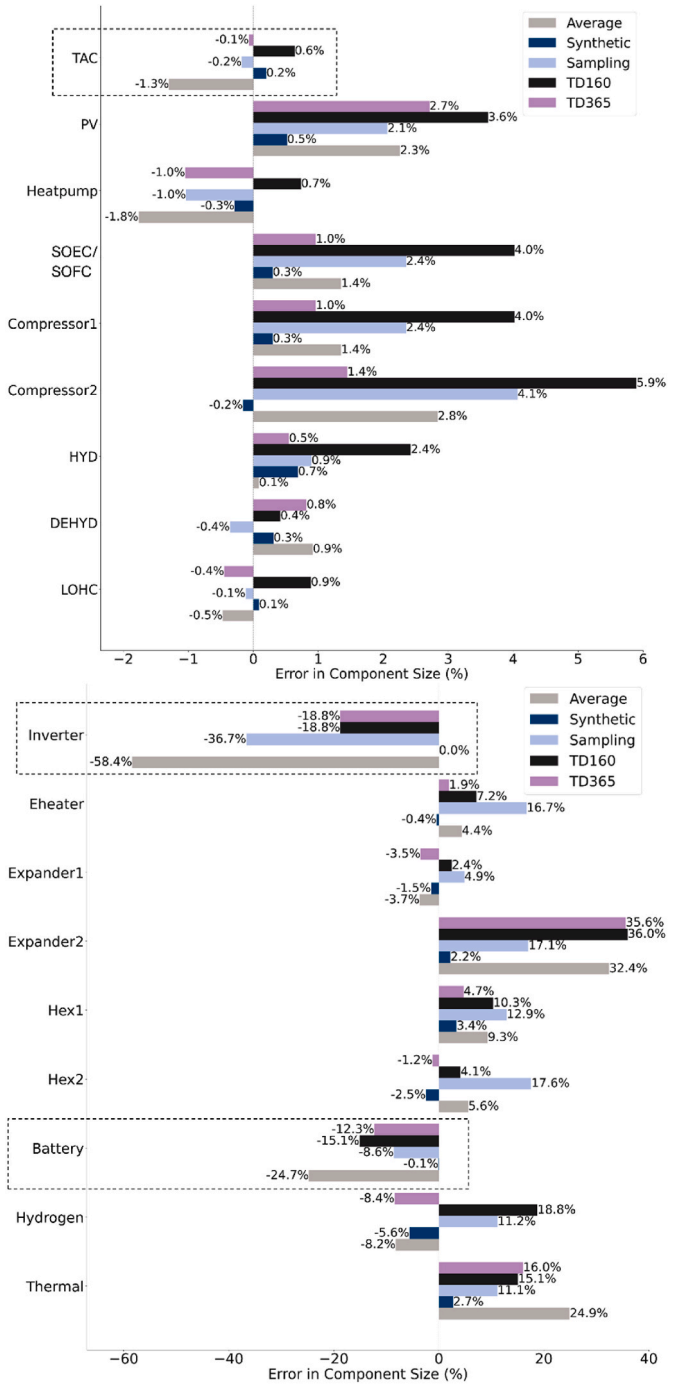


Fig. 13. Result of the components' capacity deviation of different optimization and acceleration approaches from the measured data in percentage for the self-sufficient building for Berlin in 2019. TAC – Total Annualized Cost; PV – Photovoltaic; SOEC – Solid Oxide Electrolyzer Cell; SOFC – Solid Oxide Fuel Cell; HYD – Hydrogenation; DEHYD – Dehydrogenation; Hex – Heat Exchanger; LOHC – Liquid Organic Hydrogen Carrier; TD – Typical Days.

downscaling method. For the location tested (Milan), the best-matching climate with the lowest NRMSE and KSI is its corresponding Köppen-Geiger class of Cfa, which corresponds to a temperate climate with no dry season and a hot summer. However, a Cfb corresponding to a temperate climate, no dry season, and warm summer, yields a nearly identical result. This is attributable to the marked similarity between the two climate classes. Furthermore, the effect of database size on model performance is evident in the decrease in NRMSE with increasing database size (see Table 5). For instance, the NRMSE decreases from 7.85 % with a one-year database to 6.93 % with a database spanning 19.5 years. Conversely, the KSI remains virtually constant at 0.74 % across all database sizes, indicating that a modest training database of one year is adequate for reproducing the distribution of the profiles. The negligible changes in KSI suggest that the model already captures the key distributional characteristics of the irradiance data with limited data. This outcome is indicative of the efficacy of the underlying matching algorithm in reproducing distributional characteristics.

Having a historical data of the site to downscale is useful in improving the methodology significantly. However, the developed methodology has a global application without sites' historical measurements. An improvement of the climate classes in cold and polar Köppen-Geiger climates with currently little or no data in the database may help improve the results. In addition, other machine learning algorithms may be incorporated into the methodology to capture days better, which have profiles that do not have perfect equivalents in the created database.

4.2. Energy system modeling

From the energy system modeling result, it can be inferred that comparing the synthetic result and the average hourly result, the synthetic data provides a better result in terms of the TAC as compared to the widely used hourly average result, as it deviates less from the results obtained for the real measured data (see Fig. 8). The average hourly result shows an underestimation of the TAC by 0.9 % and 1.3 % for Milan 2017 and Berlin 2019, respectively. This underestimation is attributable to the hourly resolution's incapacity to adequately capture the inherent intra-hour fluctuations present in the sub-hourly data. Consequently, the process of averaging leads to a minimization of the maxima and a maximization of the minima of the input data. The synthetic data shows an overestimation of the TAC by only 0.0 % and 0.2 % for Milan 2017 and Berlin 2019, respectively. The very low overestimation by the synthetic data is attainable through the introduction of the daily sum match between the measured and the synthetic data (as discussed in step 5 of section 2.2.1), which significantly improved the results. The computational acceleration methods [18] using the mean of the regular samples, as well as clustering-based time series aggregation of 160 and 365 typical days with 24 segments (TD160 and TD365), are also applied to the optimization (Fig. 11).

