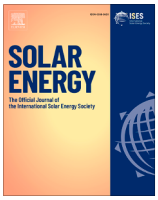
Photovoltaic energy conversion models with clustering-based classification of days from irradiance data

Solar Energy 304 (2026) 114178

Contents lists available at ScienceDirect

Solar Energy

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



Photovoltaic energy conversion models with clustering-based classification of days from irradiance data

Gabriele Malgaroli [a], Carmen Lucia Tancredo Borges [b], Filippo Spertino [a], Alessandro Ciocia [a], David Alexander [b], Gianfranco Chicco [a,*]

[a] Politecnico di Torino, Dipartimento Energia "Galileo Ferraris", Corso Duca degli Abruzzi 24, Torino, 10129, Italy
[b] Federal University of Rio de Janeiro, Electrical Engineering Dept., Rio de Janeiro, Brazil

HIGHLIGHTS

- Methodology for classifying typical days based on irradiance patterns.
- Comparison between clearness and clear sky indices for weather class identification.
- Statistical approach based on data clustering to validate day classification.
- Data collection from locations with different latitudes and PV system configurations.
- Comparison of different energy conversion models to estimate the PV power output.

ARTICLE INFO

Keywords:
Clearness index
Clear sky index
Energy models
Weather classification
Clustering
Polito model

ABSTRACT

In the literature, many contributions compare the effectiveness of different energy models for photovoltaic (PV) generators in specific installation sites. Most techniques are analytical or based on electrical equivalent circuits. However, only a few works investigate the performance of the models in different weather classes. The parameter mostly used to classify the days into weather categories is the clearness index. An alternative parameter is the clear sky index. However, information about their values and ranges in different weather conditions is missing in the literature. This paper is based on data gathered in Italy, Brazil and Portugal, at different latitudes, with fixed and sun-tracking configurations of the PV generators, and for different weather conditions, and has a twofold objective. First, the clearness index is shown to be inappropriate to represent general situations, while the clear sky index is suitable to identify consistent ranges that represent different sunny, partly cloudy and cloudy days, at different latitudes and PV configurations, by using a clustering procedure. Next, the effectiveness of the PV energy conversion models for different installation sites is evaluated, showing which model is more suitable for each site and type of day. In this context, a model based on single diode model is proposed, formulated after numerically determining the parameters of the equivalent circuit. The parameters under reference conditions and the coefficients of the equations that quantify the dependency on weather conditions have been numerically optimized starting from experimental datasets of PV modules.

1. Introduction

Nowadays, the global energy consumption has increased significantly due to factors like population growth and urbanisation. In this context, the adoption of Renewable Energy Systems (RES) has become fundamental to supply the increasing energy demand while limiting the polluting emissions, driving the energy transition [1]. PhotoVoltaic (PV) plants are among the most widespread RESs thanks to their high lifetime and reliability [2], low maintenance requirements [3], and positive impact on reducing the frequency of voltage sags in the grid [4]. Currently, one line of research in this field consists of developing models to assess the irradiance availability in sites without measurements for future PV production [5]. Another line refers to the identification of

* Corresponding author.
Email address: gianfranco.chicco@polito.it (G. Chicco).

https://doi.org/10.1016/j.solener.2025.114178
Received 16 May 2025; Received in revised form 6 November 2025; Accepted 19 November 2025
Available online 26 November 2025

Nomenclature

List of acronyms

ABC	All Back Contact
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
CDF	Cumulative Distribution Function
HJT	Heterojunction
m-Si	Monocrystalline silicon
MBE	Mean Bias Error (%)
MPP	Maximum Power Point
$NOCT$	Nominal Operating Cell Temperature (K)
p-Si	Polycrystalline silicon
PoA	Plane of Array
PV	Photovoltaic
PVGIS	PhotoVoltaic Geographical Information System
RES	Renewable Energy Systems
$NRMSE$	Normalized Root Mean Square Error (%)
SAM	System Advisor Model
SDM	Single Diode Model
STC	Standard Test Conditions
TOPCon	Tunnel Oxide Passivated Contact

List of symbols

A	Empirical coefficient of ASHRAE model (-)
A_{Bra}	Empirical coefficient of ASHRAE model for the site in Brazil ($\mathrm{W/m^2}$)
A_{SAM}	Empirical coefficient of SAM model (-)
$a_{i,j}$	Average distance between the i^{th} point and the elements of the same j^{th} cluster (-)
B	Empirical coefficient of ASHRAE model (-)
B_{Bra}	Empirical coefficient of ASHRAE model for the site in Brazil (-)
B_{SAM}	Empirical coefficient of SAM model (-)
$b_{i,j}$	Minimum average distance between the i^{th} point of the j^{th} cluster and the elements of the closest cluster (-)
C_{Bra}	Seasonality corrective coefficient of the ASHRAE model for the site in Brazil (-)
E_g	Energy gap (eV)
$E_{\mathrm{g,STC}}$	Energy gap at STC (1.121 eV)
f_k	PV power estimated by the model at k^{th} instant (W)
G	Solar irradiance ($\mathrm{W/m^2}$)
G_{low}	Threshold irradiance in the Park model (150 $\mathrm{W/m^2}$)
G_{NOCT}	Irradiance at Nominal Operating Cell Temperature conditions (800 $\mathrm{W/m^2}$)
G_{STC}	Irradiance at STC (1000 $\mathrm{W/m^2}$)
G_{T}	PoA irradiance ($\mathrm{W/m^2}$)
$G_{\mathrm{T,clear}}$	Clear sky irradiance on a tilted surface ($\mathrm{W/m^2}$)
H_{ex}	Extraterrestrial irradiation ($\mathrm{W/m^2}$)
H_{g}	Global irradiation ($\mathrm{W/m^2}$)
$H_{\mathrm{g,clear}}$	Clear sky irradiation ($\mathrm{W/m^2}$)
h_{s}	Solar altitude (°)
I	Current in the PV module (A)
I_0	Reverse saturation current (A)
$I_{\mathrm{0,STC}}$	Reverse saturation current at STC (A)
$I_{\mathrm{mpp,STC}}$	MPP current at STC (A)
I_{ph}	Photogenerated current (A)
$I_{\mathrm{ph,STC}}$	Photogenerated current at STC (A)
$I_{\mathrm{sc,STC}}$	Short-circuit current at STC (A)
k_{B}	Boltzmann constant ($1.38 \cdot 10^{-23}$ J/K)
k_{C}	Clear sky index (-)
$k_{\mathrm{R_s}}$	Irradiance-dependent term in the equation of the series resistance (-)
k_{T}	Clearness index (-)
N	Number of data (-)
N_{c}	Number of groups in the clustering technique (-)
N_{d}	Number of days of the experimental campaign (-)
N_{s}	Number of series-connected cells in a module (-)
n	Diode ideality factor (-)
n_0	Constant term in the polynomial approximation of n (-)
n_{G}	Irradiance-dependent term in the polynomial approximation of n ($\mathrm{m^2/W}$)
n_{T}	Temperature-dependent term in the polynomial approximation of n (1/K)
P_{nom}	Nominal PV power (W)
P_{PV}	PV power (W)
P_{STC}	PV power at STC (W)
q	Charge of electrons ($\approx 1.6 \cdot 10^{-19}$ C)
R^2	Coefficient of determination (-)
R_{s}	Series resistance (Ω)
$R_{\mathrm{s,STC}}$	Series resistance at STC (Ω)
R_{sh}	Shunt resistance (Ω)
$R_{\mathrm{sh,STC}}$	Shunt resistance at STC (Ω)
r_j	Number of points in the j^{th} cluster (-)
S	Surface of the PV generator ($\mathrm{m^2}$)
S_{a}	Average silhouette (-)
$s_{i,j}$	Local silhouette coefficient of the i^{th} point in the j^{th} cluster (-)
T_{air}	Air temperature (K)
$T_{\mathrm{air,NOCT}}$	Air temperature at NOCT conditions (293.15 K)
T_{PV}	Photovoltaic module temperature (K)
T_{STC}	Module temperature at STC (298.15 K)
V	Voltage of the PV module (V)
V_{f}	Wind speed (m/s)
V_{j}	Voltage of the p-n junction (V)
$V_{\mathrm{mpp,STC}}$	MPP voltage at STC (V)
$V_{\mathrm{oc,STC}}$	Open-circuit voltage at STC (V)
w	Module assembly parameter (-)
$\bar{y}$	Average PV power from measurements over the period under analysis (W)
y_{i}	Measured PV power at $\mathrm{i^{th}}$ instant (W)
$\alpha_{\mathrm{I_{sc}}}$	Temperature coefficient of short-circuit current (%/K)
$\beta_{\mathrm{V_{oc}}}$	Temperature coefficient of voltage (%/K)
γ	Temperature coefficient of maximum power (%/K)
ΔT	Temperature difference between T_{PV} and T_{STC} (K)
ΔT_{SAM}	Temperature difference between irradiated cells and their dark surface in the SAM model (K)
η_{PV}	Conversion efficiency of the PV module (%)
η_{nom}	Nominal conversion efficiency of the PV module (%)
θ	Angle between the normal to the surface and the incident irradiance (°)
θ_{t}	Tilt angle of a surface with respect to the horizontal plane (°)
θ_{z}	Zenith angle (°)

accurate energy conversion models to estimate the PV production based on irradiance and other environmental measurements.

