POLITECNICO DI TORINO
Repository ISTITUZIONALE

An Improved Model for AC Power from Grid Connected Photovoltaic Systems and Comparison with Large-Scale Hourly Measured Data

Original
An Improved Model for AC Power from Grid Connected Photovoltaic Systems and Comparison with Large-Scale Hourly Measured Data / Ciocia, A., Chicco, G., Spertino, F.. - In: IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS. - ISSN 0093-9994. - ELETTRONICO. - (2024), pp. -10. [10.1109/TIA.2024.3359121]

Availability:
This version is available at: 11583/2986107 since: 2024-02-19T16:35:44Z

Publisher:
ieee

Published
DOI:10.1109/TIA.2024.3359121

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

28 November 2026

This article has been accepted for publication in IEEE Transactions on Industry Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIA.2024.3359121

An Improved Model for AC Power from Grid Connected Photovoltaic Systems and Comparison with Large-Scale Hourly Measured Data

Alessandro Ciocia, *Member, IEEE,* Gianfranco Chicco, *Fellow, IEEE,* and Filippo Spertino, *Senior Member, IEEE*

***Abstract*—The diffusion of renewable energy sources (RES) can lead to cost reductions for the ancillary services provided to the transmission system operator, in particular, when the RES production can be estimated with appropriate accuracy. This article presents significant improvements to photovoltaic (PV) power conversion models found in the literature for grid-connected PV systems. The refinement of the conversion model is based on public hourly data of irradiance and ambient temperature, referred to the sites of three groups of PV plants, to calculate the hourly average power injected into the grid. The values obtained are compared with the hourly average power measured by the distribution system operator meters on the PV systems. A double-step optimization procedure, based on seasonal analysis, sets up the various parameters in the PV conversion model. The key result is that the deviation between the simulated and measured annual energy is reduced to less than 2%. Moreover, with the proposed model the monthly energy deviations are limited to a few percentage points. This allows for significant improvements in the estimation of the production of grid-connected PV systems.**

***Index Terms*—photovoltaic systems, modelling, optimization, stratified sampling.**

I. Introduction

THE generation from renewable energy sources (RES) such as photovoltaic (PV) systems is highly variable, challenging the operation of the power systems to which RES are connected, up to the possible need for RES curtailment [1]. The PV power generated mainly depends on the irradiance, which can be predicted with good accuracy only during clear-sky days. However, passing from irradiance to PV power is not an easy task, due to many aspects that affect the irradiance-PV power production chain (e.g., losses due to partial shading and imperfect maximum power point tracking).

At the large-scale level, such as at the interface with the transmission and sub-transmission networks, the characterization of the PV production is crucial for setting up power system control. The goal of the control is to ensure that generation and demand are balanced during time, and that adequate resources are available to provide grid services [2].

On an even larger scale, e.g., for a wide territorial area such as a region or a country, the assessment of PV productivity is essential for planning purposes. Some results are relevant for the electricity market, to understand the expected contribution of PV production at different hours, including the integration with storage systems [3]. Further results refer to procuring the resources for ancillary services (e.g., to mitigate power fluctuations from non-controllable RES [4] or provide frequency support functions [5], if needed with PV power curtailment [6], and maintain available headroom for fast reserves [7]). The assessment of PV productivity ensures better economic dispatch carried out by the Transmission System Operator (TSO), based on the provisions indicated in the grid codes [8]. If the sum of the rated powers of the PV systems located in a given area is known, in the absence of meteorological information, the PV power in the limit case could be estimated based on clear-sky conditions. In real cases, the PV production could be similar to (at clear sky) or lower than (in cloudy conditions) the estimated one, and the possible missing production must be procured during time by means of ancillary services. If the PV models can be estimated better, there is less necessity to procure ancillary services with respect to the estimations based on clear sky data. On the other hand, in the case of global overgeneration with respect to consumption, the TSO may adopt, to ensure the stability of frequency, the power curtailment for the intermittent generation as for PV power. The modeling of power curtailment is not a difficult task, as will be explained in the following, provided that the relevant information is known. In this research work, the measured AC power data from the energy meters, installed and managed by the local Distribution System Operators (DSOs), are used as the benchmark for setting up the model of AC power for grid-connected PV systems.

The photovoltaic power models can be categorized into two groups. The first group (from now "DC power models") directly gives the DC power in the maximum power point without the knowledge of the current-voltage (I-V) curve. These models use an equation in which the generated power is proportional to irradiance, including or not a worsening factor depending on the increment in the PV cells temperature, with respect to the reference temperature (25°C). The first model (PESCR, [9]) is a basic model that simply considers the influence of irradiance, without the effect of the temperature. The second model [10] includes the calculation of low irradiance losses, but it does not yet consider the temperature effect (thermal losses). The third model [11] incorporates the proportionality with irradiance and the temperature losses, but the low irradiance losses are not considered.

The second group (from now, "diode-based models") includes equivalent circuit models, which are based on the equivalent circuit of the PV cells with a single or more diodes [12][13]. These models permit to calculate, the entire I-V curve, at each operating condition, including the maximum power point. The authors in [14] present a comprehensive review of the diode-based models: their accuracy is influenced by both the quantity of model parameters and the method used for parameters' estimation. The main issue of the diode-based models is that they require the values of the equivalent circuit parameters, that can be obtained only by I-V curve measurements. Moreover, the implementation of the diode-

The authors are with Dipartimento Energia "Galileo Ferraris", Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy, (e-mail: {alessandro.ciocia, gianfranco.chicco, filippo.spertino}@polito.it)

Authorized licensed use limited to: Politecnico di Torino. Downloaded on February 19,2024 at 16:23:56 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Industry Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIA.2024.3359121

TABLE I. COMPARISON OF PHOTOVOLTAIC POWER MODELS

	Thermal losses	Low irradiance losses	Effect of wind speed on PV cell temperature	Effect of irradiance and air temperature on PV cell temperature	Independent of I-V curve measurements	Independent of measurements of PV cell temperature
[9]					X	
[10]		X			X	
[11]+[15]	X			X	X	X
[11]+[17]	X		X	X	X	
[11]+[19]	X		X	X	X	
[11]+[18]	X		X	X	X	
[12]+[15]	X	X		X		X
[12]+[17]	X	X	X	X		
[12]+[19]	X	X	X	X		
[12]+[18]	X	X	X	X		
[21]+[19]	X	X	X	X	X	
[21]+[18]	X	X	X	X	X	
[21]+[15] *	X	X		X	X	X
[21]+[17] *	X	X	X	X	X	

*Models considered and improved in the present work

based model requires a much higher computational effort with respect to the DC power models, because the I-V datapoints are results of numerical methods to solve transcendental equations.

Regarding the thermal losses, the previous models, excluding the PESCR and Park models, require a calculation to determine the cell temperature, and so the worsening factor. The Normal Operating Cell Temperature (NOCT) model [15] requires the value of NOCT parameter from PV modules datasheet; manufacturers must provide this parameter according to [16]. This NOCT value could be corrected according to the installation condition (higher values must be considered in the case of PV modules with a worse heat dissipation, typical of building integrated PV generators. The main limitation of this simple model is that the effect of wind speed is not considered. Other cell temperature models [17][18][19] are based on semiempirical formulas, which could be very accurate in the temperature calculation, but the drawback is that they require many measurements for the determination of the parameters. In the model in [17] (from now called WIND model), the PV cell temperature linearly depends on irradiance, air temperature and wind speed. The same weather inputs are used in [18], which proposes a nonlinear calculation of the cell temperature. In [19], the temperature calculation is based on a simplified energy balance. In every model, if the external conditions are changed, the empirical parameters have to be recalculated by fitting new measurement campaigns.

Starting from the model in [11], which is a good compromise between accuracy and simplicity [20] (not requiring the I-V curve measurement), in the present work, an improved version of [11] is adopted in a previous work of the authors [21], in which the losses due to low irradiance levels are taken into account. With this first improvement, the energy results are close to the results of the diode-based models [22]. Regarding the cell temperature calculation, [15] and [17] are considered in this paper; NOCT is simple and does not require measurements, while WIND takes into account also the effect of wind speed and requests temperature measurements. Among the temperature models which require measurements, model from [17] is simple, but it has the same performance of the other models [21]. A summary of the comparison of the PV power models is shown in Table I.