While TD160 successfully reduces the computational runtime, it fails to accurately capture the result of the 1-min resolution data due to over-aggregation of the model. Consequently, an overestimation of 0.9 % and 0.6 % for Milan and Berlin, respectively, is observed. Therefore, a more precise result would be one that captures the entirety of the dataset without significant aggregation. In this case, the regular samples average (sampling) and TD365 are the most appropriate metrics. The most effective computational acceleration method results in terms of TAC are obtained by these two metrics. It is important to note that TD160 considers 160 typical days and 24 segments. This results in a reduction of the input data from 525,600 datapoints (the full 1-min resolution) to 3840 (160 x 24). In contrast, the TD365 reduces the input datapoints from 525,600 to 8760 (365 x 24 or full 1-h resolution). The TD160 is less accurate than the TD365, yet it is more efficient in terms of computational runtime savings.

The optimal accuracy-complexity compromise is obtained for TD365 as recommended by Omoyele et al. [18], and thus the best approach for optimizing the minutely resolved energy system model with an aggregation-based method that reduces computational time to a level comparable to optimizations at hourly resolution. The TD365 shows an underestimation of the TAC by 0.2 % and 0.1 % for Milan 2017 and Berlin 2019, respectively. The highly dynamic components (inverters and storage, e.g., the battery) are subject to large deviations when optimized using average hourly resolution (see Figs. 12 and 13), which leads to the infeasibility of the model when such capacities are tested for feasibility using the measured data. However, this is ameliorated using

the synthetic data, as the initial inverter underestimation for Milan 2017 and Berlin 2019 of 51.2 % and 58.4 % respectively are now 0 %, and the battery underestimation for Milan 2017 and Berlin 2019 of 7.8 % and 24.7 %, respectively are now replaced by 4 % overestimation and 0.1 % underestimation yielding a more feasible energy system model design, as illustrated in Figs. 12 and 13. For the two cases of Milan 2017 and Berlin 2019, the mean of regular samples and TD365 give an improved result as compared to the mean hourly resolution for the capacities of highly dynamically operated components such as the inverter and the battery.

While the energy system analysis of this study has been on a self-sufficient building system, the proposed methodology demonstrates strong potential for scalability to grid-connected solar energy system applications. The ability of the methodology to generate high-quality, high-resolution synthetic irradiance data across diverse climatic systems underscores its generalizability and robustness. For further application to real-time grid-connected systems, the methodology can support operation forecasting and decision-making by providing high-quality sub-hourly irradiance data in locations where high-quality data are limited. Due to its computational tractability, it can be adapted to near-real-time implementation using a regularly updated database. This renders it suitable for incorporation into forecasting pipelines. Future integration with real-time satellite data or numerical weather prediction could further enhance its forecasting ability.

5. Conclusion and outlook

This work provides a globally applicable, non-dimensional methodology for increasing the temporal resolution of hourly global horizontal irradiance to minutely resolution. The methodology builds on the developed work of Peruchena et al. [22] and Larrañeta et al. [31] by matching the daily irradiance characteristics of the data to be downscaled with a robust database of non-dimensional minutely and daily parameters, across diverse Köppen-Geiger weather, ensuring its global applicability. The proposed methodology, by means of synthesized data, facilitates the consideration of transient phenomena, such as intra-hour fluctuations and passing clouds, in energy system modeling, consequently, enhancing its accuracy.

The methodology is validated using statistical metrics and a self-sufficient building energy system model. Across diverse locations, the synthetic data achieves mean hourly normalized root mean square error values between 5.6 % and 7.3 %, and mean hourly Kolmogorov-Smirnov integral test values between 0.1 % and 0.7 %. The outcomes of these analyses are contingent upon the atmospheric clarity, with clear days generally yielding more precise results compared to cloudy days. The statistical metrics reveal good representation and distribution of the synthetic data relative to the measured data. Furthermore, the synthetic data demonstrates improved system performance in terms of cost, feasibility, and component sizing when compared to the common hourly average, as evidenced by the results of the self-sufficient building energy system model. While the hourly average data underestimates the total annualized cost of the system by up to 1.3 %, the synthetic data overestimates it by up to only 0.2 %. It also improved on the significant undersizing of highly dynamic components, reducing inverter capacity underestimation from 58 % to 0 % and battery capacity underestimation from 25 % to 4 %. The methods of reducing complexities employed on the synthetic data displayed a high computational management compared to the fully resolved 1-min data, and a more accurate result compared to the hourly average data.

The main contributions of this study are summarized as follows:

- A globally applicable and high accuracy temporal downscaling method for global horizontal irradiance is proposed.
- A comprehensive database of non-dimensional curves and daily irradiance parameters, constructed across diverse Köppen-Geiger climate zones, is created.
- The high-resolution synthetic irradiance data leads to improved energy system modeling performance, including reliable cost estimation and more accurate components' sizing.
- Accelerated computation methods as alternatives to full 1-min resolution models are applied to provide an improved model accuracy at an equivalent computational time of 1 h resolution models.

The limitation of the developed method is its inability to downscale historical data prior to 2004, due to the applicability of the McClear clear sky model being limited to years from 2004 onwards. Other models to obtain the extraterrestrial irradiance could be employed in this case. Moving forward, further expansion of the database with real-time measurement for sites, particularly for underrepresented regions of cold and polar Köppen-Geiger climates, will enhance its applicability and robustness. In addition, integration of the proposed methodology with machine learning approaches holds promise for enhancing its accuracy, particularly during instances of cloud cover when precise matching days cannot be obtained from the created database. Future work could also explore the integration of satellite-based irradiance products with resolutions below 1 h as input data and machine learning-based downscaling techniques. Consequently, the augmentation of location validation and energy system optimization holds promise in the development and implementation of a more robust validation of the methodology. Despite these limitations, the developed methodology will support planners and decision makers, regardless of location, to design solar-energy-based energy systems more accurately than previous approaches.

CRediT authorship contribution statement

Olalekan Omoyele: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Maximilian Hoffmann:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Jann Michael Weinand:** Writing – review & editing, Supervision, Resources, Methodology, Formal analysis, Conceptualization. **Miguel Larrañeta:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jochen Linßen:** Writing – review & editing, Supervision. **Detlef Stolten:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Helmholtz Association under the program "Energy System Design". Miguel Larrañeta received support by Grant RYC2021-032300-I, funded by the Ministry of Science and Innovation/State Research Agency/10.13039/501100011033 and by the European Union NextGenerationEU/Recovery, Transformation and Resilience Plan.