In the literature, many techniques have been proposed to accurately predict PV power over a given time period [6], with the development of probabilistic techniques [7]. A detailed review of PV power prediction methods is outside the scope of this paper. However, a short description of the models currently used in the literature is presented to highlight the novelty of this work. The work [8] carried out a comparison of

analytical energy models for an installation site in Brazil, highlighting the overestimation of PV power by models that do not take into account the modules' temperature variation. The models with best performance were then applied to evaluate PV reliability [9]. The authors in [10] compared the physical models required for PV power prediction in Hungary, and the energy models under analysis were analytical or based on the Single Diode Model (SDM) [11]. The works [12,13] compared the energy estimated by five analytical techniques for PV modules with different technologies installed in Spain. The analyses carried out in these works rely on the application of analytical or diode-based models.

The analytical models estimate the PV production at the Maximum Power Point (MPP) without providing any information regarding the other operating points of the current-voltage (*I-V*) curve [14,15]. These models are based on equations that do not require numerical methods for their solution, with a negligible computational burden. However, their accuracy is strongly affected by the technology and electrical characteristics of PV modules under test as these models include empirical coefficients that are valid for specific modules only.

On the contrary, the diode-based models rely on electrical equivalent circuits [16]. These models permit achieving higher accuracy of power estimation with respect to analytical models, and the whole *I-V* curve of the PV generators can be determined for any operating condition. The main drawbacks of diode-based models concern the high computational burden required to solve their implicit equations and the need to know the parameters of the equivalent circuits at reference conditions, as well as the dependency of each parameter on weather conditions. The knowledge of reference parameters and weather-dependent equations is required to trace the *I-V* curve of PV modules at any operating condition. However, despite these aspects being fundamental, very few papers in the literature provide quantitative estimations based on experimental studies, and most of them use reference parameters assumed by old research works. In addition, the coefficients of equations used to estimate the parameters under real conditions are not provided, with the risk of considering values valid for old modules with outdated cell technologies only.

Some works investigated the effectiveness of energy models for days with different weather conditions. The classification of the days can be qualitative or quantitative: in the first case, the days are manually classified by observing the shape of the daily irradiance profile. This is possible with low number of days under analysis. On the contrary, studies with a quantitative classification rely on clustering techniques to group the days of the campaign into different classes. The clearness index [17] and the clear sky index can provide information about the clearness of a day; however, information about their values and ranges for different weather conditions is currently missing in literature. Other indices are based on power measurements compared with simulated power outputs from clear-sky and PV array performance models [18].

The paper also presents a model that is an extension, for different datasets and solar cell technologies, of that proposed in [19], which is based on equivalent circuits, here named Polito model. This model includes sets of reference parameters that are determined for PV modules with different cell technologies, rated power, and manufacturing years. In addition, the model includes analytical equations to estimate the parameters under variable weather conditions, with their coefficients being the result of numerical optimizations on experimental measurements. The procedure used to identify the sets of reference parameters and coefficients is presented in [19]. The model is validated for different latitude locations, subject to different weather conditions (Brazil, Italy and Portugal) using a statistical approach based on data clustering. In this context, a methodology for classifying typical days based on irradiance patterns (as sunny, partly cloudy and cloudy days) is proposed to evaluate the performance of the energy conversion models in specific weather classes. The adequacy of using the clear sky index for a global classification is evaluated.

The next sections of this paper are organized as follows. Section 2 presents the general structure of the proposed methodology. Section 3 discusses the theoretical background of each step of the methodology. Section 4 describes the PV systems and the experimental campaigns considered in this analysis. Section 5 presents the results after the application of the methodology to the three installation sites under study. Section 6 contains the conclusions.

2. Proposed methodology

This section presents a methodology to evaluate the performance of the energy models for the installation sites analyzed.

The methodology includes two procedures. The first procedure ("Classification in three weather categories and application of energy models"), permits to classify the days of an experimental campaign into three weather classes, corresponding to sunny, partly cloudy and cloudy days, previously determined through the clustering of clear sky indices. Hence, this procedure permits the evaluation of the effectiveness of energy models for installation sites with different latitudes, sun-tracking configurations, and weather conditions. The second procedure ("Application of energy models to global data and mutual comparison"), compares the different energy models under study and permits the identification of the most accurate model for each installation site.

The flowchart of the methodology is presented in Fig. 1 and consists of the following steps:

- *Step 1 - Acquisition of tilted irradiance from measurements and databases*. The two procedures require the knowledge of the global Plane of Array (PoA) irradiance G_T, i.e., the irradiance reaching a tilted surface with the same inclination of the PV modules with respect to the horizontal plane. This step aims to collect profiles of PoA irradiance for each PV system under study from measurements and from databases.
- *Step 2A - Selection of the model for clear sky irradiance estimation*.

  Several models are proposed in the literature to estimate the irradiance under clear sky conditions for specific site and installation conditions. These models are summarized in the work [20] with indication of their inputs. In this step, the most adequate clear sky model is selected considering the information available for the analysis. Two parameters commonly used to perform the classification of days into different weather classes are the clearness and the clear sky indices.

  The clear sky index has been selected for the reasons indicated in Section 3.2.
- *Step 3A - Evaluation of daily clear sky index*. This step evaluates the clear sky indices for each day of an extended dataset that contains the daily clear sky indices for a multi-year period, considering the locations corresponding to the experimental datasets.
- *Step 4A - Clustering of daily indices and identification of ranges*. The dataset of each location under study is divided into the three groups, corresponding to sunny, partly cloudy and cloudy days, by adopting an adequate clustering technique. Moreover, this step evaluates the possibility of identifying general ranges of clear sky indices to represent the different types of days for each location.
- *Step 5 - Selection of energy conversion models*. Most of the techniques used in the literature to estimate the PV energy production are analytical equations with semi-empirical coefficients. Nevertheless, advanced electrical models are becoming widely used in the research field and in commercial software. This step identifies the energy conversion models most suitable for assessing the PV energy starting from input variables like solar irradiance, air temperature, and wind speed. The two procedures of the methodology use these energy models to evaluate their effectiveness in different weather conditions and select the best model for energy estimation by comparison of global results.
- *Step 6A - Application of energy conversion models to clustered datasets*. The selected energy conversion models are applied to the three clustered datasets of each installation site, and their effectiveness in the

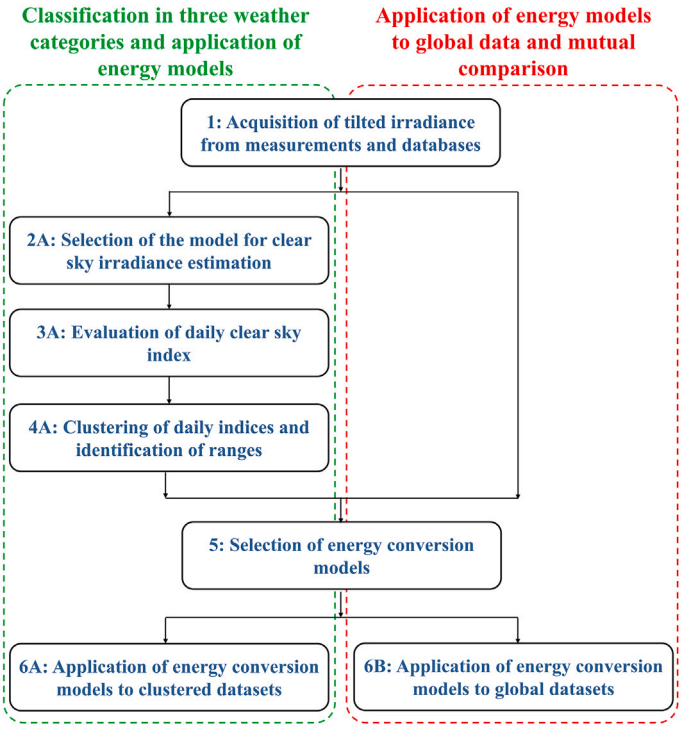


Fig. 1. Structure and steps of the proposed methodology.

different weather categories is evaluated through the calculation of proper indicators.

- *Step 6B - Application of energy conversion models to global datasets*. The selected energy conversion models are applied to the global datasets of each location, and their accuracy is assessed by comparing proper indicators.

3. Theoretical background

This section presents the theoretical background for each step of the methodology described in the previous section.

3.1. Selection of the model for clear sky irradiance estimation

The ASHRAE model is one of the most common models used to estimate the clear-sky irradiance reaching a surface with an inclination relative to the horizontal plane. The global clear-sky irradiance on a surface with tilt angle θ_t relative to the horizontal plane ($G_{T,clear}$) is estimated by the ASHRAE model as follows:

$$G_{T,clear} = A \cdot \cos\theta \cos\theta_z \cdot e^{-\frac{B}{\cos\theta_z}} \quad (1)$$

where A and B are empirical coefficients whose values are affected by the month of the year [21], θ_z is the zenith angle (i.e., the angle between the incident irradiance and the vertical plane), h_s is the solar altitude (i.e., the angle between the terrain and the incident global irradiance) and is the complementary angle to θ_z, and θ is the angle between the normal to the surface and the incident irradiance (Fig. 2).

In the literature, more complex models are available for assessing the clear sky irradiance, as in [20]. However, in this methodology, the ASHRAE model is suggested due to its good trade-off between high accuracy and simplicity.

3.2. Evaluation of the daily indices

Two indices permit quantifying the degree of clearness for a day: the clearness index k_T and the clear sky index k_C.

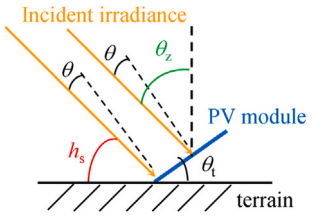


Fig. 2. Solar angles.