A mandatory step to obtain the AC power is the assessment of the inverter efficiency, including the MPPT (Maximum Power Point Tracker) effectiveness. The possible models for the conversion losses are polynomial functions [23], in which the second degree is usually the compromise solution [24].

This paper addresses large-scale PV system integration by providing accurate modelling of power conversion in PV plants. The effect of this accurate modelling is to enhance the estimation of the overall PV plant production over a relatively long-term period (e.g., one year). For this purpose, specific improvements to the existing PV power conversion model are introduced, based on the comparison between simulated results and measured data, with a large-scale validation carried out on several PV power plants with different sizes and locations.

The initial concepts were presented in the conference paper [21], of which this paper is the extended version. In [21] a comparison of different models for the calculation of PV module temperature has been presented, with a validation performed considering the data of the PV plants in an Italian region. This paper extends the analysis by introducing three main levels of novelty:

1. The improvement of the PV conversion model by changing not only the temperature parameters, but also the irradiance-PV conversion efficiency. Then, a new quadratic formula is introduced to better compute AC power during cloudy-sky days, and a global coefficient is used to include all other sources of losses not considered in the literature model.
2. The use of a double-step procedure to enable the optimal determination of the model parameters. The excellent performance of the proposed double-step optimization is shown by the comparison with the single-step optimization in which the choice of the parameters is obtained throughout the entire year, without distinguishing the seasons.
3. The application of the proposed procedure to a bigger dataset of PV systems, i.e., three groups of PV systems, respectively in Northern, Central, and Southern Italy.

The next sections of this paper are organized as follows. Section II describes how to elaborate the weather data inputs and the model for PV power generation. Section III presents the

Authorized licensed use limited to: Politecnico di Torino. Downloaded on February 19,2024 at 16:23:56 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Industry Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIA.2024.3359121

proposed improvements to better match the actual PV production patterns. Section IV describes the overall data analysis and model correction procedure, including data filtering, statistical analysis, optimal determination, and definition of the energy deviation parameters. Section V starts with a description of the PV plant groups under analysis and shows the improvements in the calculation of the AC power patterns of PV generation thanks to the proposed procedure. Section VI contains the conclusions.

II. Modelling of the PV Power Injected into the Grid

In the literature, the power conversion models for PV plants are generally created to calculate the energy output of well-working PV systems that operate in clear-sky days. Thus, the higher deviations between the real PV power patterns and the simulations of AC power occur during cloudy-sky days, especially in winter [25]. Moreover, the models do not consider additional losses, such as the losses referring to incorrect installations, low quality components, poor maintenance, etc., or include more details but refer to a single small PV plant [26].

The details of the PV conversion model are illustrated in this section, with the following organization. Subsections II.A and II.B present the elaboration of the weather data inputs for the PV production model, i.e., irradiance on the horizontal plane and air temperature, respectively. For the calculation of the PV module temperature starting from the air temperature, two different methods are presented, and the results obtained with these models are compared in the analysis of the optimization results. Subsection II.C describes in detail one of the most used models for calculating the PV production, identifying the possible improvements needed.

A. Plane of Array Irradiance from the ASHRAE Model

The irradiance data required in the PV production model refer to the global irradiance G (W/m^2) that reaches the inclined plane of the PV modules (Plane of Array Irradiance, *PAI*). In many cases, the databases (e.g., [27]) contain only irradiance data measured on the horizontal plane; these patterns shall be corrected to represent the inclined-plane irradiance. The *PAI* is obtained from the ASHRAE model [28]:

$$G = BHI\cdot\cos(\theta)/\cos(\theta_z) + DHI\cdot F_{CS} + \rho\cdot GHI\cdot(1-F_{CS}) \quad (1)$$

The physical quantities in (1) are defined as follows:

- *BHI* is the direct irradiance. It is the component of irradiance that reaches the horizontal plane without having undergone reflections in the atmosphere.
- *DHI* is the diffuse irradiance. It is the irradiance component that reaches the horizontal plane following reflections and absorption in the atmosphere.
- *GHI* is the global irradiance. It consists of the sum of the various components of irradiance, including the reflected (albedo component) irradiance G_r:

$$GHI = BHI + DHI + G_r \quad (2)$$

- θ_z is the zenith angle. It is the angle between the Earth-Sun line and the zenith direction: $\theta_z = 0°$ only in tropical sites.
- $cos(\theta)$ represents the incidence of the sunlight, where θ is the angle between the direction normal to the inclined plane and the Earth-Sun line: $\theta = 0°$ if beam is normal to the plane.
- F_{CS} is the plane-to-sky view factor.
- ρ is the albedo coefficient. It considers the reflection of the surroundings.

It is noted that the *ASHRAE* model refers to clear sky conditions, while actual irradiance components will be used in the proposed procedure.

B. Module Temperature Models from the Literature

For the calculation of the temperature T of the PV modules, the model in [15] is based on the Normal Operating Cell Temperature (*NOCT*) (in the sequel, "*NOCT model*"). The environmental data used as inputs for the cell temperature T are the air temperature T_a and the irradiance G:

$$T = T_a + (NOCT - T_{a,NOCT})\cdot G/G_{NOCT} \quad (3)$$

NOCT is the experimental temperature indicated by the manufacturer in the module datasheet. It is measured in steady state conditions, i.e., $T_{a,NOCT}$ = 20 °C, G_{NOCT} = 800 W/m^2 with module at 45° inclination in stable open circuit conditions and wind speed of 1 m/s. As a result, the *NOCT* of commercial crystalline silicon (c-Si) modules is between 42÷50°C [29].

Another possible solution is given by a formula based on a measurement campaign of weather data and PV modules temperature [17]. In this model (in the sequel, "*wind model*"), the module temperature is calculated as a function of air temperature T_a (in °C), irradiance G, and wind speed w_s (m/s):

$$T = a_T\cdot T_a + b_T\cdot G - c_T\cdot w_s + d_T \quad (4)$$

where the parameter a_T = 0.943 is dimensionless, b_T is equal to 0.028 (°C·m^2/W), c_T = 1.528 (°C·s/m), and d_T = 4.3 °C.

C. PV Power Conversion Model from the Literature

The paper [21] presents a simple equation to compute the active power P_{AC} injected into the grid by a generic PV system. This equation is useful to estimate the production of a well-working PV system. In this case, it is calculated as the product of DC power P_{DC} (supposing a MPPT stage) times the efficiency of the DC/AC converter. P_{DC} can be computed by a model proportional to the irradiance G and dependent on the temperature of the PV modules [17]:

$$P_{AC} = P_{DC}\cdot\eta_{conv} = G\cdot A\cdot\eta_{STC}\cdot\eta_{low,G}\cdot C_{Th}\cdot\eta_{mix}\cdot\eta_{l}\cdot\eta_{conv} \quad (5)$$

- $\eta_{STC}=P_{STC}/(G_{STC}\cdot A)$ is the rated PV plant efficiency defined in the Standard Test Conditions (STC) with irradiance G_{STC}=1000 W/m^2, module temperature T_{STC} = 25 °C and Air Mass AM=1.5. P_{STC} is the rated power of the PV system, and A is the area occupied by the PV modules.
- $\eta_{low,G}$ is the parameter introduced to incorporate the low-irradiance losses that are mentioned by the PV module manufacturer, but usually not quantified. In this case, instead of a complex polynomial equation [30], the PV conversion efficiency depends on the irradiance G according to a simple hyperbolic relation:

$$\eta_{low,G} = 1 - G_0/G \quad (6)$$

  where G_0 is the low-irradiance threshold, below which the PV modules do not produce due to too low irradiance (10÷50 W/m^2). As a result, the PV module efficiency increases with the increase of G and remains almost constant if G > 500 W/m^2.
- C_{Th} is a parameter useful to compute the temperature correction of P_{DC} that describes the linear dependence of the PV production on the PV module temperature:

$$C_{Th} = 1 + \gamma_T\cdot(T - T_{STC}) \quad (7)$$

When the module temperature T is above T_{STC} = 25°C, the production decreases proportionally by γ_T. It is the thermal

Authorized licensed use limited to: Politecnico di Torino. Downloaded on February 19,2024 at 16:23:56 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Industry Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIA.2024.3359121

coefficient of power, which is in the range -(0.3÷0.5) %/K for c-Si technology [15].

