Hybrid Modeling for Photovoltaic Module Operating Temperature Estimation

IEEE Journal of Photovoltaics

Santos, Leticia O.; Souza, Francisco A.A.; Carvalho Filho, Clodoaldo O.; Carvalho, Paulo C.M.; Alskaif, Tarek et al

https://doi.org/10.1109/JPHOTOV.2024.3372328

Hybrid Modeling for Photovoltaic Module Operating Temperature Estimation

Leticia O. Santos, *Graduate Student Member, IEEE*, Francisco A. A. Souza, Clodoaldo O. Carvalho Filho, Paulo C. M. Carvalho, Tarek AlSkaif, *Senior Member, IEEE*, and Renata I. S. Pereira

***Abstract*—The performance and efficiency of photovoltaic (PV) modules are significantly impacted by their operating temperature. Therefore, accurately estimating the PV module temperature (T_m) is a crucial factor in the assessment of PV systems. This article introduces a hybrid model for T_m estimation that combines both physical and data-driven modeling. The primary objective of our research is to enhance long-term T_m estimation, a domain where steady-state physical models are conventionally applied. Model parameters are extracted for poly-Si modules using Bayesian optimization. The adaptivity of our approach is validated using data from three distinct PV plants, each featuring different installation types and operating under different climatic conditions. To evaluate the effectiveness of our model, we compare its results with two widely used models for T_m estimation: the Sandia and Faiman models. The comparative analysis further confirms that our model provides more accurate T_m estimations. Our model shows a mean absolute error (MAE) of 2.44 °C, surpassing the 3.82 °C and 4.14 °C MAE values obtained using Faiman and Sandia models, respectively. The results suggest a superior T_m estimation even in scenarios of short-term irradiance variations. Model validation demonstrates its potential to improve the accuracy of PV conversion efficiency estimation by up to 1.05% compared with reference models.**

***Index Terms*—Dynamic thermal model, machine learning (ML), photovoltaic (PV) hybrid models.**

I. Introduction

PHOTOVOLTAIC (PV) efficiency and performance are intricately linked to the PV module operating temperature (T_m). As T_m increases, PV efficiency decreases, impacting electricity production [1]. For instance, crystalline silicon PV modules exhibit an efficiency reduction of approximately 0.5% for every 1 °C increase in the PV cell operating temperature [2]. Accurate T_m estimation is crucial for designing and operating PV systems, as inaccuracies can lead to overestimation or underestimation of PV output power, directly affecting financial outcomes of PV investments.

Three common approaches for modeling the PV operating temperature include physics-based, data-driven, and hybrid models. The industry commonly employs steady-state physical models for modeling T_m in new PV systems and for time horizons of a day ahead and longer [2]. However, these models, which do not account for transient climatic conditions, can deviate by up to 25 °C from actual conditions [3]. Physics-based transient models offer a more precise simulation of the thermal behavior of PV modules, with an accuracy improvement ranging between 1.2 °C and 2 °C compared with steady-state models [4]. One main advantage of dynamic models is their superior responsiveness under varying weather conditions, even in the face of abrupt changes in solar irradiance and wind speed. Various studies have explored different approaches to implementing transient models for T_m estimation [5]. These approaches include considering the influence of wind [2], accounting for the PV module layers [6], assessing T_m during irrigation cycles [7], and addressing modules equipped with Sun-tracking systems [8]. However, a constraint associated with physical models is the requirement for parameters that are either not readily available or difficult to measure [9]. Typically, this challenge is addressed by resorting to mathematical approximations, which can limit the model accuracy.

Data-driven models avoid the complexity of physical modeling, as they do not require information on PV modules and utilize data without considering the inherent mathematical relationships within input parameters. Various machine learning (ML) techniques have been employed for T_m estimation [10], [11], [12]. However, these models may struggle to properly interpret the physical interactions of the input parameters concerning T_m [13]. Many of these models use steady-state physical models as benchmarks [9], [14], leading to potentially unfair comparisons. Physical models are recommended for long-term estimations, whereas ML models tend to excel when trained with local data and for short-term forecasting. Furthermore, physical models distinguish between values of T_m, PV cell temperature, back surface module temperature, wind speed measurement, and the PV configuration, details sometimes neglected in certain ML-based works.

Manuscript received 11 December 2023; revised 5 February 2024; accepted 23 February 2024. Date of publication 21 March 2024; date of current version 19 April 2024. This work was supported in part by Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) and in part by *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq)-Brazil. *(Corresponding author: Leticia O. Santos.)*

Leticia O. Santos and Paulo C. M. Carvalho are with the Graduate Program of Electrical Engineering, Federal University of Ceará, Fortaleza 60020-181, Brazil (e-mail: leticia@fisica.ufc.br).

Francisco A. A. Souza is with OnePlanet Research Center, imec-NL, 6708 WE Wageningen, The Netherlands.

Clodoaldo O. Carvalho Filho is with the Department of Mechanical Engineering, Federal University of Ceará, Fortaleza 60020-181, Brazil.

Tarek AlSkaif is with the Information Technology Group, Wageningen University, 6708 PB Wageningen, The Netherlands.