For a specific inclination and time period, the clearness index is defined in the following way:

$$k_T = H_g / H_{ex} \quad (2)$$

where H_g is the global irradiation, and H_{ex} is the extraterrestrial irradiation. The clearness index is a stochastic variable that can be defined by using probability distributions based on the meteorology [22].

Conversely, the *clear sky index* is defined as follows [23]:

$$k_C = H_g / H_{g,clear} \quad (3)$$

where $H_{g,clear}$ is the clear sky irradiation. This index can be defined for different time frames. The work [23] defines this parameter as the ratio between irradiances; however, the daily clear sky index is, generally, calculated for this type of analyses as the ratio between irradiation values. The clearness index can be determined at different time steps as well. For a fixed photovoltaic plant, the clearness index ranges from 0 to 1 [24,25]. However, the validity of using the clearness index might be limited in PV installations with sun-tracking systems, because values higher than unity could be reached depending on the type of sun-tracking option considered. The clear sky index ranges from 0 to 1 also in case of sun-tracking PV installations and as such is appropriate for generalized comparisons. For this reason, the clear sky index is used in this paper and is determined for each day.

3.3. Clustering of daily indices and identification of ranges

Let us consider the daily values of the clear sky index obtained over a long period of time (e.g., N_d days spanning several years). From these data, the N_d days are partitioned into N_c groups by using a clustering technique. The number of groups is user-defined. In particular, without loss of generality, $N_c = 3$ types of days are considered in this paper, corresponding to sunny, partly cloudy and cloudy days.

The main goal is to identify possible ranges of variation of the clear sky index that can be considered as a common ground under different conditions. The following procedure is used:

- Apply a clustering algorithm whose execution depends on random number extractions, such that different results may be obtained for different executions and a statistical analysis of the results can be applied.
- For a given PV site, run the clustering algorithm for a statistically significant number of executions and construct the cumulative distribution functions (CDFs) of the limits of the intervals between cloudy and partly cloudy days, as well as between partly cloudy and sunny days. From the results, determine the ranges of the three types of days that can be interpreted (using a terminology borrowed from fuzzy logic) as *membership degrees* for each type of day.
- Repeat the calculations for different PV sites to obtain a set of results for each site.
- Compare the results obtained from the different sites and synthesize the ranges of the three types of days with global membership degrees.

The choice of the clustering technique and the assessment of the effectiveness of the results by using an appropriate clustering validity indicator are illustrated in the sequel.

3.3.1. Clustering technique

Among the many clustering algorithms available, the k-medoids algorithm has been selected, for the following main reasons [26]:

1. The algorithm is based on random number extractions and enables the user to conduct a statistical analysis based on a significant number of executions.
2. Each cluster is represented by a medoid, that is, the *actual* clear sky index closest to the average clear sky index determined for the days clustered into the same group, as the average clear sky index could not correspond to any actual value.
3. The kmedoids algorithm is robust to the presence of outliers or noise in the data.

3.3.2. Clustering validity indicators

The clustering validity indicator considered is the average silhouette S_a [27], evaluated by computing the local silhouette coefficients $s_{i,j}$ for each cluster $j = 1,\ldots, N_c$ that contains r_j points:

$$S_a = \frac{1}{N_c} \cdot \sum_{j=1}^{N_c} \left(\frac{1}{r_j} \cdot \sum_{i=1}^{r_j} s_{i,j} \right) \tag{4}$$

The local silhouette coefficient $s_{i,j}$ of the i^{th} point of the j^{th} cluster is evaluated according to the following equation:

$$s_{i,j} = \frac{b_{i,j} - a_{i,j}}{\max\{b_{i,j}, a_{i,j}\}} \tag{5}$$

where $a_{i,j}$ is the average distance between the i^{th} point and the elements of the same j^{th} cluster, and $b_{i,j}$ is the minimum average distance between the i^{th} point of the j^{th} cluster and the elements of the closest cluster to the j^{th} cluster.

The local silhouette coefficients vary from -1 to 1, with higher values corresponding to appropriate location of the day in the cluster to which it has been assigned. For the average silhouette index, indicatively, values higher than 0.7 represent a strong clustering structure, values higher than 0.5 a reasonable clustering structure, while values lower than 0.5 (including negative values) represent weak or poor clustering structures.

3.4. Selection of the energy conversion models

In the literature, many irradiance-to-power models permit to determine the PV power starting from the incident solar irradiance. This subsection describes the most common models in the literature, namely the Osterwald [28], Park [29], SAM [30] and Skoplaki [31] models. In addition, a model proposed by Politecnico di Torino, called "Polito model", is included in this description.

3.4.1. Polito model

The *I-V* curve of PV modules can be described by electrical equivalent circuits with a number of parameters ranging between three and nine. The Single Diode Model (SDM) is a good compromise between high accuracy and low computational burden [11] and its equation at cell level [16] is the following:

$$I = I_{ph} - I_0 \cdot \left(\exp\left(\frac{q \cdot V_j}{n \cdot k_B \cdot T_{PV}} \right) - 1 \right) - \frac{(V + R_s \cdot I)}{R_{sh}} \tag{6}$$

where q is the charge of electrons ($\approx 1.6 \cdot 10^{-19}$ C), k_B the Boltzmann constant ($1.38 \cdot 10^{-23}$ J/K), V_j is the p-n junction voltage, T_{PV} is the module temperature expressed in K, I_{ph} is the photogenerated current, I_0 is the reverse saturation current, n is the diode ideality factor, R_s is the series resistance, and R_{sh} is the shunt resistance. As the resistances are affected by the number of cells N_s in a module, their values at the module level are obtained by multiplying the cell resistances by N_s.

SDM-based models determine the power of PV modules starting from the knowledge of the SDM parameters. However, these quantities are affected by weather conditions like irradiance and air temperature, and, for each parameter, the quantitative knowledge of such a dependency is required to trace the *I-V* curves of PV modules under any operating condition. In particular, SDM-based models require the knowledge of the parameters at reference conditions, as well as coefficients to quantify their dependency on irradiance and module temperature. However, most of the models proposed in the literature might be effective for old modules with old cell technologies only. This might be due to the source of sets of reference parameters and coefficients since these models are, generally, formulated starting from old research works.

On the contrary, the Polito model includes sets of reference parameters that are determined for PV modules with different cell technologies, rated power, and manufacturing years. In addition, the model includes analytical equations to estimate the parameters under variable weather conditions, with the coefficients of such equations being the result of numerical optimizations based on experimental measurements. The Polito model consists of Eqs. (6)–(11) and the coefficients ($I_{ph,STC}$, $I_{0,STC}$, n_0, n_G, n_T, $R_{s,STC}$, k_{Rs}, and $R_{sh,STC}$) are identified by following the procedure reported in [19], where the model was applied to a high-efficiency PV module with all-back contact monocrystalline silicon technology.

The equations describing the dependency of each equivalent circuit's parameter on weather conditions in the Polito model are as follows [32]:

$$I_{ph} = I_{ph,STC} \cdot \left[1 + \alpha_{I_{sc}} \cdot \left(T_{PV} - T_{STC} \right) \right] \cdot \frac{G_T}{G_{STC}} \tag{7}$$

where $\alpha_{I_{sc}}$ is the temperature coefficient for short-circuit current, T_{STC} is 298.15 K, G_T is PoA irradiance (W/m^2), G_{STC} is 1000 W/m^2, and $I_{ph,STC}$ is the photogenerated current at Standard Test Conditions (STC);

$$I_0 = I_{0,STC} \cdot \left(\frac{T_{PV}}{T_{STC}}\right)^3 \cdot \exp\left[\left(\frac{E_{g,STC}}{T_{STC}} - \frac{E_g(T_{PV})}{T_{PV}}\right) \cdot \frac{1}{k_b}\right] \tag{8}$$

where $E_g(T_{PV})$ and $E_{g,STC}$ are the energy gaps of the semiconductor material evaluated at temperature T_{PV} and at STC, respectively, and $I_{0,STC}$ is the reverse saturation current at STC;

Table 1
Coefficients of the Polito model.

Coefficients	p-Si	old m-Si	m-Si
$I_{ph,STC}$(A)	8.64–9.42	5.74–10.08	10.47
$I_{0,STC}$ (A) (10^{-9})	8.90	20.8	0.17
n_0 (-)	0.80	1.14	2.21
n_G (m^2/W)(10^{-5})	−25.7	5.42	15.9
n_T (1/K)(10^{-4})	20.3	3.74	−37.4
$R_{s,STC}$ (Ω) (10^{-5})	5 ·N_s	5 ·N_s	0.05 ·N_s
k_{Rs} (-)	0.15	0.04	−0.11
$R_{sh,STC}$ (Ω)	3.31 ·N_s	16.7 ·N_s	3.76 ·N_s

$$n = n_0 + n_G \cdot G_T + n_T \cdot T_{PV} \tag{9}$$

where the dependence of n on weather conditions is not unique in the literature: indeed, a quantitative dependence of the diode ideality factor with respect to the weather variables, irradiance and temperature, is not provided. Previous studies [33,34] showed a weak dependence of such a parameter on weather conditions, and most of the papers assume a constant value for it [32,35]. In this work, the proposed model attempts to find a more accurate modeling of the diode ideality factor by assuming a linear dependence on irradiance and module temperature. In this context, the term n_0 represents the constant value, and the terms n_G and n_T quantify the mentioned dependence on irradiance and module temperature, respectively. This assumption was confirmed by the results of the analysis carried out in [19];

$$R_s = R_{s,STC} \cdot \frac{T_{PV}}{T_{STC}} \cdot \left(1 - k_{Rs} \cdot \ln\left(\frac{G_T}{G_{STC}}\right)\right) \tag{10}$$

where $R_{s,STC}$ is the series resistance at STC and k_{Rs} is a coefficient that represents the impact of the irradiance on R_s;

$$R_{sh} = R_{sh,STC} \cdot \frac{G_{STC}}{G_T} \tag{11}$$

where $R_{sh,STC}$ is the shunt resistance at STC.