Appendix A

Table 6
Summary of the Five-Step Downscaling Procedure

Step	Step Name	Description	Input(s)	Output(s)
1	**Collection and Preparation of Input Parameters**	Site metadata and simulation duration are used to compute solar angles and extraterrestrial irradiance using McClear model.	Latitude, longitude, simulation start and end time.	Extraterrestrial irradiance, solar angles, climate class.
2	**Extraction of Defining Parameters from Low-Resolution Data**	Compute the daily defining parameters (k_d, VI, NVI, F_m, and ICCDF) from hourly irradiance data.	Hourly-resolution data, extraterrestrial irradiance, solar angles, climate class.	Daily defining parameters (k_d, VI, NVI, F_m, and ICCDF) and weather class
3	**Matching Algorithm**	Match daily defining parameters with the closest day in the daily database using k-nearest neighbor algorithm with 1 neighbor and Euclidean distance.	Daily defining parameters, weather class	Closest-matching day
4	**Selection of High-Resolution Data**	Retrieve non-dimensional irradiance profile corresponding to the matched day from the daily database.	Matched days, highly resolved database	Selection of high resolution non-dimensional data from the highly resolved database
5	**Unpacking the 1 Minute Data**	Scale the non-dimensional data back to the standard irradiance value in W/m^2 and optimize daily scaling factor, k, to minimize the daily cummulative energy difference between the synthetic and the measured irradiance.	Selected high resolution profile, daily sunrise and sunset, extraterrestrial irradiance, measured irradiance, daily scaling factor, k.	Synthetic 1-min irradiance in W/m^2

Appendix B

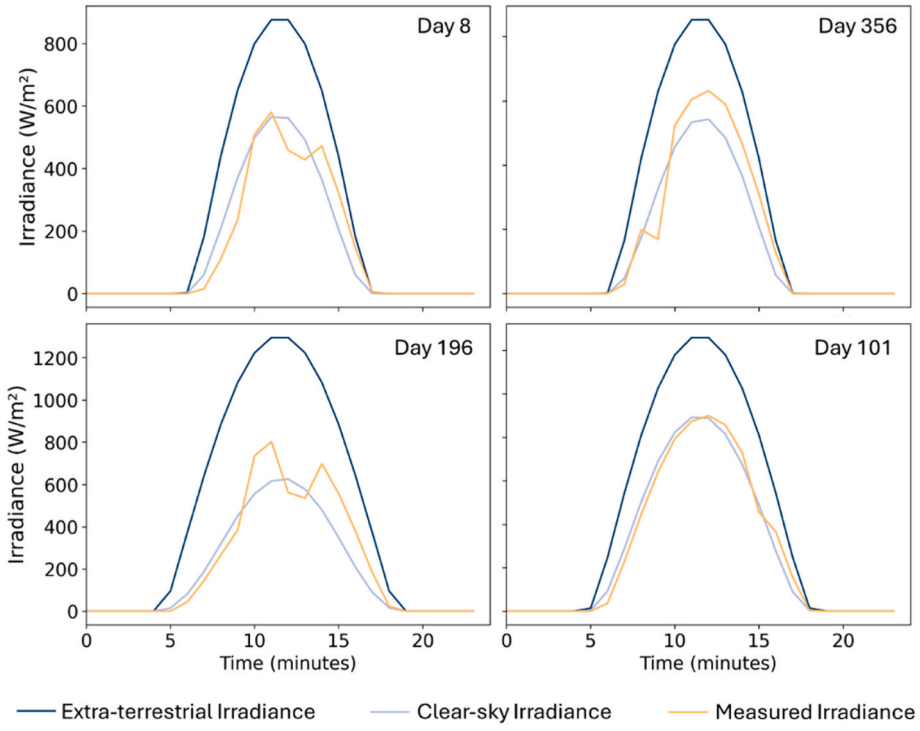


Fig. 14. Four random days in Delhi, India, to see the effect of extraterrestrial irradiance and clear sky irradiance in a polluted environment for variability capture.

Appendix C

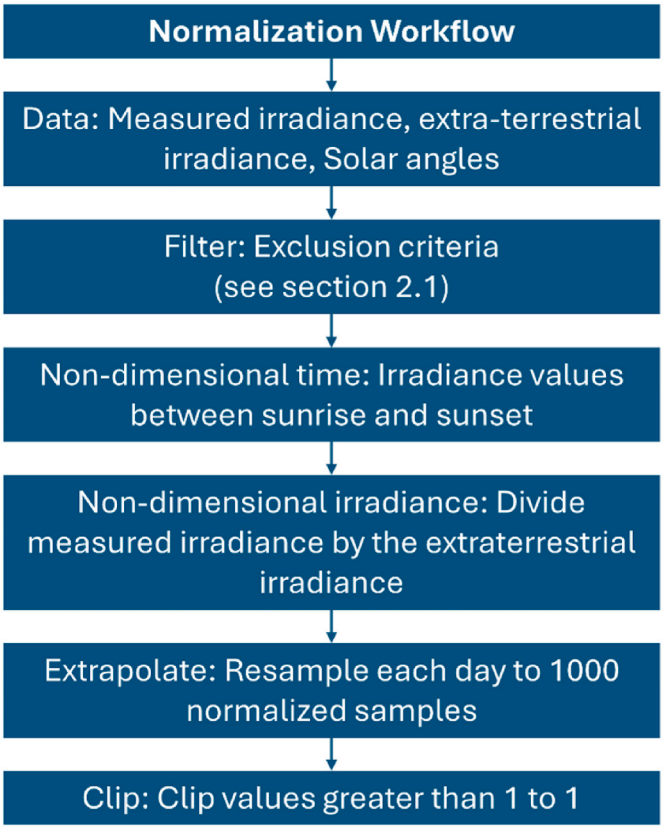


Fig. 15. Normalization workflow of the non-dimensional curves.

Appendix D

Table 7
Cost parameters of the self-sufficient building model by Knosala et al. [60].