- $\eta_{mix} = \eta_{life}\ \eta_{dirt}\ \eta_{refl}\ \eta_{mis}\ \eta_{cable}$ is the formula of the overall performance of the plant. These experimental loss parameters are considered constant for all plants [31]. In particular, the PV production decreases due to the reflection effect (η_{refl}=0.973) and dirt deposited on the glass of the modules (η_{dirt}=0.976), the mismatch of current-voltage (*I-V*) curves (η_{mis}=0.97) and joule losses in cables (η_{cable}=0.99).
- The loss by ageing $\eta_l = 1-\gamma_{life}\cdot n_{life}$ is proportional to the product of the age of the plant n_{life} times the annual loss coefficient γ_{life}. According to the specifications included in the datasheet of the modules, the modules are certified to reduce their efficiency down to 80% of the initial value after 20 years (γ_{life}= -1%/year). Nevertheless, experimental results reported in [32] and [33] show that the efficiency reduction is lower, so that γ_{life}=-0.5%/year.
- η_{conv} represents the non-linear performance of the AC/DC converter that takes into account the efficiency of the Maximum Power Point Tracking (MPPT) and the inverter losses. The generic yield at the AC side is determined by a quadratic model, based on experimental data in [24][34]:

$$P_{loss} = P_0 + C_L\cdot P_{DC} + C_Q\cdot P_{DC}^2 \quad (8)$$

The model includes no-load losses P_0 (W) due to the supply of auxiliary circuits. Linear losses are expressed by using the parameter C_L (W^{-1}) that considers the conduction of diodes and IGBTs and the switching losses. Finally, the quadratic losses are expressed by using the parameter C_Q (W^{-2}) for the conduction of MOSFETs and the resistive contribution. As a result, in commercial devices, the maximum efficiency is about 98% [34]. The formula for the calculation of the converter efficiency is the ratio between the AC power output (P_{AC}) and the DC power input (P_{DC}). By expressing P_{DC} as the difference between input and losses, the resulting formula for the calculation of η_{conv} is:

$$\eta_{conv} = (P_{DC} - P_0 - C_L\cdot P_{DC} - C_Q\cdot P_{DC}^2)/\ P_{DC} \quad (9)$$

To use the same model for different plants, the previous formula is normalized by the rated power of the inverter.

The formula (5) permits to calculate the AC production for a PV plant, but it does not include limitations in the power injections due to the adequate operation of the grid. Requests of power limitation can be remotely sent to the PV plant by the transmission system operator in any moment. The simplest curtailment technique leads to the disconnection of the plants (or of parts of the plants) from the grid. In this case, to take into account the curtailment, it is necessary to import in the model the power limitation values $P_{AC,lim}$. For each considered time step, the power production will be the minimum value between P_{AC} and $P_{AC,lim}$.

III. Proposed Improvements in the Modelling of the Photovoltaic Production

The proposed improvements in the model parameters are the adjustment of the low-irradiance threshold G_0, the thermal coefficient of power, and the parameters used for determining the module temperature in both models presented in Subsection II.B. Moreover, two adjustment coefficients are proposed to better compute the production of PV plants during cloudy days and in case of other sources of losses such as component failures, poor maintenance, etc. The next subsections explain the effects of changing the parameters under analysis.

A. Adjustment of the Low-Irradiance Threshold G_0

The optimization of G_0 permits to better include the non-linearities occurring at low irradiance levels in the irradiance-electrical power conversion.

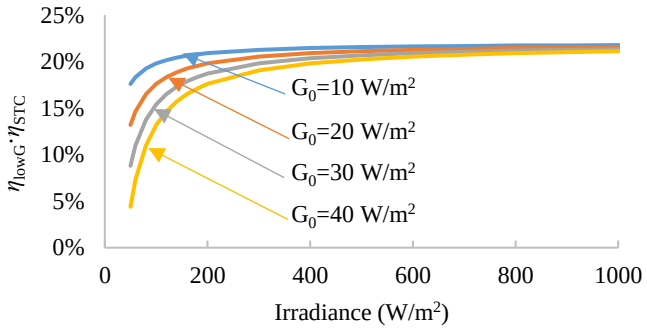

Fig. 1. Effect of G_0 variation on the efficiency $\eta_{STC}\cdot\eta_{low,G}$.

Fig. 1 shows the effect of the variation of G_0 on the relation between irradiance and the product of the rated efficiency of the modules η_{STC}=22% times the efficiency $\eta_{low,G}$. By increasing G_0, the model can better consider not only the irradiance-electrical power conversion non-linearities, but also additional optical losses that increase at low irradiance level and with sunbeams away from orthogonality with the PV module plan.

B. Improvement in the Calculation of the Module Temperature

The improvement in the calculation of the module temperature consists of changing the literature parameters with new ones. In the "*NOCT model*", the *NOCT* value is the only parameter to be changed. On the other hand, in the "*wind model*", it is possible to change one up to all the four semi-empirical parameters (e.g., a_T, b_T, c_T and d_T). A priori selection of the most effective parameter variation is not easy: it depends on the PV plants installation and on the weather conditions. For example, a ground-mounted PV plant installed in a windy site is strongly affected by wind speed. A building-integrated plant can be better modelled by changing the parameters proportional to irradiance and temperature. On the other hand, changing all the four parameters has a higher computational burden, without knowing if there will be an improvement of the model. A comparison of the different options is presented in this work.

C. Adjustment of the Thermal Coefficient of Power

Fig. 2 is created for a PV plant (P_{STC}=1 kW) in a clear sky day. To present the differences, it shows data from sunrise to midday, since the rest of the patterns are almost mirrored.

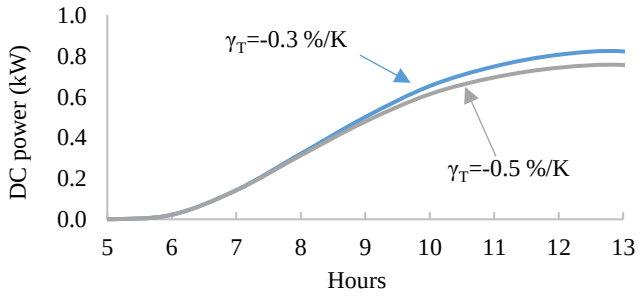

Fig. 2. Effect of the variation of the thermal coefficient of power γ_T on the DC power pattern of a generator with rated power of 1 kW during a clear-sky day.

Authorized licensed use limited to: Politecnico di Torino. Downloaded on February 19,2024 at 16:23:56 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Industry Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIA.2024.3359121

At midday, the irradiance $G \approx 1$ kW/m^2, while the air temperature $T_a \approx 27$°C. The DC power output pattern is calculated with a thermal factor γ_T in the range -0.3÷-0.5%/K. The significant effect of the variation of γ_T on the power production is demonstrated by the energy deviation, mainly occurring at high temperature and irradiance levels. By changing γ_T from the standard value (-0.5%/K for c-Si modules) to -0.3%/K, C_{Th} increases from 0.88 to 0.93. Thus, the effect of adjusting the parameter γ_T is the stretching of the PV power pattern mainly at midday, with higher effects in the summer.

D. Adjustment Coefficient for Non-Ideal Plants

The proposed *Ideality Factor IF* is multiplied to the nominal power of the system P_{STC} to consider the phenomena not included in the original model created for perfectly working plants (for which IF=1). For example, an aspect to consider is a high mismatch in the *I-V* curves of the PV modules. It can be due to a non-uniform installation (different tilts and azimuths), to the use of PV modules with different specifications, or to PV modules with unusual ageing and failure. Other losses identifiable with a low AC power can be the disconnection of parts of the PV plant [35], or large amounts of dirt [36].