For each (G_T, T_{PV}) data, the equivalent circuit's parameters are evaluated according to Eqs. (7)–(11). Such parameters are used to solve Eq. (6) and trace the *I-V* curve at module level; then, the Maximum Power Point (MPP) is identified.

The analysis presented in the paper [19] was applied to seven PV modules with different cell technologies, efficiency, rated power, and manufacturing years. The modules were classified in three groups: three old modules with monocrystalline silicon (m-Si) technology, one recent module with newer m-Si technology, and three modules with polycrystalline silicon (p-Si) technology. The modules were tested under natural sunlight in long experimental campaigns to determine one set of parameters and coefficients ($I_{ph,STC}$, $I_{0,STC}$, n_0, n_G, n_T, $R_{s,STC}$, k_{Rs}, $R_{sh,STC}$) for each PV generator. Except for $I_{ph,STC}$, the value of the coefficients was the average between that of the modules in each group. The results of such an analysis are reported in Table 1. The coefficient $I_{ph,STC}$ has been assumed equal to the short-circuit current of the PV generators at STC ($I_{sc,STC}$). The reduction of $I_{0,STC}$ in newer m-Si modules is due to the better performance of PV modules under high-temperature conditions, resulting in better temperature coefficients. The values of the diode ideality factor are almost in the range between 1 and 2: in this case, the dependence on weather conditions is weak, and the highest impact is due to the constant term n_0. The most evident change related to the series and shunt resistances is the reduction of $R_{s,STC}$ in newer m-Si modules. This effect is mainly due to the new technology of the PV module analysed [19], which has back-contact cells that enhance the exploitation of the area and the conversion efficiency of the module. In addition, the series and shunt resistances at STC are provided as functions of the number of cells in the module N_s to make the model applicable to PV modules with different dimensions and numbers of cells.

3.4.2. Osterwald model

The Osterwald model is a simple irradiance-proportionality model suggested in [28]:

$$P_{PV} = P_{STC} \cdot \frac{G_T}{G_{STC}} \cdot (1 + \gamma \cdot \Delta T) \tag{12}$$

$$T_{PV} = T_{air} + G_T \cdot \frac{NOCT - T_{air,\,NOCT}}{G_{NOCT}} \tag{13}$$

where T_{air} is the ambient temperature, $NOCT$ is the Nominal Operating Cell Temperature, $T_{air,\,NOCT}$ is the air temperature at NOCT conditions (293.15 K), G_{NOCT} is the solar irradiance at NOCT conditions (800 W/m^2), P_{PV} is the PV output power, P_{STC} is the PV power under STC conditions, γ is the temperature coefficient for power (in %/K), and $\Delta T = T_{PV} - T_{STC}$.

3.4.3. SAM model

With respect to the Osterwald model, this technique includes additional coefficients affected by the installation of PV modules, which are determined empirically. Indeed, PV power is estimated as follows [30]:

$$P_{PV} = G_T \cdot S \cdot \eta_{PV} \tag{14}$$

$$\eta_{PV} = \eta_{nom} \cdot [1 + \gamma \cdot (T_{PV} - T_{STC})] \tag{15}$$

$$T_{PV} = T_{air} + G_T \cdot \exp^{A_{SAM} + B_{SAM} \cdot V_f} + G_T \cdot \frac{\Delta T_{SAM}}{G_{STC}} \tag{16}$$

where S is the area of the photovoltaic array, η_{PV} is the PV efficiency, η_{nom} is the nominal PV efficiency, ΔT_{SAM} is the temperature difference between the irradiated cells and the dark-coloured surface of the module, A_{SAM} and B_{SAM} are empirical coefficients tabulated in [30], and V_f is the wind speed.

In this work, the coefficients for modules with glass/cell/polymer sheet structure and installed in an open rack are selected from King et al. [30], resulting in $A_{SAM} = -3.56$, $B_{SAM} = -0.0750$, and $\Delta T_{SAM} = 3$ K.

3.4.4. Park model

The Park model [29] considers different equations for PV power estimation according to the solar irradiance value:

$$\begin{cases} P_{STC}\left(\frac{G_T^2}{G_{STC} G_{low}}\right) & \text{for} \quad 0 \le G_T \le G_{low} \\ P_{STC}\left(\frac{G_T}{G_{STC}}\right) & \text{for} \quad G_{low} < G_T \le G_{STC} \\ P_{STC} & \text{for} \quad G_T > G_{STC} \end{cases} \tag{17}$$

where G_{low} is a threshold irradiance, typically ≈ 150 W/m^2.

3.4.5. Skoplaki model

Proposed in [31], the Skoplaki model takes into account the effects of reduced efficiency due to wind and module assembly, according to the following equation:

$$P_{PV} = 0.12 \cdot S \cdot G_T \cdot [1 - 0.004 \cdot (T_{air} + w \frac{0.32}{8.91 + 2.0 \cdot V_f} G_T - 25)] \tag{18}$$

where w is a module assembly parameter provided in [31], usually close to unity.

3.5. Application of the energy conversion models to clustered and global datasets

The above described models are applied to the available datasets to assess the PV power and their effectiveness is evaluated through the estimation of the three statistical indicators [36–38] reported below.

- Coefficient of determination: denoted as R^2, ranges between 0 (bad approximation) and 1 (good approximation) and quantifies

the adherence of the PV power determined by the model to the N measurements as follows:

$$R^2 = 1 - \frac{\sum_{k=1}^{N} (y_k - f_k)^2}{\sum_{k=1}^{N} (y_k - \bar{y})^2} \quad (19)$$

where y_k is the measured PV power at the k^{th} time instant, f_k is the PV power determined by the model at the k^{th} time instant, and $\bar{y}$ is the average value of the measured PV power over the period under analysis.

- Normalized Root Mean Square Error ($NRMSE$): estimates the square deviations between the model outcomes and the measured data, normalized with respect to the average measured value. Low values correspond to good approximation by the model under analysis:

$$NRMSE = 100 \frac{\sqrt{\frac{\sum_{k=1}^{N} (f_k - y_k)^2}{N}}}{\bar{y}} \quad (20)$$

- Mean Bias Error (MBE): calculates the deviations between the model outcomes and the measurements, with low values corresponding to good approximation by the model under analysis:

$$MBE = 100 \frac{\frac{\sum_{k=1}^{N} (f_k - y_k)}{N}}{\bar{y}} \quad (21)$$

These indicators can be calculated considering the global dataset or any clustered dataset. When the energy models are applied to the three clustered datasets, the three indicators are computed for each of them, as $(R^2, NRMSE, MBE)_{\text{sunny}}$, $(R^2, NRMSE, MBE)_{\text{partly}}$ and $(R^2, NRMSE, MBE)_{\text{cloudy}}$. Such indicators permit quantifying the effectiveness of the energy models in different weather conditions for each installation site under study.

4. Description of the measurement systems

The experimental campaigns were conducted in three different countries (Italy, Brazil and Portugal) using outdoor measurement stations. The specifications of the modules under test are reported in Table 2.

4.1. Experimental campaign in Italy

The Italian experimental campaign was carried out in Torino (Italy) for one PV module with rated power of 410 W, and monocrystalline silicon half cells technology.

The module was South-oriented with an inclination with respect to the ground θ_t of 30°. The measurement setup used consists of the following components:

- One programmable electronic load with maximum power of 800 W (maximum voltage 120 V and maximum current 16 A), connected to a single module under test. This device has a resolution of 16 bits, and relative uncertainty lower than ±0.15 % (voltage) and ±0.25 % (current).
- One PT100 sensor (resolution of 0.1 °C, and relative uncertainty lower than ±0.3 °C) gathers the PV rear temperature;
- One secondary standard pyranometer (measurement range 0–2000 W/m^2 and relative uncertainty of ±20 W/m^2) gathers the in-plane solar irradiance G;
- One weather station includes one anemometer to acquire the wind speed (resolution of 0.01 m/s and relative uncertainty of ±1 %) and its direction (resolution 0.1° and relative uncertainty of ±1 %), one piezoresistive sensor to acquire the atmospheric pressure (resolution of 0.1 mbar, relative uncertainty of ±0.4 mbar), one PT100 sensor to measure the ambient temperature (resolution of 0.1 °C, relative uncertainty lower than ±0.3 °C), and a capacitive sensor to measure the relative humidity (resolution of 0.1 %, relative uncertainty ±2.5 %).

Table 2
Manufacturer data of the PV modules under test at STC.

Installation site	Italy	Brazil	Portugal
PV technology	m-Si	m-Si	p-Si
P_{nom}	410 W	255 W	200 W
η_{STC}	21.0 %	15.8 %	14.2 %
$V_{\text{mpp,STC}}$	31.45 V	30.0 V	26.3 V
$I_{\text{mpp,STC}}$	13.04 A	8.50 A	7.61 A
$V_{\text{oc,STC}}$	37.32 V	37.2 V	32.9 V
$I_{\text{sc,STC}}$	13.95 A	8.85 A	8.21 A
$\alpha_{I_{sc}}$	0.045 %/K	0.042 %/K	0.038 %/K
$\beta_{V_{oc}}$	−0.275 %/K	−0.306 %/K	−0.374 %/K
γ	−0.35 %/K	−0.42 %/K	−0.49 %/K

- One personal computer runs the control software in LabVIEW environment to store the electrical and environmental quantities in a dedicated database.