	Capex				Opex		Lifetime	
Components	Fixed		Capacity-Specific		Fixed + Capacity-Specific			
Photovoltaic Ground	–	–	4000.00	€/kW$_p$	1.00	% Inv./a	20	a
Photovoltaic Rooftop	–	–	769.00	€/kW$_p$	1.00	% Inv./a	20	a
Inverter	–	–	75.00	€/kW$_p$	–	–	20	a
Battery	–	–	301.00	€/kWh$_p$	–	–	15	a
Reversible Solid Oxide Cell	5000.00	€	2400.00	€/kW$_{el}$	1.00	% Inv./a	15	a
Heat Pump	4230.00	€	504.90	€/kW$_{th}$	1.50	% Inv./a	20	a
Thermal Storage	–	–	90.00	€/kWh$_{th}$	0.01	% Inv./a	25	a
E-Heater & E-Boiler	–	–	60.00	€/kW$_{th}$	2.00	% Inv./a	30	a
Tank	–	–	0.79	€/kWh$_{H2}$	–	–	25	a
Dibenzyltoluene	–	–	1.25	€/kWh$_{H2}$	–	–	25	a
Hydrogen Vessels	–	–	15.00	€/kWh$_{H2}$	–	–	25	a
Hydrogenizer	2123.30	€	761.10	€/kW$_{H2}$	1.00	% Inv./a	20	a
Dehydrogenizer	1140.00	€	408.60	€/kW$_{H2}$	1.00	% Inv./a	20	a
Low Pressure Compressor	–	–	1716.71	€/kW$_p$	1.00	% Inv./a	25	a
High Pressure Compressor	560.00	€	1329.80	€/kW$_p$	1.00	% Inv./a	25	a
Heat-Exchangers 1 and 2	–	–	1.00	€/kW$_{th}$	1.00	% Inv./a	–	a
Expanders 1 and 2	–	–	1.00	€/kW$_{th}$	1.00	% Inv./a	25	a

Appendix E

Table 8
Building demand parameters.

		Building1 (Milan 2017)	Building2 (Berlin 2019)
Number of occupants		4	4
H_ceiling/ww_ratio		2.4/0.2	2.5/0.25
Size (m^2)		200	200
Height (Stories)		1	2
Annual demand (kWh)	**Electricity**	3087.85	3050.98
	Heat	16257.05	16611.26
Mean (kWmin)	**Electricity**	0.3525	0.3483
	Heat	1.8558	1.8963
Standard deviation (kWmin)	**Electricity**	0.5109	0.4930
	Heat	1.9761	2.0088

Appendix F

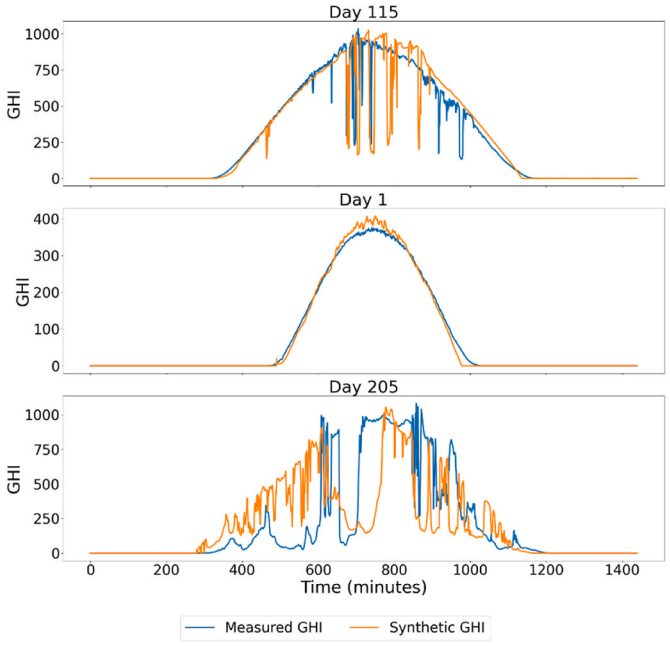


Fig. 16. Comparison of measured and synthetic GHI in Milan 2017 for three different days.

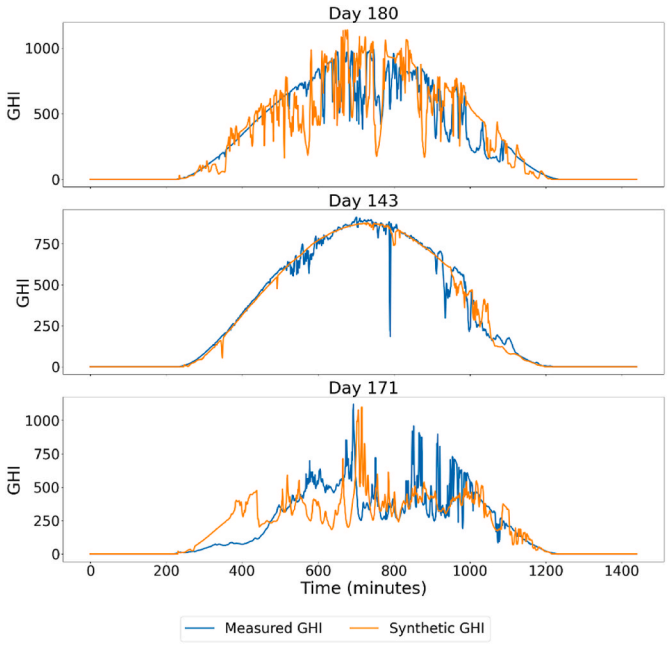


Fig. 17. Comparison of measured and synthetic GHI in Berlin 2019 for three different days.