E. Adjustment Coefficient for Non-Clear-Sky Days

The PV conversion model has been created to simulate clear-sky days rather than cloudy ones. It has been designed to estimate the PV production more accurately during the most productive days. Thus, there is a production overestimation in the winter. A multiplicative corrective factor called the "Non-Clear-Sky Day Factor" ($f_{ncsd,i}$) for the i-th day, is introduced. This corrective factor corrects the AC power production P_{AC} to obtain the new weighted value $P_{AC,w}$, as follows:

$$P_{AC,w}=P_{AC}\cdot(1- f_{ncsd,i}) \quad (10)$$

This corrective factor is calculated for each plant and for each i-th day, with the following quadratic relation:

$$f_{ncsd,i} = a_{DID}\cdot DID_i^2 + b_{DID}\cdot DID_i + c_{DID} \quad (11)$$

where a_{DID}, b_{DID} and c_{DID} are the coefficients of the quadratic expression. The parameter Daily Irradiation Deviation (DID) is the relative deviation between the global daily measured irradiation H_i and the global daily irradiation calculated in clear sky conditions $H_{ClearSky,i}$ [37]:

$$DID_i=(H_{ClearSky,i} - H_i)/H_{ClearSky,i} \quad (12)$$

Thus, for each plant under analysis, the hourly irradiation pattern in clear sky days is calculated for each i-th day of the year. The clear sky irradiation calculation can be performed according to [38] and [39]. Fig. 3 shows the calculation of $f_{ncsd,i}$ as a function of the DID. The values used in these quadratic formulas are the optimized values in two configurations of the case study explained in Section V.

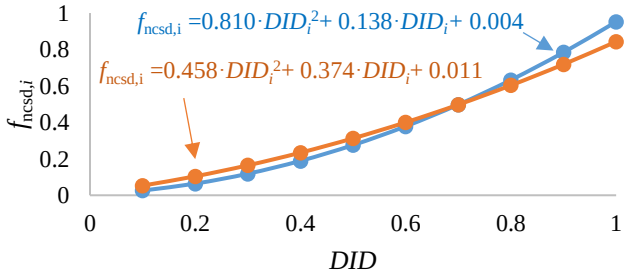

Fig. 3. $f_{ncsd,i}$ as a function of Daily Irradiation Deviation (DID).

IV. Procedure for the Analysis of PV Plant Groups

The procedure for the analysis of a large set of PV plants requires the collection of the data about the plants, the measured production patterns, and the weather data. After that, the production patterns are filtered to remove bad data, and the PV plant grouping is performed. The grouping is necessary to check the statistical validation of the data remaining after the filtering with respect to the whole set of plants. After the definition of the group of PV plants that represents the whole dataset, the models are optimized to calculate the production patterns with reduced energy deviation with respect to the measured energy. Finally, the energy deviations are presented to compare the model performance and the optimization steps (single or double) under analysis. The key points of this procedure are shown in the next subsections.

A. Import of Weather Data and PV Plant Information

The measured production patterns of the PV plants can be analyzed, and simulation can be performed, only if the same basic information is available. First at all, the rated power P_{STC} is necessary to analyze the measured patterns and to perform simulations. As explained in the next subsection, the rated powers are also the variables used to classify the plants. The site coordinates are essential to obtain meteorological data [27][33]. The tilt and azimuth of the PV modules can be obtained by the layout of the plants, by inspections, or by satellite/street images. This information is necessary to obtain the irradiance on their surface at each considered time step. Finally, the installation year is necessary to calculate the degradation losses [40].

B. Production Pattern Filtering

The production patterns obtained by distributed meters can be affected by errors that occur in data measurement, transmission, and storage. In many cases, it is not possible to define the source of the bad data. Patterns with unexpected trends could be the results of plant shutdowns and failures of the components. Thus, an accurate filtering process of the data is necessary. The procedure presented in [41] deals with the filtering of thousands of patterns, without the possibility of accessing detailed information of the plants. The only available information is the basic information described in the previous sub-section. The first criterion is the check of the absence of production during night hours, as such production is obviously physically impossible. The second criterion is the check of the absence of days of data, due to the failure of the monitoring infrastructure. In fact, the patterns with missing weeks or months of data (especially during summer) affect the calculation of the performance of the plant. The third criterion is the check of the typical territorial range productivities. In particular, if the yearly specific production in kWh/kW/year is out of the accepted range, the performance is too low or too high and the pattern is removed. Productivity maps, such as those available in [33], can be used to define the boundaries of the acceptable performance ranges, depending on the site location.

C. Data Classification Based on PV Plant Rated Power

Inferential statistical methods allow us generalizing the results to an entire population of elements with an acceptable confidence, by analyzing a small sample of elements. These methods can be used in the present work, i.e., in case of large groups of PV systems, where many data about the PV plants (the power patterns) are affected by errors. Thus, the representative elements of the population are the rated powers

Authorized licensed use limited to: Politecnico di Torino. Downloaded on February 19,2024 at 16:23:56 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Industry Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIA.2024.3359121

of the PV plants. The Neyman's Stratified Sampling (SS) technique [42] has been applied. As described in [41], the SS technique indicates how many PV production patterns (among the ones that remained after data filtering) are significantly enough to statistically represent the entire population of PV plants.

D. Model Optimization and Energy Deviations Calculation

The goal of the optimization is to find the best parameters (described in Section III) to minimize the differences, over a defined time frame, between the measured and calculated production patterns. In this paper, the parameters optimization is performed in a double-step procedure. The first step involves the optimization of most of the parameters (with the exception of the parameters appearing in the quadratic formula of the corrective factor $f_{\mathrm{ncsd},i}$ described in subsection IV.F). In this case, the period under analysis is from April to September (from now on, "summer semester"). In the second step, the optimization of the parameters in the formula of $f_{\mathrm{ncsd},i}$ is carried out only from October to March (from now on, "winter semester") keeping constant the other previously optimized parameters. The rationale of this choice is the largest presence of non-clear-sky days in the winter semester.

Regarding the determination of the optimal parameters, the deviation between measured and simulated patterns is calculated as the sum of their root mean square differences at each time step of the period under analysis. For the analysis of a portfolio of $j = 1, \ldots, J$ plants, the average quadratic deviation is calculated on the difference between the estimated power values P_{AC} and the measured values P_{ACm}. For each plant, this quantity is normalized with respect to the rated power P_{STC} of the PV plant. The calculation is performed over the time optimization horizon t_{opt} discretized with k time steps. Considering the vector $\mathbf{x}$ that includes all the variables defined in Section III, the optimization problem is formulated as:

$$\min \sigma(\mathbf{x}) \quad \text{subject to } \mathbf{x}_{\mathrm{lb}} \le \mathbf{x} \le \mathbf{x}_{\mathrm{ub}} \tag{13}$$

$$\sigma(\mathbf{x}) = \sum_{j=1}^{J} \sqrt{\frac{1}{T_{\mathrm{opt}}} \sum_{k=1}^{K} \left(P_{\mathrm{AC},k}^{(j)}(\mathbf{x}) - P_{\mathrm{ACm},k}^{(j)}\right)^2} \Big/ P_{\mathrm{PV},j} \tag{14}$$

where $\mathbf{x}_{\mathrm{lb}}$ and $\mathbf{x}_{\mathrm{ub}}$ are the lower and upper bounds, respectively, set up on the entries of the vector $\mathbf{x}$.

To perform the double-step procedure more easily, the vector $\mathbf{x}$ is divided into three sub-vectors of variables. The first sub-vector $\mathbf{x}_1$ includes the thermal factor $\gamma_{T\%}$, the low irradiance threshold G_0, and the ideality factor IF:

$$\mathbf{x}_1 = [\gamma_{T\%},\ G_0,\ IF]^{\mathrm{T}} \tag{15}$$

The second sub-vector $\mathbf{x}_2$ includes the variables to be modified for better calculation of the cell temperature T. In case of the "*NOCT model*" (3), only the *NOCT* value is optimized. Such equilibrium condition is strongly affected by the installation conditions, which often do not permit an optimal heat dissipation in actual PV plants. This happens, e.g., in case of building-applied or building-integrated plants, in which the PV modules have steady-state temperatures at T_a=20°C and G_{NOCT} = 800 W/m² higher than the declared *NOCT*.

$$\mathbf{x}_{2_\mathrm{NOCT}} = [NOCT] \tag{16}$$

In the "*wind model"* defined in (4), all the four parameters can be modified. The change of the parameter a_{T} quantifies the effect of air temperature on the module temperature. The same approach is individually applied by considering b_{T} and c_{T} for the effect of irradiance and wind speed, respectively. If all the four parameters are changed, the vector is:

$$\mathbf{x}_{2w} = [a_{\mathrm{T}}, b_{\mathrm{T}}, c_{\mathrm{T}}, d_{\mathrm{T}}]^{\mathrm{T}} \tag{17}$$

The third sub-vector $\mathbf{x}_3$ includes the parameters for the calculation of the coefficient $f_{\mathrm{ncsd},i}$ for non-clear-sky days. They are modified for better calculation of the winter production:

$$\mathbf{x}_3 = [a_{\mathrm{DID}}, b_{\mathrm{DID}}, c_{\mathrm{DID}}]^{\mathrm{T}} \tag{18}$$

Concerning the constraints, for the parameters referring to technical aspects of PV plants, the lower and upper bounds $\mathbf{x}_{\mathrm{lb}}$ and $\mathbf{x}_{\mathrm{ub}}$ are defined in the literature. For example, the thermal coefficient of power is in the range -(0.3÷0.5) %/K for c-Si modules. The coefficient *IF* could be close to unity, but a specific range cannot be defined at priori, so that large limits are imposed to avoid reaching the limits.