The measurements were gathered with resolution of 15 min during the period 27 June – 19 September 2023.

4.2. Experimental campaign in Brazil

The analysis is based on the actual PV system located in Brazil, namely Petrolina-ENERQ, which consists of 10 monocrystalline PV modules (rated power 255 W each), with total installed capacity of 2.55 kW. The modules have horizontal inclination (θ_t = 0° with respect to the ground). The measurements are available in [39] with temporal resolution of 5 min for summer days (December 2013) and winter days (June 2014).

4.3. Experimental campaign in Portugal

The analysis is based on the experimental campaign conducted in Porto (Portugal) for one PV module during the months of August and October in 2013. The device under test belongs to GECAD – Knowledge Engineering and Decision-Support Research Group of the Electrical Engineering Institute of Porto. In particular, the polycrystalline silicon PV module has rated power 200 W, rated efficiency about 14 % and horizontal inclination with respect to the ground (θ_t = 0°). The module is equipped with an East-West horizontal single-axis sun tracking system and the data were acquired with a temporal resolution of 5 min.

5. Results

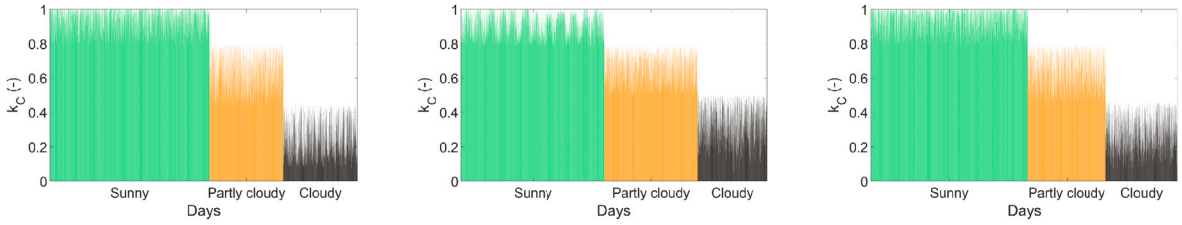
This section presents the results for the application of the two procedures described in Section 2 to the experimental datasets of the three installation sites under study.

In the first step of the methodology, the Plane of Array (PoA) irradiance required by the two procedures was measured in the three locations using three secondary standard pyranometers with the same inclination of the PV modules. In addition, historical series of PoA irradiance and PoA irradiance under clear sky conditions were obtained for the three sites in the period 1 January 2012 - 31 December 2023 from the Solcast database [40], which provides global irradiance and weather data records.

5.1. Classification in weather categories

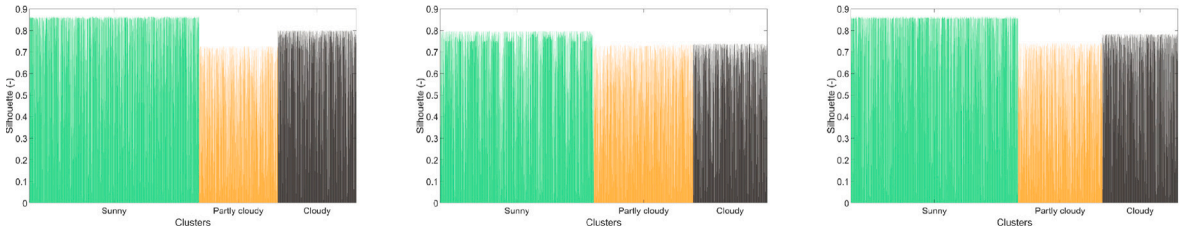
In steps 2A and 3A of the methodology, the clear sky indices were calculated on a daily basis. The clear sky irradiance required for the calculation of the clear sky indices was taken from two different sources. Regarding the installation sites in Italy and Portugal, average profiles of clear sky irradiance for a typical day of each month under study were taken from the PhotoVoltaic Geographical Information System (PVGIS) database [41].

The clear sky irradiance on the tilted surface ($G_{\text{T,clear}}$) for Brazil was calculated using the following modified equation of the ASHRAE Model,



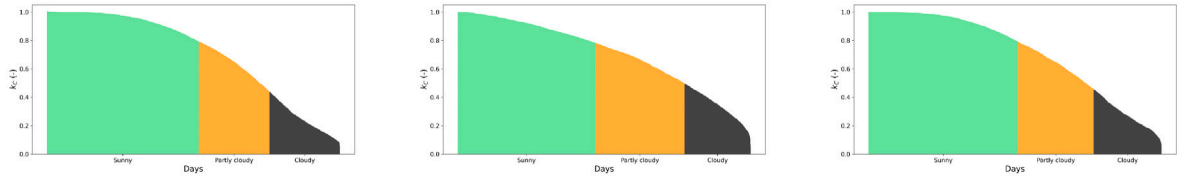
(a) Italy. (b) Brazil. (c) Portugal.

Fig. 3. Clear sky indices for the three installation sites (solution with the highest average silhouette index for each site).



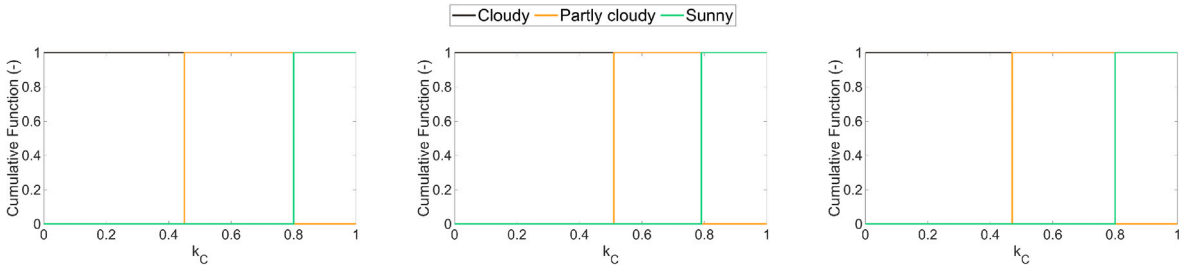
(a) Italy. (b) Brazil. (c) Portugal.

Fig. 4. Silhouette indices for the three installation sites (solution with the highest average silhouette index for each site).



(a) Italy. (b) Brazil. (c) Portugal.

Fig. 5. Clear sky indices in descending order for the three installation sites (solution with the highest average silhouette index for each site).



(a) Italy. (b) Brazil. (c) Portugal.

Fig. 6. Membership degrees from the assessment of the data from the three installation sites.

Table 3
Definition of the clear sky index ranges from 12-year data.

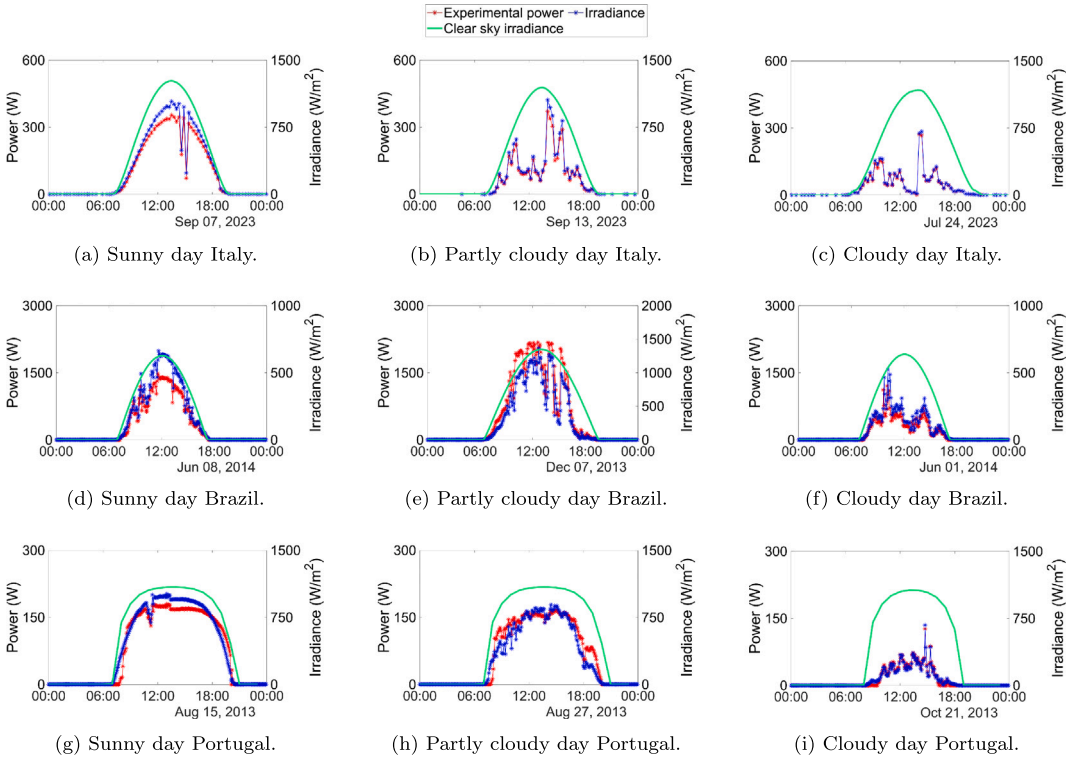
Weather class	Italy	Brazil	Portugal
Cloudy	0 – 0.44	0 – 0.5	0 – 0.46
Partly cloudy	0.44 – 0.79	0.5 – 0.78	0.46 – 0.79
Sunny	0.79 – 1	0.78 – 1	0.79 – 1

which is valid for PV modules with horizontal inclination (θ_t for Brazil $= 0°$):

$$G_{T,clear} = C_{Bra} \cdot A_{Bra} \cdot \cos\theta_z \cdot e^{-\frac{B_{Bra}}{\cos\theta_z}} \quad (22)$$

where A_{Bra} and B_{Bra} (1098 W/m^2 and -0.057, respectively) are the empirical coefficients for Brazil, while C_{Bra} is a seasonality corrective coefficient that is assumed equal to 1.3 and 1.1 for summer and winter days, respectively.