Appendix G

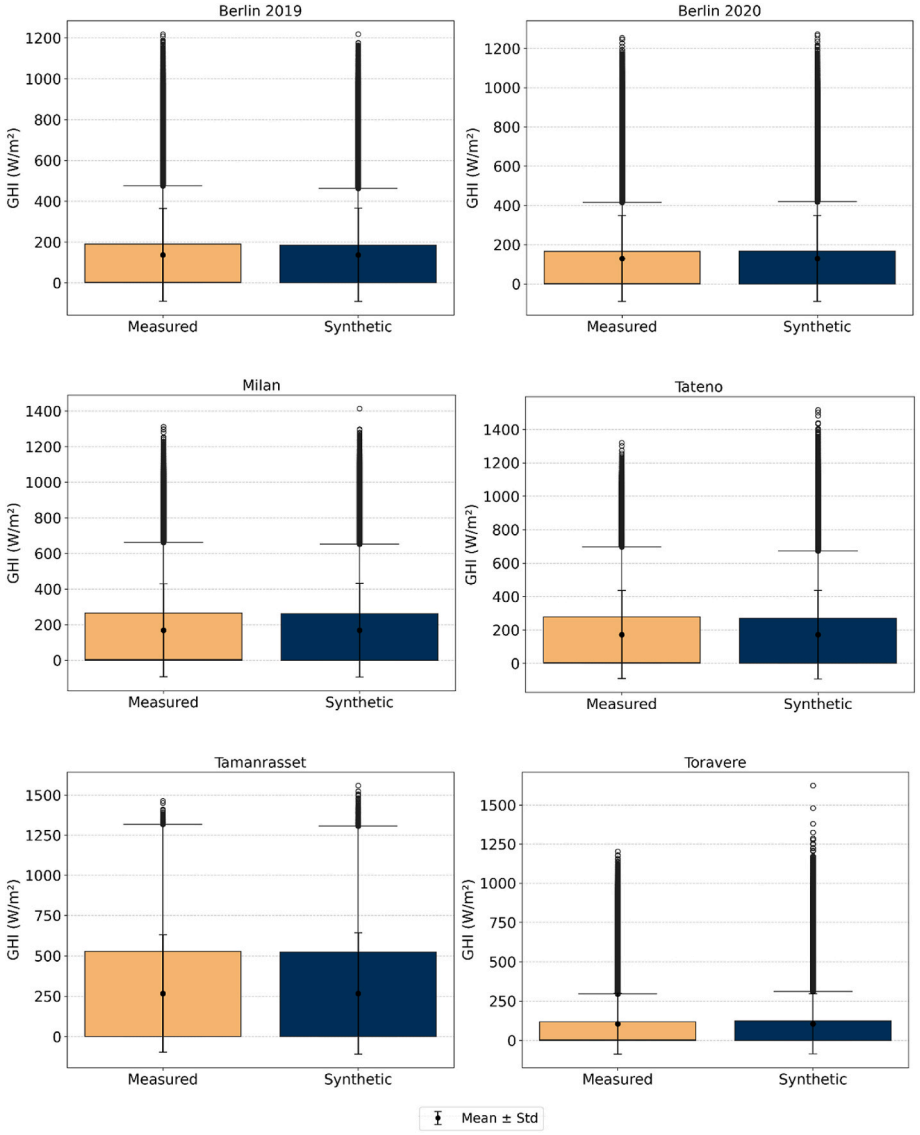


Fig. 18. Box plot showing the statistical properties between the measured and synthetic data for different validation locations.

Appendix H

Table 9
Influence of location diversity on the accuracy of the downscaling method.

Correlated Köppen-Geiger Climate[2]	NRMSE Hour (W/m^2)	KSI Hour (%)
Aw	7.76	0.84
Bsh	7.49	0.76
Bwh	7.50	0.84
Cfa	6.93	0.74
Cfb	6.95	0.77
Csa	7.85	0.79
Csb	7.94	0.76
Cfa/Cfb/Ocean	7.14	0.75

Table 10
Influence of database size on the accuracy of the downscaling method.

Database size (Years)	NRMSE Hour (W/m^2)	KSI Hour (%)
1	7.85	0.74
4	7.83	0.74
8	7.88	0.74
12	7.11	0.74
16	6.94	0.75
Full (19.52)	6.93	0.74

Appendix I

Table 11
Table of results for Milan 2017 using different data in measured, average hourly, synthetic data, average of 60 hourly samples in 1-min, typical days of 160, and typical days of 365.

Components	Unit	Measured	Average	Synthetic	Sampling	TD160	TD365
TAC	€	3629.417	3595.703	3629.454	3632.78	3662.524	3620.727
PV	kW_p	11.59839	11.58171	11.56093	11.65135	11.8213	11.58258
Inverter	kW_p	6.852589	3.342544	6.852589	5.309767	5.441717	5.629491
Heatpump	kW_{th}	6.196846	6.159355	6.208509	6.2066	6.091244	6.200098
Eheater	kW_{th}	0.417623	0.420109	0.410928	0.393729	0.391022	0.411217
Eboiler	kW_{th}	0.354864	0.291255	0.082257	0.038161	1.69E-07	0.006359
rSOEC	kW_{el}	3.889021	3.909791	3.881362	3.863531	4.000791	3.881812
rSOFC	kW_{el}	3.889021	3.909791	3.881362	3.863531	4.000791	3.881812
Compressor1	kW_p	0.222103	0.223289	0.221537	0.213067	0.214523	0.221108
Compressor2	kW_p	0.09834	0.098906	0.097189	0.094056	0.095881	0.097174
Expander1	kW_{th}	2.21419	2.195638	2.199464	2.186055	2.059239	2.218102
Expander2	kW_{th}	0.884721	0.873775	0.887396	0.901041	0.620621	0.89411
HYD	kW_{H2}	1.630156	1.637925	1.646672	1.570328	1.553913	1.639502
DEHYD	kW_{H2}	3.569166	3.56825	3.574803	3.566195	3.75101	3.573592
Hex1	kW_{th}	0.782836	0.703242	0.774399	0.743904	0.77608	0.802651
Hex2	kW_{th}	0.778922	0.699725	0.770527	0.785389	0.7722	0.798637
Battery	kWh_p	4.693275	4.326669	4.881476	5.362296	4.944058	4.918613
Hydrogen	kWh_{H2}	27.39216	28.00643	25.52363	27.42508	25.78829	25.05461
Thermal	kWh_{th}	14.57971	14.64694	13.87878	17.04399	19.55655	14.22859
LOHC	kWh_{H2}	5692.606	5706.05	5720.426	5615.955	5645.661	5704.336

Table 12
Table of results for Berlin 2019 using different data in measured, average hourly, synthetic data, average of 60 hourly samples in 1-min, typical days of 160, and typical days of 365.