E. Power Patterns Comparison and Energy Deviations

The results are analyzed by comparing the hourly estimated pattern with the measured production. The patterns are calculated by using the energy model (5) and one of the temperature models (3) and (4). The error between the different patterns is assessed in energy terms:

$$\Delta E_{\%,\mathrm{s},i}^{\Delta t} = 100 \cdot \left(E_{\mathrm{s},i} - E_{\mathrm{m},i}\right) / E_{\mathrm{m},i}\big|_{\Delta t} \tag{19}$$

where $E_{\mathrm{s},i}$ is the simulated energy in the sth model and $E_{\mathrm{m},i}$ is the measured energy for the ith plant, for the same timeframe Δt (one year). To compute errors for shorter times (e.g., one day), it is useful to compute the residual ratio for each configuration:

$$R_{\mathrm{s},i}^{\Delta t} = E_{\mathrm{s},i} / E_{\mathrm{m},i}\big|_{\Delta t} \tag{20}$$

The error calculated with (20) is positive when the model overestimates the production with respect to the measurements. In that case, the residual ratio is higher than unity.

V. Case Study and Optimization Results

The set of PV plants under analysis includes plants installed in Lombardy, a region in northern Italy, Latium (central Italy) and Sicily (southern Italy). In 2018, there were 125,284 plants (2.3 GW) in Lombardy, 54,285 in Latium (1.3 GW), and 52,731 in Sicily (1.4 GW). Thus, the sum of the rated powers of the whole dataset is ≈5 GW and the total number of plants is 232,300. Only ≈2% of these plants has rated power higher than 120 kW but correspond to about 64% of the total installed power. Moreover, 89% of these plants have polycrystalline silicon modules, while ≈10% have m-Si modules. The general data available for the PV plants are the coordinates of the installation sites, rated power, and installation year. The hourly production patterns are measured by the DSOs and transferred to the TSO, according to a regulated procedure. These hourly AC powers are not available for all the plants; they are provided for a subgroup of 20,471 plants for Lombardy, 11,061 for Latium and 11,113 for Sicily (total 42,645 patterns).

Subsection V.A presents the plants classification based on the 232,300 rated powers, and the result of the filtering procedure on the 42,645 hourly patterns. Subsection V.B presents the results of the optimization of the production models on the plants remaining after the filtering procedure.

A. Rated Power Clustering and Hourly Pattern Filtering

The clustering procedure described in [41] is applied to the 232,300 values of rated power. The result is the definition of 10 classes of plants (Table II). In parallel to the creation of the classes, the data filtering procedure is applied to the 42,645

Authorized licensed use limited to: Politecnico di Torino. Downloaded on February 19,2024 at 16:23:56 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Industry Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIA.2024.3359121

power patterns. Following the procedure in [41], patterns affected by issues (i.e., errors in data measurement and storage, or poorly functioning generators) are removed, and a sample of PV plants is selected for optimizing the model parameters. A drastic selection process that rejects any pattern having even a single inaccuracy has been applied, so that most patterns have been removed, and 366 plants are used in the optimization.

TABLE II. CLASSES OF PLANTS AND RELATED POWER RANGES

	Power ranges (kW)	Number of plants		
Class		**Lombardy**	**Latium**	**Sicily**
1	$P \leq 3.5$	49,290	22,170	15,710
2	$3.5 < P \leq 6.5$	49,605	22,377	24,242
3	$6.5 < P \leq 12.5$	7,482	4,320	4,870
4	$12.5 < P \leq 25$	8,118	2,923	4,443
5	$25 < P \leq 70$	4,693	1,086	1,476
6	$70 < P \leq 120$	3,316	586	1,079
7	$120 < P \leq 500$	2,021	439	403
8	$500 < P \leq 1200$	696	248	393
9	$1200 < P \leq 3600$	54	81	66
10	$3600 < P$	9	55	49

Table III shows the distribution of the 366 plants in the classes (the empty classes are not shown). Finally, to statistically validate the subgroups of 139, 73, and 154 PV plants with respect to the whole dataset, the stratified sampling procedure has been applied, as defined in Section III.

TABLE III. DISTRIBUTION OF PLANTS AFTER FILTERING IN THE CLASSES

Class		**4**	**5**	**6**	**7**	**8**	**9**	**10**
Number of plants after filtering	**Lombardy**	1	10	26	25	66	9	2
	Latium	1	6	15	11	18	15	7
	Sicily	2	8	68	15	45	7	9

The result is that the statistical validation is confirmed mainly for the classes including the biggest PV plants (e.g., in Lombardy, classes 6, 8, 9 and 10), as shown in Table IV. The first consideration is that generally bigger PV plants have better maintenance and individual monitoring of the energy production; thus, the production patterns are correct. Nevertheless, in case of Sicily, there is a lack of correct patterns also for big plants. Secondly, the data collection infrastructure should be improved, especially for small PV plants. In fact, classes from 1 to 3 are empty, and also class 4 is never validated.

TABLE IV. DISTRIBUTION OF PLANTS AFTER FILTERING IN THE CLASSES

Class		**4**	**5**	**6**	**7**	**8**	**9**	**10**
Validation	**Lombardy**	no	no	yes	no	yes	yes	yes
	Latium	no	yes	yes	yes	yes	yes	yes
	Sicily	no	yes	yes	no	yes	no	no

B. Reduction of the Energy Deviations

In the present work, the optimization is performed minimizing the differences between measurements and simulations on hourly values. The key point for the improvement is the use of a double-step procedure, in which the parameter optimization is performed on different periods. After analyzing different time frames, the best results were obtained with the following configuration. In the first step, the vectors of parameters related to the literature model ($\mathbf{x}_1$ and $\mathbf{x}_2$) are optimized comparing measurements and simulated patterns only during the summer semester. The reason is that the model was created mainly for clear sky days. In the second step, the parameters for the calculation of $f_{ncsd,i}$ are adjusted only in the winter semester. Following this procedure, the performance of all the models increases. As an example, Fig. 4 and Fig. 5 show the improvement of the use of this double-step procedure in both summer and winter days for a PV plant in Sicily. The day with passing clouds (Fig. 5, top) exhibits the highest variations in hourly AC power. The proposed double-step optimization matches the measurements better than the single step optimization also in this case, as for all the other days. As expected, the clear-sky daily production patterns (already close to the measurements) are less affected by the use of $f_{ncsd,i}$ with respect to cloudy days.

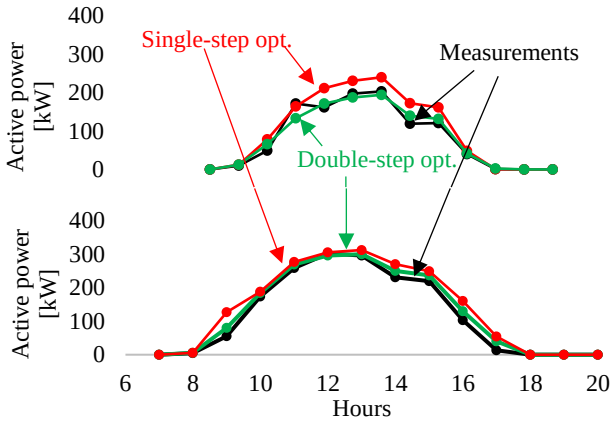

Fig. 4 Examples of optimized production patterns vs. measurements in a cloudy day (top) and in an almost clear-sky day (bottom) in winter.