Step 4A of the methodology aims at identifying three ranges for the clear sky indices, corresponding to sunny, partly cloudy, and cloudy days. Therefore, the daily clear sky indices calculated from the irradiance values gathered for 12 years from the Solcast database were grouped into the three clusters for each location using the kmedoids tool (from Matlab, with 100 repetitions for each execution). To obtain a statistically significant set of outcomes, the kmedoids clustering was executed 1000 times for each site. The partitioning into the three groups in the solution with the highest average silhouettes (0.644 for Italy, 0.593



(a) Sunny day Italy. (b) Partly cloudy day Italy. (c) Cloudy day Italy.

(d) Sunny day Brazil. (e) Partly cloudy day Brazil. (f) Cloudy day Brazil.

(g) Sunny day Portugal. (h) Partly cloudy day Portugal. (i) Cloudy day Portugal.

Fig. 7. Experimental power vs. experimental and clear sky irradiance.

for Brazil, and 0.642 for Portugal) is shown in Fig. 3 for the clear sky indices and in Fig. 4 for the local silhouette values. The results show that the local silhouette values are positive for each day and have relatively high values, confirming the effective clustering structure obtained in the three installation sites. In Fig. 3, the indices are sorted in chronological order, while Fig. 5 presents the distribution of the clear sky indices with descending order. Hence, Fig. 5 demonstrates the need to perform clustering to identify the three weather classes since the distribution of the indices in descending order is quite regular without evident gaps among the weather classes.

The results shown in Fig. 6 contain the partitionings obtained in the three installation sites, constructed by taking the CDFs of the higher values for the cloudy days and for the partly cloudy days, as well as the lower values for the partly cloudy days and the sunny days. Interestingly, the separation into the three clusters is the same for all executions in the same installation site.[1] Table 3 summarizes the clear sky index ranges obtained from a statistically significant analysis.

Based on the previous clear sky index ranges, the experimental data have been processed by assigning to the three types of days (sunny, partly cloudy and cloudy) all the days with clear sky index (calculated by integrating the irradiance measured on-site and the clear sky irradiance gathered from the PVGIS or calculated from Eq. (22)) belonging to the corresponding range. Then, the medoids of each type of day have been identified. From the experimental data, each medoid corresponds to a PV power pattern. Fig. 7 shows the experimental PV power (red points) and PoA irradiance (blue points) for the three medoids for each installation site. The green curve is the estimated clear sky irradiance. The patterns shown in the figure belong to different periods of the year, so that the sunny day does not necessarily belong to the season with high daily irradiation.

[1] By using data for a shorter period, e.g., daily values for one year or less, the separation was not the same in all cases, having higher dispersion of the data; this is a reason for choosing a multi-year set of data.

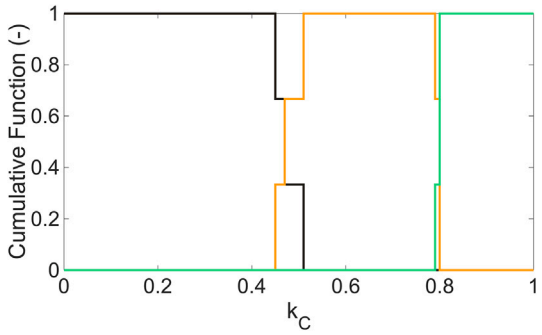


Fig. 8. Membership degrees from the global assessment of the data from the three installation sites.

A further result is shown in Fig. 8, in which all the clear sky indices of the three sites are considered together. It is interesting to note that the separation between partly cloudy and sunny days is almost similar in all cases, notwithstanding the clear differences in latitude, technology and type of installation (fixed vs. sun-tracking). The similarity is lower for the separation from cloudy to partly cloudy days. However, there are well-defined ranges of the clear sky index in which the type of day is consistently established with membership function equal to unity.

5.2. Comparison among the energy conversion models

In step 6A of the methodology, the energy conversion models defined in Section 3.4 are applied to the classified datasets (sunny, partly cloudy and cloudy days). This operation has the goal of evaluating the effectiveness of the energy models in different weather conditions. The results for the indicators defined in Section 3.5 are reported in Tables 4–6. For each weather class of the different sites, the best indicators are presented in bold. Among the three considered statistical indicators,

Table 4
Results of the R^2 indicator.

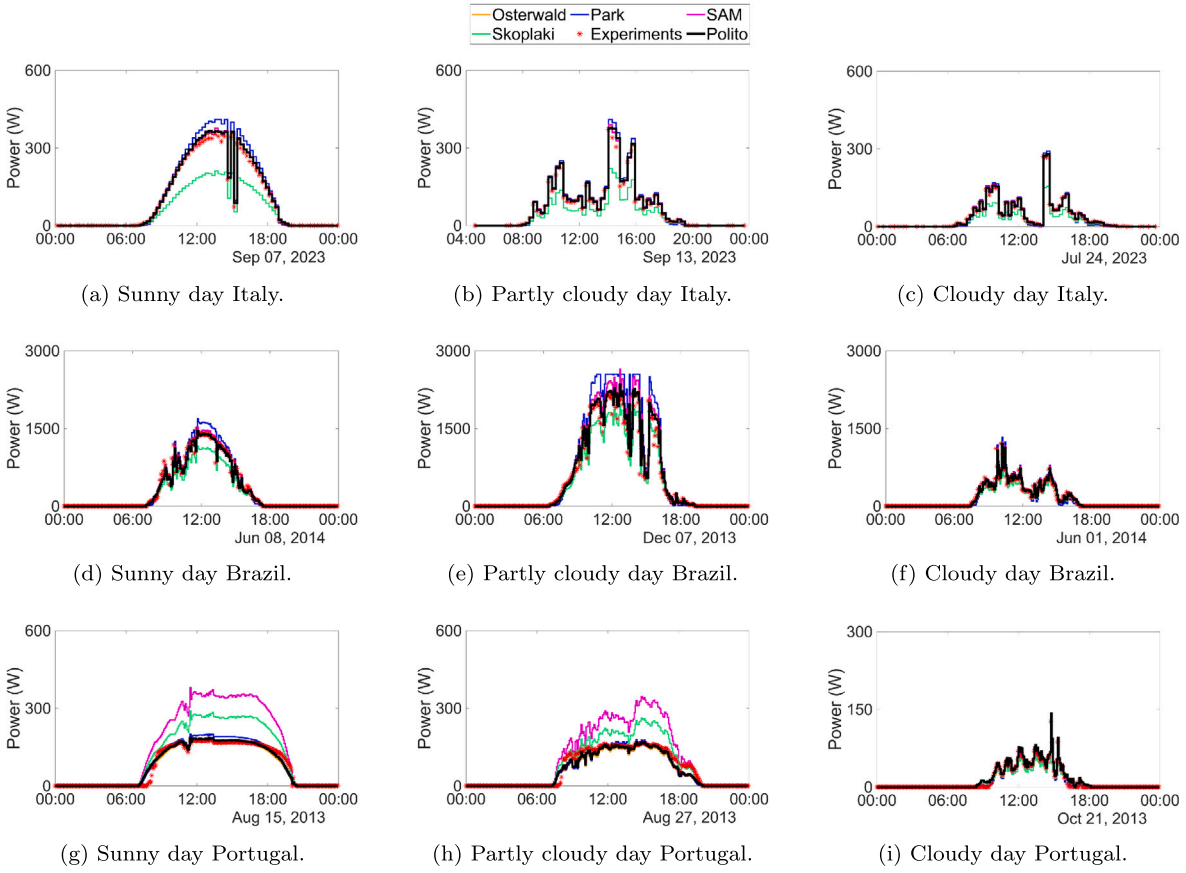
	Italy			Brazil			Portugal		
	Sunny	Partly cloudy	Cloudy	Sunny	Partly cloudy	Cloudy	Sunny	Partly cloudy	Cloudy
Polito model	0.977	**0.995**	0.929	0.984	**0.898**	**0.829**	**0.948**	**0.974**	**0.962**
Osterwald	**0.980**	**0.995**	**0.933**	**0.986**	0.873	0.811	**0.948**	0.968	0.942
Park	0.943	0.964	0.906	0.970	0.769	0.713	0.931	0.962	0.954
SAM	0.979	**0.995**	0.932	**0.986**	0.865	0.805	0.162	0.141	0.764
Skoplaki	0.701	0.709	0.741	0.900	0.886	0.819	0.728	0.747	0.914

Table 5
Results of the $NRMSE$ indicator.