Components	Unit	Measured	Average	Synthetic	Sampling	TD160	TD365
TAC	€	3451.345	3406.54	3458.167	3445.172	3473.29	3449.053
PV	kW_p	12.25917	12.53615	12.32266	12.51231	12.70191	12.59147
Inverter	kW_p	7.767394	3.234016	7.767394	4.920022	6.310086	6.310086
Heatpump	kW_{th}	5.141412	5.050618	5.126538	5.0881	5.17923	5.087491
Eheater	kW_{th}	0.207596	0.216709	0.20672	0.242299	0.222506	0.211598
Eboiler	kW_{th}	0.705865	1.114323	0.880538	0.925693	0.049019	1.115286
rSOEC	kW_{el}	2.796084	2.833836	2.804396	2.861852	2.908456	2.822959
rSOFC	kW_{el}	2.796084	2.833836	2.804396	2.861852	2.908456	2.822959
Compressor1	kW_p	0.159685	0.161841	0.16016	0.163441	0.166103	0.16122
Compressor2	kW_p	0.05589	0.057476	0.055799	0.058164	0.059185	0.056699
Expander1	kW_{th}	1.512302	1.457043	1.489925	1.585914	1.549227	1.459782
Expander2	kW_{th}	0.615835	0.815505	0.6296	0.720863	0.837233	0.834884
HYD	kW_{H2}	1.512302	1.513616	1.522705	1.525827	1.548969	1.520595
DEHYD	kW_{H2}	2.560854	2.584322	2.568795	2.551617	2.571468	2.581835
Hex1	kW_{th}	0.410794	0.449026	0.424612	0.463865	0.453261	0.430166
Hex2	kW_{th}	0.433205	0.457439	0.422492	0.509458	0.451067	0.42802
Battery	kWh_p	6.509155	4.898356	6.505433	5.951915	5.526527	5.709567
Hydrogen	kWh_{H2}	19.04732	17.48451	17.98086	21.17836	22.62168	17.444
Thermal	kWh_{th}	11.79202	14.73149	12.11029	13.10544	13.56881	13.68301
LOHC	kWh_{H2}	5311.545	5286.443	5316.186	5305.21	5358.64	5287.662

Table abbreviations.
TAC – Total Annualized Cost.
PV – Photovoltaic.
SOEC – Solid Oxide Electrolyzer Cell.
SOFC – Solid Oxide Fuel Cell.
HYD – Hydrogenation.

DEHYD – Dehydrogenation.
Hex – Heat Exchanger.
LOHC – Liquid Organic Hydrogen Carrier.
TD – Typical Day.

Data availability

All data supporting the findings of this study are openly available at: https://github.com/FZJ-IEK3-VSA/ETHOS.TISED. The repository contains the non-dimensional databases, code, and documentation necessary to replicate the results and conduct further analysis.

References

[1] G. Salazar, C. Gueymard, J.B. Galdino, O. de Castro Vilela, N. Fraidenraich, Solar irradiance time series derived from high-quality measurements, satellite-based models, and reanalyses at a near-equatorial site in Brazil, Renew. Sustain. Energy Rev. 117 (2020) 109478, https://doi.org/10.1016/j.rser.2019.109478.

[2] T. Levin, P.L. Blaisdell-Pijuan, J. Kwon, W.N. Mann, High temporal resolution generation expansion planning for the clean energy transition, Renew. Sustain. Energy Transit. 5 (2024) 100072, https://doi.org/10.1016/j.rset.2023.100072.

[3] G.M. Njoka, L. Mogaka, A. Wangai, Enhancing grid stability and resilience through BESS optimal placement and sizing in VRES-dominated systems, Energy Rep. 13 (2025) 1764–1779, https://doi.org/10.1016/j.egyr.2025.01.028.

[4] J. Salom, J. Widén, J. Candanedo, K.B. Lindberg, Analysis of grid interaction indicators in net zero-energy buildings with sub-hourly collected data, Adv. Build. Energy Res. 9 (2015) 89–106, https://doi.org/10.1080/17512549.2014.941006.

[5] O. Omoyele, et al., Impact of temporal resolution on the design and reliability of residential energy systems, Energy Build. 319 (2024) 114411, https://doi.org/10.1016/j.enbuild.2024.114411.

[6] J.E. Bistline, The importance of temporal resolution in modeling deep decarbonization of the electric power sector, Environ. Res. Lett. 16 (8) (2021) 084005, https://doi.org/10.1088/1748-9326/ac10df.

[7] J. Deane, G. Drayton, B.Ó. Gallachóir, The impact of sub-hourly modelling in power systems with significant levels of renewable generation, Appl. Energy 113 (2014) 152–158, https://doi.org/10.1016/j.apenergy.2013.07.027.

[8] H. Gangammanavar, S. Sen, V.M. Zavala, Stochastic optimization of sub-hourly economic dispatch with wind energy, IEEE Trans. Power Syst. 31 (2) (2015) 949–959, https://doi.org/10.1109/TPWRS.2015.2410301.

[9] N. Troy, D. Flynn, M. O'Malley, The importance of sub-hourly modeling with a high penetration of wind generation, in: 2012 IEEE Power and Energy Society General Meeting, IEEE, 2012, pp. 1–6, https://doi.org/10.1109/PESGM.2012.6345631.

[10] A.V. Klokov, E.Y. Loktionov, Temporal resolution of input weather data strongly affects an off-grid PV system layout and reliability, Solar 3 (1) (2023) 49–61, https://doi.org/10.3390/solar3010004. MDPI.

[11] C. O'Dwyer, D. Flynn, Using energy storage to manage high net load variability at sub-hourly time-scales, IEEE Trans. Power Syst. 30 (4) (2014) 2139–2148, https://doi.org/10.1109/TPWRS.2014.2356232.

[12] I.D. Lopez, D. Flynn, M. Desmartin, M. Saguan, T. Hinchliffe, Drivers for sub-hourly scheduling in unit commitment models, in: 2018 IEEE Power & Energy Society General Meeting (PESGM), IEEE, 2018, pp. 1–5, https://doi.org/10.1109/PESGM.2018.8586262.

[13] M. Ernst, J. Gooday, Methodology for generating high time resolution typical meteorological year data for accurate photovoltaic energy yield modelling, Sol. Energy 189 (2019) 299–306, https://doi.org/10.1016/j.solener.2019.07.069.

[14] M. Hofmann, G. Seckmeyer, Influence of various irradiance models and their combination on simulation results of photovoltaic systems, Energies 10 (10) (2017) 1495, https://doi.org/10.3390/en10101495.

[15] A. Villoz, B. Wittmer, A. Mermoud, M. Oliosi, A. Bridel-Bertomeu, S. Pvsyst, A model correcting the effect of sub-hourly irradiance fluctuations on overload clipping losses in hourly simulations, in: 8th World Conference on Photovoltaic Energy Conversion, December, 2024 2022 [Online]. Available: https://www.pvsyst.com/wp-content/publications/2022_PVsyst_WCPEC8_SubHourlyClipping_Article.pdf.

[16] M.J. Mayer, Effects of the meteorological data resolution and aggregation on the optimal design of photovoltaic power plants, Energy Convers. Manag. 241 (2021) 114313, https://doi.org/10.1016/j.enconman.2021.114313.