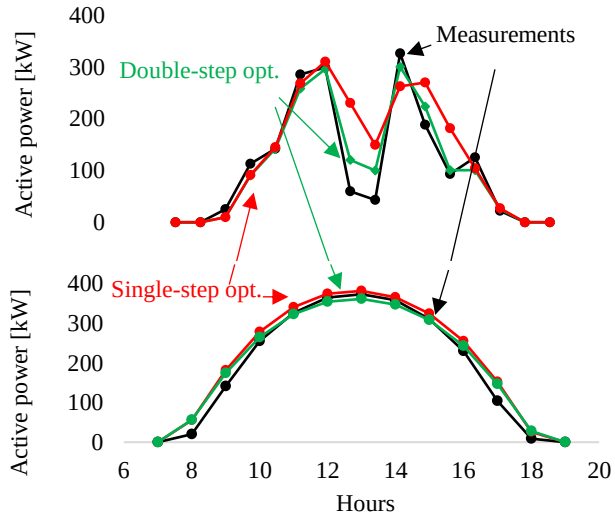

Fig. 5 Examples of optimized production patterns vs. measurements in a day with passing clouds (top) and in a clear-sky day (bottom) in summer.

The improvements due to the double-step optimization are significant both for the single plant and the whole group. Fig. 6 shows a box plot related to the whole group of plants in Sicily. It shows the monthly energy deviation (%), where a positive value corresponds to an overestimation made by the model. Each rectangle is the interquartile range (IQR) and includes 50% of the population, while segments are delimited by the maximum and minimum values. As shown in Fig. 6 (top), a single-step optimization (SSO), carried out on the whole year, leads to an overestimation of the winter semester production and a slight underestimation during the summer semester, which compensate for each other on a yearly basis. The proposed procedure mainly solves these issues reducing the deviation down to values generally lower than ±8% for the IQRs in most of the months (bottom of Fig. 6). In December, the deviation is higher but the energy contribution is low (about 5% of the PV annual group production).

Authorized licensed use limited to: Politecnico di Torino. Downloaded on February 19,2024 at 16:23:56 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Industry Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIA.2024.3359121

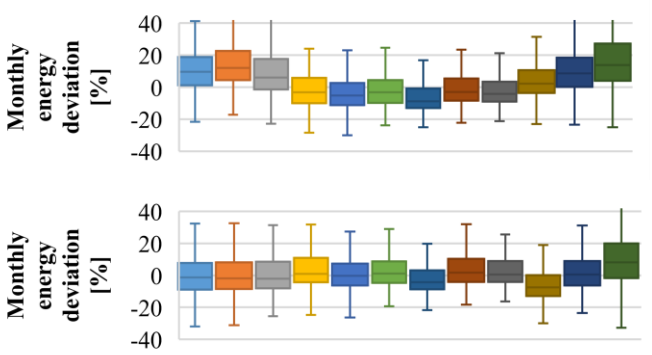

Fig. 6 Montly energy deviations between measurements and simulated patterns with a single-step (top), and double-step optimization (bottom).

C. *Energy Deviations for PV Plants Classes in Sicily*

The improvements described in the previous subsection are quantified in Table V for the group of plants in Sicily. This table shows, for each class of plants, the comparison of the yearly energy deviations (simulations vs. measurements). The optimization of each class is independent of the other classes, and the obtained parameters are different. The analyzed configurations are the following:

- CONFIG#1 and #2: NOCT (3) and wind (4) models are used with default parameters and no optimization is performed.
- CONFIG#3: in addition to the parameters $\gamma_{T\%}$, G_0, IF, the NOCT is adjusted in a single step, $\mathbf{x}=[\gamma_{T\%}, G_0, IF, NOCT]^T$.
- CONFIG#4: in addition to the parameters $\gamma_{T\%}$, G_0, IF, the parameter depending on air temperature in the wind model (4) is adjusted in a single step $\mathbf{x}=[\gamma_{T\%}, G_0, IF, a_t]^T$.
- CONFIG#5: with respect to CONFIG#4, in the wind model (4) the parameter depending on irradiance is the only one adjusted (in a single step) $\mathbf{x}=[\gamma_{T\%}, G_0, IF, b_t]^T$.
- CONFIG#6: with respect to CONFIG#4, in the wind model (4) the parameter depending on wind speed is the only one adjusted (in a single step) $\mathbf{x}=[\gamma_{T\%}, G_0, IF, c_t]^T$.
- CONFIG#7: with respect to CONFIG#4, the difference is that all the parameters in the wind model (4) are adjusted in a single step $\mathbf{x}=[\gamma_{T\%}, G_0, IF, a_t, b_t, c_t, d_t]^T$.
- CONFIG #8: the first step optimizes $\mathbf{x}_{step1}=[\gamma_{T\%}, G_0, IF, NOCT]^T$ in the summer semester. The second step optimizes $\mathbf{x}_{step2}=[a_{DID}, b_{DID}, c_{DID}]^T$ in the winter semester.
- CONFIG #9: the first step optimizes $\mathbf{x}_{step1}=[\gamma_{T\%}, G_0, IF, a_t]^T$ in the summer semester. The second step optimizes $\mathbf{x}_{step2}=[a_{DID}, b_{DID}, c_{DID}]^T$ in the winter semester.
- CONFIG #10: the first step optimizes $\mathbf{x}_{step1}=[\gamma_{T\%}, G_0, IF, b_t]^T$ in the summer semester. The second step optimizes $\mathbf{x}_{step2}=[a_{DID}, b_{DID}, c_{DID}]^T$ in the winter semester.
- CONFIG #11: the first step optimizes $\mathbf{x}_{step1}=[\gamma_{T\%}, G_0, IF, a_t, b_t, c_t, d_t]^T$ in the summer semester. The second step optimizes $\mathbf{x}_{step2}=[a_{DID}, b_{DID}, c_{DID}]^T$ in the winter semester.

The lowest yearly deviations are obtained with the double-step optimizations (<2%). Deviations in the ranges ±(2÷5)% are obtained in the classes in which the number of plants is very low (in class 4 and 5, there are only 2 and 8 plants, respectively). Among all, CONFIG #8 and #11 are the best configurations.

The improvement due to the double-step optimization is a lower deviation not only on yearly basis, but also on both the winter and summer semesters. Table VI shows the energy deviations in the winter semester, generally with an important overestimation of the production by using literature models.

TABLE V. ANNUAL ENERGY DEVIATIONS (%) FOR THE CLASSES OF PV PLANTS IN SICILY

Config#	Class						
	4	**5**	**6**	**7**	**8**	**9**	**10**
Literature models without optimization and default parameters							
1	1.1	-9.2	-7.1	5.8	-2.9	-6.3	-11
2	5.8	-4.6	-2.5	11.2	2.1	-1.3	-6.5
Single-step optimization							
3	1.3	9.7	5.4	9.1	6.3	6.5	2.6
4	-3.7	9.9	5.6	10.7	5.5	7.5	3.7
5	-4.7	9.6	6.6	10.5	6.6	7.5	4
6	-6.4	10.1	4.9	11	6.1	8	3.5
7	5.7	11.9	6.5	11.6	6.6	8.8	5.5
Double-step optimization							
8	0	5	0.9	0.8	1	0.8	-2.5
9	-4.9	4.7	0.9	1.4	0.9	0.8	-2.5
10	-4.9	4.7	1	1.2	0.9	1	-2.2
11	-0.1	4.6	1	1.7	0.8	1	-2.1

TABLE VI. ENERGY DEVIATIONS (%) IN WINTER SEMESTER (OCTOBER-MARCH) FOR THE CLASSES OF PV PLANTS IN SICILY

Config#	Classes						
	4	**5**	**6**	**7**	**8**	**9**	**10**
Literature models without optimization and default parameters							
1	-3.6	-1.5	-1.0	20.7	6.8	4.0	-2.7
2	0.9	3.7	4.2	27.3	12.7	9.7	2.4
Single-step optimization							
3	-2.2	18.5	14.1	23.7	15.9	17.6	11.8
4	-3.1	20.1	14.1	26.4	14.4	19.9	14.7
5	-6.2	19.1	17.0	26.4	17.4	19.5	15.0
6	-11.5	20.8	12.2	27.0	15.9	20.1	13.9
7	11.7	26.9	16.7	28.4	17.2	23.0	18.9
Double-step optimization							
8	-6.2	4.7	0.9	-0.9	1.8	2.2	-2.1
9	-6.9	4.7	0.5	-1.1	2.3	2.2	-2.1
10	-6.7	4.6	0.9	-0.8	2.3	2.3	-2.0
11	-6.3	5.1	0.8	-0.9	1.9	2.1	-1.9

D. *Energy Deviations and Optimized Parameters Obtained for the PV Plant Group in Sicily*

After the optimization of the PV production models class by class, the procedure is repeated considering all the PV plants of the group, regardless of their rated power.