	Italy			Brazil			Portugal		
	Sunny	Partly cloudy	Cloudy	Sunny	Partly cloudy	Cloudy	Sunny	Partly cloudy	Cloudy
Polito model	12.1 %	**8.1 %**	**16.2 %**	18.6 %	**42.0 %**	**53.5 %**	**23.4 %**	**18.3 %**	**23.9 %**
Osterwald	**10.7 %**	8.3 %	17.5 %	**17.5 %**	46.9 %	56.2 %	23.8 %	20.3 %	26.1 %
Park	25.3 %	22.7 %	29.0 %	25.6 %	63.4 %	69.0 %	27.3 %	22.7 %	24.6 %
SAM	11.5 %	8.83 %	17.7 %	**17.5 %**	48.3 %	57.0 %	86.7 %	96.5 %	42.9 %
Skoplaki	64.1 %	65.4 %	74.9 %	46.5 %	46.0 %	55.3 %	53.6 %	56.3 %	39.7 %

Table 6
Results of the MBE indicator.

	Italy			Brazil			Portugal		
	Sunny	Partly cloudy	Cloudy	Sunny	Partly cloudy	Cloudy	Sunny	Partly cloudy	Cloudy
Polito model	**4.33 %**	**2.74 %**	2.07 %	−4.91 %	**4.62 %**	**3.07 %**	2.38 %	**−0.03 %**	**−2.12 %**
Osterwald	5.69 %	5.13 %	6.03 %	−1.36 %	11.0 %	7.99 %	**−0.68 %**	−4.33 %	−10.7 %
Park	13.4 %	11.9 %	**1.27 %**	3.42 %	22.1 %	11.1 %	7.75 %	1.69 %	−10.8 %
SAM	6.25 %	5.56 %	6.24 %	**−0.91 %**	12.0 %	8.51 %	50.1 %	56.4 %	14.1 %
Skoplaki	−40.2 %	−40.5 %	−39.6 %	−25.1 %	−14.9 %	−17.8 %	18.9 %	22.5 %	−7.57 %



(a) Sunny day Italy. (b) Partly cloudy day Italy. (c) Cloudy day Italy.

(d) Sunny day Brazil. (e) Partly cloudy day Brazil. (f) Cloudy day Brazil.

(g) Sunny day Portugal. (h) Partly cloudy day Portugal. (i) Cloudy day Portugal.

Fig. 9. Power prediction by the energy models vs. experimental data.

the MBE better represents the accuracy of the energy assessed by the models, as it calculates the PV power deviations without considering the absolute value:

- In cloudy conditions, the Polito model is better in the Brazilian and Portuguese sites and the Park model performs better with the data from the Italian site.
- In partly cloudy conditions, the Polito model has shown better performance than the other models in all the sites tested.
- In sunny conditions, the Polito model performs better with the data from the Italian site, while SAM performs better in the Brazilian site and Osterwald in the Portuguese site.

In partly cloudy conditions, the better performance of the Polito model is confirmed by the highest values of R^2 and the lowest values of $NRMSE$ in all sites. In addition, for the Polito model the $NRMSE$ is always better than the other models for all sites in cloudy conditions. In summary, the Polito model is suggested for partly cloudy days and cloudy days at all sites.

In sunny conditions, regarding the $NRMSE$, the best performance is achieved by different models. Actually, the Polito model has exhibited better performance for the Portuguese site, while Osterwald and SAM have obtained better values for the Brazilian site (lowest $NRMSE$ for both) and the Italian site (Osterwald). These results are confirmed by the R^2: the Polito model has the highest value in Portugal, while Osterwald and SAM achieve the best performance in the other sites. As a conclusion, it is not possible to suggest one model for all the locations tested in sunny conditions.

Fig. 9 presents the experimental power and the estimated power by the different energy models for the three medoids, i.e., the most representative sunny, partly cloudy, and cloudy days for each installation site. The figure confirms that the power estimation by the Polito model (black curve) is the closest to the experimental data (red points) for partly cloudy and cloudy conditions. In sunny conditions, the best performance is obtained by the Polito model, while SAM (purple curve) has shown the best PV power estimation for Brazil and Osterwald (orange curve) for Portugal.

In step 6B of the methodology, the energy models presented in Section 3 were applied to the global datasets available for Italy, Brazil, and Portugal. The results for the statistical indicators described in Section 3.5 are shown in Table 7. For each site, the best model is highlighted in bold. The Polito model has achieved the best performance for the Brazilian and the Portuguese sites, with the highest R^2 and the lowest $NRMSE$ and MBE. In Italy, despite SAM has shown better values of R^2 and $NRMSEs$, the MBE by the Polito model is the lowest. Therefore, considering the overall results, the Polito model is suggested for all the three installation sites.

Table 7
Indicators for power estimation.

Installation site	Italy	Brazil	Portugal
Index R^2			
Polito model	0.981	**0.913**	**0.968**
Osterwald	0.979	0.894	0.889
Park	0.948	0.814	0.957
SAM	**0.983**	0.890	0.302
Skoplaki	0.704	0.901	0.784
Index $NRMSE$			
Polito model	16.3 %	**46.1 %**	**23.8 %**
Osterwald	17.0 %	50.9 %	44.3 %
Park	27.0 %	67.4 %	27.5 %
SAM	**16.0 %**	52.2 %	111 %
Skoplaki	64.6 %	49.3 %	61.9 %
Index MBE			
Polito model	**4.56 %**	**2.68 %**	**1.35 %**
Osterwald	5.60 %	8.44 %	14.9 %
Park	12.0 %	16.6 %	2.54 %
SAM	6.00 %	9.31 %	47.0 %
Skoplaki	−40.2 %	−9.04 %	16.4 %

6. Concluding remarks

The methodology proposed in this paper has enabled reaching some significant results. Starting from the inappropriateness of the clearness index to represent PV systems with sun-tracking structures, the clear sky index has emerged as an appropriate indicator for creating consistent groups of sunny, partly cloudy and cloudy days considering PV sites at different latitudes and technologies, and with fixed or sun-tracking configurations. The results obtained from kmedoids clustering executed for a statistically significant number of times have identified similar ranges of the clear sky index that represent sunny days in different conditions, while the ranges for the distinction between cloudy and partly cloudy days are slightly different in the sites analyzed. This part of the methodology can be used in a general way to identify the ranges of the types of days, allowing a sound classification that can be considered for assigning the days based on simulations, predictions or experimental assessments. The day type indications may be used in future works in this research line to classify days according to weather conditions in experimental campaigns for sites with latitudes close to that of the locations investigated in this study. Furthermore, the evaluation of the effectiveness of the PV energy conversion models in different weather conditions has shown the remarkable performance of the Polito model for the Italian and Brazilian sites in any weather condition, as well as in the Portuguese site campaign under partly cloudy conditions. In the comparison of the performance of the PV energy conversion models for the global datasets, the Polito model was the most accurate for energy assessment, with mean bias errors lower than 5 % for all the three sites. The Polito model has been formulated after numerically determining the reference parameters of the equivalent circuit and the coefficients of equations to quantify their weather dependency from experimental datasets of modules with different cell technology, efficiency and manufacturing year. However, newer technologies like the Heterojunction (HJT), Tunnel Oxide Passivated Contact (TOPCon), and All Back Contact (ABC) modules are arising in the global market and, currently, the effectiveness of the Polito model, as well as that of other models in the literature, could be lower while analyzing PV modules with such technologies. Hence, in future works, the Polito model will be extended to be applicable also for these PV technologies. Moreover, the Polito model has the possibility of being extended to estimate the maximum power in the case of partial shading. Since the parameters and coefficients of the Polito model result from numerical optimizations applied to experimental studies, other sets for different operating conditions, like at partial shading, will be identified by focusing on datasets affected by that condition. The accurate estimation of the maximum power under partial shading will be achieved thanks to the possibility of obtaining the whole *I-V* curve by the Polito model, avoiding the need to select local maxima in the case of multiple peaks.

CRediT authorship contribution statement

Gabriele Malgaroli: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Carmen Lucia Tancredo Borges:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Filippo Spertino:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Alessandro Ciocia:** Writing – review & editing, Formal analysis, Data curation, Conceptualization. **David Alexander:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Gianfranco Chicco:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization.

Data availability

No - I have not included a data availability in my manuscript

References

[1] International Renewable Energy Agency (IRENA), World Energy Transitions Outlook 2023: 1.5°c Pathway, Tech. rep., IRENA, Abu Dhabi, 2023, https://www.irena.org/Publications/2023/June/World-Energy-Transitions-Outlook-2023.