[17] A. Zurita, C. Mata-Torres, J.M. Cardemil, R.A. Escobar, Assessment of time resolution impact on the modeling of a hybrid CSP-PV plant: a case of study in Chile, Sol. Energy 202 (2020) 553–570, https://doi.org/10.1016/j.solener.2020.03.100.

[18] O. Omoyele, M. Hoffmann, J. M. Weinand, and D. Stolten, "Accelerating computational efficiency in sub-hourly renewable energy systems modeling," Available at SSRN 5004752, doi: https://dx.doi.org/10.2139/ssrn.5004752.

[19] M. Hoffmann, L. Kotzur, D. Stolten, M. Robinius, A review on time series aggregation methods for energy system models, Energies 13 (3) (2020) 641, https://doi.org/10.3390/en13030641.

[20] L. Kotzur, et al., A modeler's guide to handle complexity in energy systems optimization, Adv. Appl. Energy 4 (2021) 100063, https://doi.org/10.1016/j.adapen.2021.100063.

[21] O. Omoyele, et al., Increasing the resolution of solar and wind time series for energy system modeling: a review, Renew. Sustain. Energy Rev. (2024) 113792, https://doi.org/10.1016/j.rser.2023.113792.

[22] C.F. Peruchena, M. Blanco, A. Bernardos, Generation of series of high frequency DNI years consistent with annual and monthly long-term averages using measured DNI data, Energy Proc. 49 (2014) 2321–2329, https://doi.org/10.1016/j.egypro.2014.03.246.

[23] Y. Sabzevari, S. Eslamian, Reference evapotranspiration in water requirement: theory, concepts, and methods of estimation, in: Handbook of Hydroinformatics, Elsevier, 2023, pp. 269–289, https://doi.org/10.1016/B978-0-12-821961-4.00005-1.

[24] M.J. Reno, C.W. Hansen, Identification of periods of clear sky irradiance in time series of GHI measurements, Renew. Energy 90 (2016) 520–531 [Online]. Available: https://www.osti.gov/servlets/purl/1239983.

[25] P. Lauret, R. Alonso-Suárez, J. Le Gal La Salle, M. David, Solar forecasts based on the clear sky index or the clearness index: which is better? Solar 2 (4) (2022) 432–444, https://doi.org/10.3390/solar2040026. MDPI.

[26] F. Antonanzas-Torres, R. Urraca, J. Polo, O. Perpiñán-Lamigueiro, R. Escobar, Clear sky solar irradiance models: a review of seventy models, Renew. Sustain. Energy Rev. 107 (2019) 374–387, https://doi.org/10.1016/j.rser.2019.02.032.

[27] X. Sun, J.M. Bright, C.A. Gueymard, B. Acord, P. Wang, N.A. Engerer, Worldwide performance assessment of 75 global clear-sky irradiance models using principal component analysis, Renew. Sustain. Energy Rev. 111 (2019) 550–570, https://doi.org/10.1016/j.rser.2019.04.006.

[28] C.A. Gueymard, REST2: high-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation–validation with a benchmark dataset, Sol. Energy 82 (3) (2008) 272–285, https://doi.org/10.1016/j.solener.2007.04.008.

[29] A. Inness, et al., The CAMS reanalysis of atmospheric composition, Atmos. Chem. Phys. 19 (6) (2019) 3515–3556, https://doi.org/10.5194/acp-19-3515-2019.

[30] M. Lefèvre, et al., McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions, Atmos. Meas. Tech. 6 (9) (2013) 2403–2418, https://doi.org/10.5194/amt-6-2403-2013.

[31] M. Larrañeta, C. Fernandez-Peruchena, M.A. Silva-Pérez, I. Lillo-Bravo, Methodology to synthetically downscale DNI time series from 1-h to 1-min temporal resolution with geographic flexibility, Sol. Energy 162 (2018) 573–584, https://doi.org/10.1016/j.solener.2018.01.064.

[32] C. Fernández-Peruchena, M. Blanco, A. Bernardos, Increasing the temporal resolution of direct normal solar irradiance series in a desert location, Energy Proc. 69 (2015) 1981–1988, https://doi.org/10.1016/j.egypro.2015.03.199.

[33] C.M. Fernández-Peruchena, M. Blanco, M. Gastón, A. Bernardos, Increasing the temporal resolution of direct normal solar irradiance series in different climatic zones, Sol. Energy 115 (2015) 255–263, https://doi.org/10.1016/j.solener.2015.02.017.

[34] C.M. Fernández-Peruchena, M. Gastón, A simple and efficient procedure for increasing the temporal resolution of global horizontal solar irradiance series, Renew. Energy 86 (2016) 375–383, https://doi.org/10.1016/j.renene.2015.08.004.

[35] C.M. Fernández-Peruchena, M. Gastón, M. Schroedter-Homscheidt, I.M. Marco, J. L. Casado-Rubio, J.A. García-Moya, Increasing the temporal resolution of direct normal solar irradiance forecasted series, AIP Conf. Proc. 1850 (1) (2017), https://doi.org/10.1063/1.4984515. AIP Publishing.

[36] C.F. Peruchena, M. Larrañeta, M. Blanco, A. Bernardos, High frequency generation of coupled GHI and DNI based on clustered dynamic paths, Sol. Energy 159 (2018) 453–457, https://doi.org/10.1016/j.solener.2017.11.024.

[37] M.J.L.G. Caminero, Synthetic Generation of high-temporal Resolution Direct Normal Irradiation Time Series, Universidad de Sevilla, 2018 [Online]. Available: https://api.semanticscholar.org/CorpusID:139782693.

[38] M. Larrañeta, C. Fernandez-Peruchena, M. Silva-Pérez, I. Lillo-Bravo, A. Grantham, J. Boland, Generation of synthetic solar datasets for risk analysis, Sol. Energy 187 (2019) 212–225, https://doi.org/10.1016/j.solener.2019.05.042.

[39] H.E. Beck, N.E. Zimmermann, T.R. McVicar, N. Vergopolan, A. Berg, E.F. Wood, Present and future köppen-geiger climate classification maps at 1-km resolution, Sci. Data 5 (1) (2018) 1–12, https://doi.org/10.1038/sdata.2018.214.

[40] S. Moreno-Tejera, M.A. Silva-Pérez, L. Ramírez-Santigosa, I. Lillo-Bravo, Classification of days according to DNI profiles using clustering techniques, Sol. Energy 146 (2017) 319–333, https://doi.org/10.1016/j.solener.2017.02.031.