TABLE VII. ENERGY DEVIATIONS FOR ALL PV PLANTS IN SICILY

Config#	Energy deviation (%)		
	Yearly	**Apr-Sept**	**Oct-March**
Literature models without optimization and default parameters			
1	-1.3	-7.4	9.4
2	3.8	-2.8	15.4
Single-step optimization			
3	2.1	3.3	11.4
4	2.6	-2.9	12.1
5	4.2	-3	16
6	-2.7	-2.8	13.7
7	3.7	-3.1	15.3
Double-step optimization			
8	-2.4	-3.3	-0.8
9	-2.2	-3	-0.8
10	-2.2	-3	-0.8
11	-2.3	-3.1	-0.8

Authorized licensed use limited to: Politecnico di Torino. Downloaded on February 19,2024 at 16:23:56 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Industry Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIA.2024.3359121

The double-step optimization reduces the deviations, especially in the winter semester, from values higher than 10% down to <1%. On a yearly basis, the deviations are in the range -2÷-3% (Table VII). The optimized parameters for the two best configurations, i.e., CONFIG #8 and CONFIG #11, are shown in Table VIII. There are minor differences in the parameters in the two configurations. Moreover, *IF* is >1: the reason could be an ageing parameter overestimation. Only the parameters of the coefficient $f_{ncsd,i}$ are different; nevertheless, the effect on these differences is negligible on the production pattern computation.

TABLE VIII. OPTIMIZED PARAMETERS FOR THE WHOLE GROUP IN SICILY

Parameter		CONFIG #8	CONFIG #11
Conversion model	$\gamma_{T\%}$ (1/K)	-0.004	-0.005
	G_0 (W/m^2)	11	14
	IF (-)	1.022	1.036
Module temperature	*NOCT* (°C)	47	-
	a_t (-)	-	0.537
	b_t (°C·m^2/W)	-	0.053
	c_t (°C·s/m)	-	2.513
Cloudy days	a_{DID} (-)	0.810	0.458
	b_{DID} (-)	0.138	0.374
	c_{DID} (-)	0.004	0.011

Also, the SSO cannot drastically reduce the deviations; indeed, the annual SSO leads to a better match of the summer production, with a consequent worse match in the winter. On the contrary, the double-step optimization always reduces the deviations. Excluding classes 4 and 5 characterized by a low number of small plants, the deviations are in the range ±2.5%.

E. Results for the Other Groups of PV Plants

The results shown for Sicily are similar to the main results for the other two groups of PV plants in Latium and Lombardy. In all the analyzed regions, the double-step procedure reduces the energy deviations on annual basis, both on the winter and summer semesters. The best results are obtained for the classes with the highest number of PV plants. Considering the whole groups, regardless of the nominal power classes, the energy deviations after the double-step optimization are always <2%, and the deviations in the winter semester are <3%. Also in these regions, the differences between the use of the double-step optimized NOCT (3) and wind (4) models are negligible. The only noticeable difference in the parameters of the entire groups is the ideality factor *IF*, which is in the range 1.04÷1.07 in Latium, and 1.06÷1.09 in Lombardy. Thus, the geographical position of the three groups did not affect the procedure.

F. Smoothing Effect of the Aggregation of PV Plant Profiles

Besides the daily PV energy production, the difference between the AC average power at two successive hours is relevant to establish the need for the TSO to procure ancillary service at different hours and at different scales (substation, region, national). In this respect, the ratio χ between the hourly power variation and the PV rated power can provide useful information. From the real data of 154 hourly profiles of Sicilian PV systems in a day with partially cloudy sky, this ratio could reach about 0.5 for a single plant, especially for small plants (Fig. 7 left). However, if the PV plants are aggregated, there is a smoothing effect that reduces the hourly average power variations and the corresponding ratio χ_{agg}. From the aggregate daily pattern (Fig. 7 right), the sum of the rated power values is 112 MW, and the maximum ratio χ_{agg} is 0.19 at hour 16:00.

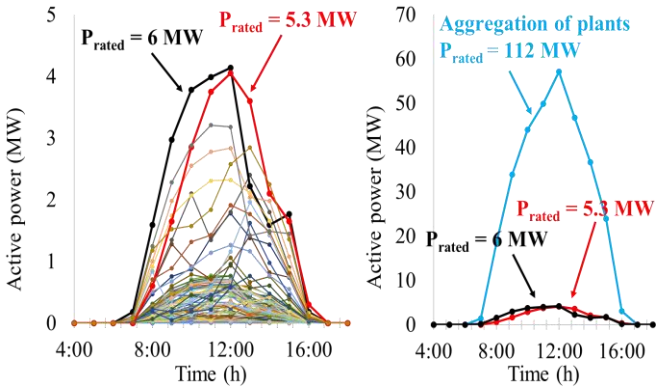

Fig. 7 Daily profiles for 154 PV plants (left) and their aggregation (right).

By considering all the days in the year, Fig. 8 shows that the CDF of the ratio χ_{agg} is at most 0.26, while it is lower than 0.2 for 97.7% of the hours and lower than 0.077 for 50% of the hours. In practice, taking the aggregate average power at a given hour, in most cases the variation of the aggregate average power at the next hour is lower than 20%. This result is significant for assisting the ancillary service procurement by the TSO.

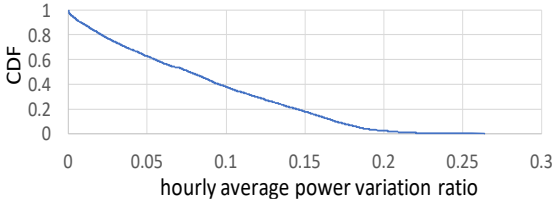

Fig. 8 CDF of the hourly average power ratio for the 154 PV plants.

VI. CONCLUSIONS

This paper has proposed a procedure that enhances the PV production model at first in the summer semester, then in the winter semester. The first optimization step includes the parameter of irradiance to power conversion, and the temperature coefficient of maximum power. Moreover, an ideality factor is introduced to consider all the other phenomena not included in the literature model. The second optimization step reduces the deviations between simulation and measurements in the winter: the parameters of a novel quadratic formula are optimized to better simulate the PV production in cloudy days. The results demonstrate the effectiveness of the proposed procedure: the energy deviations are generally <2% on an annual basis, and <3% during winter months. After optimizing the model for classes of PV plants based on the rated power, the procedure is repeated for the PV groups in three different regions of Italy. Increasing the number of PV plants and considering together generators with different rated powers lead to yearly deviation <3% in the different geographical areas. Future work will address the data quality of the PV production patterns, adjusting the data filtering to obtain a statistically significant number of data for all the classes of PV plants under analysis.

ACKNOWLEDGMENT

This study was carried out within the Ministerial Decree no. 1062/2021 and received funding from the FSE REACT-EU - PON Ricerca e Innovazione 2014-2020. This manuscript reflects only the authors' views and opinions, neither the

Authorized licensed use limited to: Politecnico di Torino. Downloaded on February 19,2024 at 16:23:56 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Industry Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIA.2024.3359121

European Union nor the European Commission can be considered responsible for them.

REFERENCES

[1] I. Bonilla-Campos, F.J. Sorbet, and D. Astrain, "Radical change in the Spanish grid: Renewable energy generation profile and electric energy excess," *Sust. Energy, Grids and Networks*, vol. 32, art. 100941, 2022.

[2] D. Remon, A.M. Cantarellas, and P. Rodriguez, "Equivalent Model of Large-Scale Synchronous Photovoltaic Power Plants," *IEEE Trans Ind Appl*, vol. 52, no. 6, pp. 5029–5040, 2016.

[3] A. Saez-de-Ibarra, V.I. Herrera, A. Milo, H. Gaztañaga, I. Etxeberria-Otadui, S. Bacha, and A. Padrós, "Management Strategy for Market Participation of Photovoltaic Power Plants Including Storage Systems," *IEEE Trans Ind Appl*, vol. 52, no. 5, pp. 4292–4303, 2016.