[2] P. Hacke, S. Lokanath, P. Williams, A. Vasan, P. Sochor, G.S. TamizhMani, H. Shinohara, S. Kurtz, A status review of photovoltaic power conversion equipment reliability, safety, and quality assurance protocols, Renew. Sustain. Energy Rev. 82 (2018) 1097–1112, https://doi.org/10.1016/J.RSER.2017.07.043

[3] G. Di Lorenzo, R. Araneo, M. Mitolo, A. Niccolai, F. Grimaccia, Review of o&m practices in PV plants: failures, solutions, remote control, and monitoring tools, IEEE J. Photovoltaics 10 (2020) 914–926, https://doi.org/10.1109/JPHOTOV.2020.2994531

[4] J.E. Ribeiro Baptista, A. Barbosa Rodrigues, M. Da Guia Da Silva, Probabilistic analysis of PV generation impacts on voltage sags in LV distribution networks considering failure rates dependent on feeder loading, IEEE Trans. Sustain. Energy 10 (2019) 1342–1350, https://doi.org/10.1109/TSTE.2018.2866931

[5] C.O. Inacio, C.L. Tancredo Borges, Stochastic model for generation of high-resolution irradiance data and estimation of power output of photovoltaic plants, IEEE Trans. Sustain. Energy 9 (2018) 952–960, https://doi.org/10.1109/TSTE.2017.2767780

[6] B. Yang, T. Zhu, P. Cao, Z. Guo, C. Zeng, D. Li, Y. Chen, H. Ye, R. Shao, H. Shu, T. Yu, Classification and summarization of solar irradiance and power forecasting methods: a thorough review, CSEE J. Power Energy Syst. 9 (2023) 978–995, https://doi.org/10.17775/CSEEJPES.2020.04930

[7] N. Zhou, X. Xu, Z. Yan, M. Shahidehpour, Spatio-temporal probabilistic forecasting of photovoltaic power based on monotone broad learning system and copula theory, IEEE Trans. Sustain. Energy 13 (2022) 1874–1885, https://doi.org/10.1109/TSTE.2022.3174012

[8] R. Oliveira, C.L. Tancredo Borges, A comparison of photovoltaic models for estimating power generation: a case study of Brazilian data, Clean Technol. Environ. Policy 23 (2020) 517–530, https://doi.org/10.1007/S10098-020-01986-W

[9] R. Oliveira, C.L. Tancredo Borges, Influence of photovoltaic generation model and time resolution on the reliability evaluation of distribution systems, Int. J. Energy Res. 45 (2021) 864–878, https://doi.org/10.1002/er.5971

[10] M.J. Mayer, G. Gróf, Extensive comparison of physical models for photovoltaic power forecasting, Appl. Energy 283 (2021) 116239, https://doi.org/10.1016/j.apenergy.2020.116239

[11] A. Ciocia, P. Di Leo, S. Fichera, F. Giordano, G. Malgaroli, F. Spertino, A novel procedure to adjust the equivalent circuit parameters of photovoltaic modules under shading, in: Proc. 2020 Int. Symp. On Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM), 2020, pp. 711–715, https://doi.org/10.1109/SPEEDAM48782.2020.9161878

[12] Á. Fernández-Solas, J. Montes-Romero, L. Micheli, F. Almonacid, E.F. Fernández, Estimation of soiling losses in photovoltaic modules of different technologies through analytical methods, Energy 244 (2022) 123173, https://doi.org/10.1016/J.ENERGY.2022.123173

[13] M. Fuentes, G. Nofuentes, J. Aguilera, D.L. Talavera, M. Castro, Application and validation of algebraic methods to predict the behaviour of crystalline silicon PV modules in Mediterranean climates, Sol. Energy 81 (2007) 1396–1408, https://doi.org/10.1016/j.solener.2006.12.008

[14] A. Ciocia, G. Chicco, F. Spertino, An improved model for AC power from grid connected photovoltaic systems and comparison with large-scale hourly measured data, IEEE Trans. Ind. Appl. 60 (2024) 4458–4469, https://doi.org/10.1109/TIA.2024.3359121

[15] D.D. Milosavljević, T.S. Kevkić, S.J. Jovanović, Review and validation of photovoltaic solar simulation tools/software based on case study, Open Phys. 20 (2022) 431–451, https://doi.org/10.1515/PHYS-2022-0042

[16] A.M. Humada, M. Hojabri, S. Mekhilef, H.M. Hamada, Solar cell parameters extraction based on single and double-diode models: a review, Renew. Sustain. Energy Rev. 56 (2016) 494–509, https://doi.org/10.1016/J.RSER.2015.11.051

[17] C.S. Lai, Y. Jia, M.D. McCulloch, Z. Xu, Daily clearness index profiles cluster analysis for photovoltaic system, IEEE Trans. Ind. Inf. 13 (2017) 2322–2332, https://doi.org/10.1109/TII.2017.2683519

[18] N.A. Engerer, F.P. Mills, Kpv: a clear-sky index for photovoltaics, Sol. Energy 105 (2014) 679–693, https://doi.org/10.1016/j.solener.2014.04.019

[19] F. Spertino, G. Malgaroli, A. Amato, M.A.E. Qureshi, A. Ciocia, H. Siddiqi, An innovative technique for energy assessment of a highly efficient photovoltaic module, Solar 2 (2022) 321–333, https://doi.org/10.3390/SOLAR2020018

[20] 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

[21] M.J. Ahmad, G.N. Tiwari, Solar radiation models-a review, Int. J. Energy Res. 35 (2011) 271–290, https://doi.org/10.1002/er.1690

[22] S. Roy, Worst-case photovoltaic generation and power change distribution under dense cloud cover, IEEE Trans. Sustain. Energy 8 (2017) 1021–1028, https://doi.org/10.1109/TSTE.2016.2645383

[23] T. Zhu, H. Wei, X. Zhao, C. Zhang, K. Zhang, Clear-sky model for wavelet forecast of direct normal irradiance, Renew. Energy 104 (2017) 1–8, https://doi.org/10.1016/J.RENENE.2016.11.058

[24] A. Mellit, S.A. Kalogirou, S. Shaari, H. Salhi, A.H. Arab, Methodology for predicting sequences of mean monthly clearness index and daily solar radiation data in remote areas: application for sizing a stand-alone PV system, Renew. Energy 33 (2008) 1570–1590, https://doi.org/10.1016/J.RENENE.2007.08.006

[25] M. Piliougine, D. Elizondo, L. Mora-López, M. Sidrach-De-Cardona, Modelling photovoltaic modules with neural networks using angle of incidence and clearness index, Prog. Photovoltaics Res. Appl. 23 (4) (2015) 513–523, https://doi.org/10.1002/pip.2449

[26] L. Kaufman, P.J. Rousseeuw, Partitioning around medoids (program pam), in: Finding Groups in Data: an Introduction to Cluster Analysis, John Wiley & Sons, 1990, pp. 68–125, https://doi.org/10.1002/9780470316801.ch2

[27] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65, https://doi.org/10.1016/0377-0427(87)90125-7

[28] C.R. Osterwald, Translation of device performance measurements to reference conditions, Sol. Cells 18 (1986) 269–279, https://doi.org/10.1016/0379-6787(86)90126-2

[29] J. Park, W. Liang, J. Choi, A.A. El-Keib, M. Shahidehpour, R. Billinton, A probabilistic reliability evaluation of a power system including solar/photovoltaic cell generator, in: Proc. IEEE Power and Energy Society General Meeting (PES '09), 2009, pp. 1–8, https://doi.org/10.1109/PES.2009.5275722

[30] D.L. King, W.E. Boyson, J.A. Kratochvil, Photovoltaic Array Performance Model, Tech. Rep. SAND2004-3535, Sandia National Laboratories, Albuquerque, NM, 2004, https://doi.org/10.2172/919131. https://www.osti.gov/biblio/919131.

[31] E. Skoplaki, A.G. Boudouvis, J.A. Palyvos, A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting, Sol. Energy Mater. Sol. Cells 92 (2008) 1393–1402, https://doi.org/10.1016/j.solmat.2008.05.016

[32] A.K. Tossa, Y.M. Soro, Y. Azoumah, D. Yamegueu, A new approach to estimate the performance and energy productivity of photovoltaic modules in real operating conditions, Sol. Energy 110 (2014) 543–560, https://doi.org/10.1016/J.SOLENER.2014.09.043

[33] D.M. Fébba, R.M. Rubinger, A.F. Oliveira, E.C. Bortoni, Impacts of temperature and irradiance on polycrystalline silicon solar cells parameters, Sol. Energy 174 (2018) 628–639, https://doi.org/10.1016/J.SOLENER.2018.09.051

[34] C.S. Ruschel, F.P. Gasparin, A. Krenzinger, Experimental analysis of the single diode model parameters dependence on irradiance and temperature, Sol. Energy 217 (2021) 134–144, https://doi.org/10.1016/j.solener.2021.01.067

[35] W. De Soto, S.A. Klein, W.A. Beckman, Improvement and validation of a model for photovoltaic array performance, Sol. Energy 80 (1) (2006) 78–88, https://doi.org/10.1016/j.solener.2005.06.010, https://www.sciencedirect.com/science/article/pii/S0038092X05002410.

[36] T. Ma, H. Yang, L. Lu, Solar photovoltaic system modeling and performance prediction, Renew. Sustain. Energy Rev. 36 (2014) 304–315, https://doi.org/10.1016/J.RSER.2014.04.057

[37] E. Ogliari, M. Sakwa, P. Cusa, Enhanced convolutional neural network for solar radiation nowcasting: all-sky camera infrared images embedded with exogeneous parameters, Renew. Energy 221 (2024) 119735, https://doi.org/10.1016/j.renene.2023.119735

[38] J.Y. Han, P. Vohnicky, An optimized approach for mapping solar irradiance in a mid-low latitude region based on a site-adaptation technique using Himawari-8 satellite imageries, Renew. Energy 187 (2022) 603–617, https://doi.org/10.1016/j.renene.2022.01.027

[39] Open Data Sets, Dataset available at, 2025. https://site.ieee.org/pes-iss/data-sets/. (Accessed 20 March 2025).

[40] Historical Solar Irradiance Time Series, Dataset available at, 2025. https://solcast.com/time-series. (Accessed 20 March 2025).

[41] JRC Photovoltaic Geographical Information System (PVGIS), Online resource available at. 2025. https://re.jrc.ec.europa.eu/pvg_tools/en/. (Accessed 20 March 2025).