[41] M. Larrañeta, C. Cantón-Marín, M.A. Silva-Pérez, I. Lillo-Bravo, Use of the ND tool: an open tool for the synthetic generation of 1-min solar data from hourly means with geographic flexibility, in: AIP Conference Proceedings, AIP Publishing, 2022, https://doi.org/10.1063/5.0085901 vol. 2445, no. 1.

[42] P. Jiménez-Valero, M. Larrañeta, E. López-García, S. Moreno-Tejera, M.A. Silva-Pérez, I. Lillo-Bravo, Synthetic generation of plausible solar years for long-term forecasting of solar radiation, Theor. Appl. Climatol. 150 (1) (2022) 649–661, https://doi.org/10.1007/s00704-022-04163-9.

[43] I. Balog, G. Caputo, D. Iatauro, P. Signoretti, F. Spinelli, Downscaling of hourly climate data for the assessment of building energy performance, Sustainability 15 (3) (2023) 2762, https://doi.org/10.3390/su15032762.

[44] A. Frimane, J.M. Bright, D. Yang, B. Ouhammou, M. Aggour, Dirichlet downscaling model for synthetic solar irradiance time series, J. Renew. Sustain. Energy 12 (6) (2020), https://doi.org/10.1063/5.0028267.
[45] W. Zhang, W. Kleiber, A.R. Florita, B.-M. Hodge, B. Mather, A stochastic downscaling approach for generating high-frequency solar irradiance scenarios, Sol. Energy 176 (2018) 370–379, https://doi.org/10.1016/j.solener.2018.10.019.
[46] B.W. Forgan, A new method for calibrating reference and field pyranometers, J. Atmos. Ocean. Technol. 13 (3) (1996) 638–645, https://doi.org/10.1175/1520-0426(1996)013%3C0638:ANMFCR%3E2.0.CO;2.
[47] K. Perry, W. Vining, K. Anderson, M. Muller, C. Hansen, Pvanalytics: a Python Package for Automated Processing of Solar Time Series Data, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2022 [Online]. Available: https://www.osti.gov/biblio/1887283.
[48] C.A. Gueymard, Revised composite extraterrestrial spectrum based on recent solar irradiance observations, Sol. Energy 169 (2018) 434–440, https://doi.org/10.1016/j.solener.2018.04.067.
[49] F. Solarspeichersysteme, HTW Berlin Weather Data with a Temporal Resolution of 1 Hz and 1/60Hz (2017-2021), Zenodo, 2024, https://doi.org/10.5281/zenodo.6675646.
[50] S. Leva, A. Nespoli, S. Pretto, M. Mussetta, E. Ogliari, Photovoltaic power and weather parameters, IEEE Dataport 23 (2020), https://doi.org/10.21227/42v0-jz14. September.
[51] A. Driemel, et al., Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017), Earth Syst. Sci. Data 10 (3) (2018) 1491–1501, https://doi.org/10.1594/PANGAEA.880000.
[52] C. Mantuano, O. Omoyele, M. Hoffmann, J.M. Weinand, M. Panella, D. Stolten, Data imputation methods for intermittent renewable energy sources: implications for energy system modeling, Energy Convers. Manag. 339 (2025) 119857, https://doi.org/10.1016/j.enconman.2025.119857.
[53] W.F. Holmgren, C.W. Hansen, M.A. Mikofski, Pvlib python: a python package for modeling solar energy systems, J. Open Source Softw. 3 (29) (2018) 884, https://doi.org/10.21105/joss.00884.
[54] J. Stein, C. Hansen, M.J. Reno, The Variability Index: a New and Novel Metric for Quantifying Irradiance and PV Output Variability, Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA, 2012 [Online]. Available: https://www.osti.gov/servlets/purl/1078490.
[55] S. Moreno-Tejera, M. Larrañeta, I. Lillo-Bravo, M. Silva-Pérez, A normalized variability index of daily solar radiation, in: AIP Conference Proceedings, AIP Publishing, 2020, https://doi.org/10.1063/5.0028919 vol. 2303, no. 1.
[56] R. Blaga, M. Paulescu, Quantifiers for the solar irradiance variability: a new perspective, Sol. Energy 174 (2018) 606–616, https://doi.org/10.1016/j.solener.2018.09.034.
[57] M. Kleinebrahm, J.M. Weinand, E. Naber, R. McKenna, A. Ardone, W. Fichtner, Two million European single-family homes could abandon the grid by 2050, Joule 7 (11) (2023) 2485–2510, https://doi.org/10.1016/j.joule.2023.09.012.
[58] T. Klütz, et al., ETHOS. FINE: a framework for integrated energy System assessment, J. Open Source Softw. 10 (105) (2025) 6274, https://doi.org/10.21105/joss.06274.
[59] D. Fischer, T. Wolf, J. Scherer, B. Wille-Haussmann, A stochastic bottom-up model for space heating and domestic hot water load profiles for German households, Energy Build. 124 (2016) 120–128, https://doi.org/10.1016/j.enbuild.2016.04.069.
[60] K. Knosala, et al., Hybrid hydrogen home storage for decentralized energy autonomy, Int. J. Hydrogen Energy 46 (42) (2021) 21748–21763, https://doi.org/10.1016/j.ijhydene.2021.04.036.
[61] M. Hoffmann, et al., A review of mixed-integer linear formulations for framework-based energy system models, Adv. Appl. Energy (2024) 100190, https://doi.org/10.1016/j.adapen.2024.100190.
[62] M. Kittel, H. Hobbie, C. Dierstein, Temporal aggregation of time series to identify typical hourly electricity system states: a systematic assessment of relevant cluster algorithms, Energy 247 (2022) 123458, https://doi.org/10.1016/j.energy.2022.123458.
[63] L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, Impact of different time series aggregation methods on optimal energy system design, Renew. Energy 117 (2018) 474–487, https://doi.org/10.1016/j.renene.2017.10.017.
[64] P. Esling, C. Agon, Time-series data mining, ACM Comput. Surv. 45 (1) (2012) 1–34, https://doi.org/10.1145/2379776.2379788.
[65] K. Cugerone, C. De Michele, A. Ghezzi, V. Gianelle, Aerosol removal due to precipitation and wind forcings in milan urban area, J. Hydrol. 556 (2018) 1256–1262, https://doi.org/10.1016/j.jhydrol.2017.06.033.