[4] S. Patel, M. Ahmed, and S. Kamalasadan, "A Novel Energy Storage-Based Net-Load Smoothing and Shifting Architecture for High Amount of Photovoltaics Integrated Power Distribution System," *IEEE Trans Ind Appl*, vol. 56, no. 3, pp. 3090–3099, 2020.

[5] B.I. Crăciun, T. Kerekes, D. Séra, and R. Teodorescu, "Frequency Support Functions in Large PV Power Plants with Active Power Reserves," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 2, no. 4, pp. 849–858, 2014.

[6] T.T. Ku *et al*., "Enhancement of Power System Operation by Renewable Ancillary Service," *IEEE Trans Ind Appl*, vol. 56, no. 6, pp. 6150–6157, 2020.

[7] S.A. Julien, A. Sajadi, and B.M. Hodge, "Hierarchical Control of Utility-Scale Solar PV Plants for Mitigation of Generation Variability and Ancillary Service Provision," *IEEE Trans. on Sustainable Energy*, vol. 13, no. 3, pp. 1383–1395, 2022.

[8] ENTSO-E, Network codes, https://www.entsoe.eu/network_codes/, (access. 16 June 2023).

[9] M. Fuentes *et al*., "Application and validation of algebraic methods to predict the behaviour of crystalline silicon PV modules in Mediterranean climates," *Solar Energy*, vol. 81, no. 11, 2007, pp.1396-1408.

[10] J. Park *et al*., "A probabilistic reliability evaluation of a power system including Solar/Photovoltaic cell generator," *IEEE Power and Energy Society General Meeting-PES'09*, pp. 1–6, 2009.

[11] C.R., Osterwald, "Translation of device performance measurements to reference conditions," *Solar Cells*, vol. 18, pp. 269–279, 1996.

[12] Humada, Ali M., Mojgan Hojabri, Saad Mekhilef, and Hussein M. Hamada. "Solar cell parameters extraction based on single and double-diode models: A review." *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 494-509, 2016.

[13] S. Prakash *et al*., "Modeling and Performance Analysis of Simplified Three-Diode Photovoltaic Module." *Journal of Electrical Engineering* vol. 1, pp. 55-64, 2022.

[14] A. Kumar and R. Agarwal, "Mathematical Modeling and Analysis of Single-Diode, Double-Diode and Triple Diode based PV Module," *2023 International Conference for Advancement in Technology (ICONAT)*, Goa, India, pp. 1-6, 2023.

[15] J.H. Bae, D.Y. Kim, J.W. Shin, S.E. Lee, and K.C. Kim, "Analysis on the Features of NOCT and NMOT Tests with Photovoltaic Module," *IEEE Access*, vol. 8, pp. 151546–151554, 2020.

[16] International Electrotechnical Commission, "IEC 61215-1, Terrestrial photovoltaic (PV) modules - Design qualification and type approval", Available online: https://webstore.iec.ch/. Accessed on December, 2023.

[17] G. Tamizhmani *et al*., "Photovoltaic Module Thermal/Wind Performance: Long-Term Monitoring and Model Development For Energy Rating," *NCPV Solar Program Rev. Meeting*, Denver, Colorado, USA, 2003, pp. 936–939.

[18] D.L. King, W.E. Boyson and Kratochvil, "Report: Photovoltaic Array Performance Model, " Sandia National Laboratories, 2004 [online] Available: https://www.osti.gov/servlets/purl/919131sca5ep/, accessed on December 2023.

[19] M. Mattei *et al*., "Calculation of the polycrystalline PV module temperature using a simple method of energy balance", *Ren. Energy*, vol. 31, no. 4, pp. 553-567, 2006.

[20] R. Oliveira, C.L.T. Borges, "A comparison of photovoltaic models for estimating power generation: a case study of Brazilian data," *Clean Techn Environ Policy*, vol. 23, pp. 517–530, 2021.

[21] A. Ciocia, G. Chicco, and F. Spertino, "Optimisation of Generation Models for Clusters of Photovoltaic Plants," *Proc. MELECON 2022 - IEEE Mediterranean Electrotechnical Conference*, pp. 849–854, 2022.

[22] E. Bompard *et al*., "Assessing the role of fluctuating renewables in energy transition: Methodologies and tools," *Applied Energy*, vol. 314, 2022.

[23] L. Kaci *et al*., "Solar inverter performance prediction," *2020 6th International Symposium on New and Renewable Energy (SIENR)*, Ghadaia, Algeria, pp. 1-5, 2021.

[24] F. Spertino *et al*., "An experimental procedure to check the performance degradation on-site in grid-connected photovoltaic systems", *Proceedings of the 2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*, pp. 2600-2604, 2014.

[25] G.G. Kim *et al*., "Prediction Model for PV Performance With Correlation Analysis of Environmental Variables," *IEEE J Photovolt*., vol. 9, no. 3, pp. 832-841, May 2019.

[26] B. Hashemi *et al*., "Systematic photovoltaic system power losses calculation and modeling using computational intelligence techniques," *Applied Energy*, vol. 284, art. 116396, 2021.

[27] "HOME - SoDa." https://www.soda-pro.com/, (access. on Jun. 2023).

[28] A.A.M. Omara, *et al*., "Estimation of Solar Radiation of Khartoum City Using ASHRAE Model," *Proceedings of: 2020 International Conference on Computer, Control, Electrical, and Electronics Eng., ICCCEEE 2020*, Feb. 2021.

[29] S.B. Schujman *et al.*, "Evaluation of protocols for temperature coefficient determination," *2015 IEEE 42nd Photovoltaic Specialist Conference, PVSC 2015*, Dec. 2015.

[30] F. Mavromatakis, F. Vignola, B. Marion, "Low irradiance losses of photovoltaic modules," *Solar Energy*, vol. 157, pp. 496-506, 2017.

[31] F. Spertino *et al.*, "A power and energy procedure in operating photovoltaic systems to quantify the losses according to the causes," *Solar Energy*, vol. 118, pp. 313–326, 2015.

[32] F. Spertino *et al*., " Maintenance Activity, Reliability, Availability, and Related Energy Losses in Ten Operating Photovoltaic Systems up to 1.8 MW," *IEEE Trans Ind Appl*, vol. 57, no. 1, pp. 83–93, 2021.

[33] "PVGIS Photovoltaic Geographical Information System." https://joint-research-centre.ec.europa.eu/pvgis-photovoltaic-geographical-information-system_en (accessed Dec. 31, 2022).

[34] F. Spertino *et al*., " Reliability analysis and repair activity for the components of 350 kW inverters in a large scale grid-connected photovoltaic system," *Electronics (Switzerland)*, vol. 10, no. 5, pp. 1–13, 2021.

[35] F. Spertino and F. Corona, "Monitoring and checking of performance in photovoltaic plants: A tool for design, installation and maintenance of grid-connected systems," *Renew Energy*, vol. 60, pp. 722–732, 2013.

[36] W. J. Jamil *et al*., "Modeling of Soiling Derating Factor in Determining Photovoltaic Outputs," *IEEE J Photovolt*, vol. 10, no. 5, pp. 1417–1423, Sep. 2020.

[37] G. Chicco *et al*., "Error assessment of solar irradiance forecasts and AC power from energy conversion model in grid-Connected photovoltaic systems," *Energies*, vol. 9, no. 1, 2016.

[38] J.S. Stein, C.W. Hansen, and M.J. Reno, "Global horizontal irradiance clear sky models: implementation and analysis," Mar. 2012.

[39] V. Badescu *et al.*, "Computing global and diffuse solar hourly irradiation on clear sky. Review and testing of 54 models," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 3, pp. 1636–1656, Apr. 2012.

[40] A. Carullo *et al.*, "In-field monitoring of eight photovoltaic plants: Degradation rate over seven years of continuous operation," *Acta IMEKO*, vol. 7, no. 4, pp. 75–81, 2018.

[41] G. Alba *et al*., "Statistical Validation and Power Modelling of Hourly Profiles for a Large-Scale Photovoltaic Plant Portfolio," in *6th Intern. Forum on Research and Technology for Society and Industry, RTSI 2021 – Proc.*, 2021, pp. 18–23.

[42] J. Neyman, "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," *J R Stat Soc*, vol. 97, no. 4, p. 558, 1934.

Authorized licensed use limited to: Politecnico di Torino. Downloaded on February 19,2024 at 16:23:56 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.