Renata I. S. Pereira is with the Federal Institute of Alagoas - Campus Arapiraca, Maceio 57317-291, Brazil.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JPHOTOV.2024.3372328.

Digital Object Identifier 10.1109/JPHOTOV.2024.3372328

Hybrid[1] models combine both physics-based and ML modeling, presenting a promising strategy that harnesses the accuracy of data-driven models and the interpretability and generalization ability of physics-based models. In [15], a hybrid model is proposed, coupling a steady-state physics model with a radial basis function artificial neural network to estimate cell temperature. In [16], a hybrid stepwise linear regression is presented, enhancing the accuracy of a thermal-optical simulation to estimate T_m. In [17], T_m is modeled using an adaptive neural fuzzy inference system, considering variations in weather conditions. Additionally, a hybrid thermal model is developed in [3], extending steady-state models with an exponential smoothing kernel to include the heat capacity effect. This model, featuring optimizable parameters, enhances performance by nearly half compared with the steady-state model. Similarly, Prilliman et al. [4] introduced a transient thermal modeling approach, which involved weighting and averaging steady-state estimations from previous timesteps to estimate T_m at finer time scales. These hybrid models outperformed the physics-based models, demonstrating improved performance.

Although hybrid models for estimating PV temperature show potential advantages over traditional physics-based and purely data-driven models, current methodologies have limitations. These approaches often fall short of capturing the dynamic behavior of PV systems, merely offering incremental improvements to existing simplified models. To elaborate, the few physic-guided hybrid models for T_m estimation found in literature rely on simplified thermal expressions. As demonstrated in [18], these expressions can be expressed as linear equations representing T_m as a function of meteorological factors. The use of ML to determine the terms of a linear function often results in these hybrid models resembling linear regressions more than true hybrid models. Moreover, these methods are frequently validated within a constrained scope of data, commonly confined to single-site evaluations over limited periods, which does not adequately address seasonal variability. Therefore, there is a clear necessity for the development of more robust models capable of reliably describing T_m across diverse conditions and over extended timescales.

To address the aforementioned challenges, this study presents a novel hybrid model for long-term estimation of T_m. Our contributions are threefold. First, we build upon the works outlined in [3] and [4] by introducing a more comprehensive physics dynamic thermal PV model. This model is grounded in a series of expressions detailing the electro-thermal processes, considering specific module characteristics. Second, we employ Bayesian inference to estimate unknown parameters within the thermal model, thereby enhancing the precision of our estimations. Third, we validate our model by evaluating its performance across three distinct PV plants, each featuring different configurations and climatic conditions. The results demonstrate a strong model resilience to incident irradiance fluctuations, such as those resulting from cloud movement. The research outcomes indicate a step forward in accurate T_m modeling and estimation.

[1]The term hybrid models in this work refers to models that combine physics-based and ML modeling, although the term is often referred to in the literature as an ensemble of different ML models.

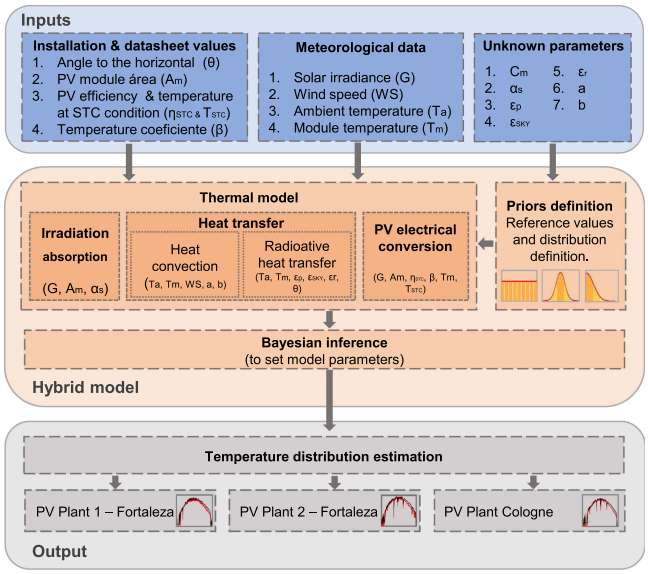

Fig. 1. Flowchart of the hybrid data-driven model training and test.

The rest of this article is organized as follows. Section II provides a detailed presentation of the dynamic thermal PV model. The utilization of Bayesian inference for determining the model's parameters, along with a description of the data and the adopted performance indices are given in Sections III and IV, respectively. Section V presents a comparison of the estimated results with the measurements and the benchmark models. Finally, Section V concludes this article.

II. Thermal Model

The methodology employed in this study begins with the description of the thermal model, outlined in this section. Further methodological steps are elucidated in Section III, encompassing the details of Bayesian inference, and in Section IV, covering both data preprocessing and postprocessing, as well as the model accuracy analysis used to assess its performance. An overview of the methods is provided in Fig. 1.

Since heat is dissipated from the PV module's front and back surfaces, a temperature gradient in the direction of the heat flow is observed inside the module. Consequently, the daytime back surface module temperature (T_b) is typically lower than the PV cell temperature (T_c) [19]. However, the discrepancies between T_c and T_b are on the order of 1 °C, well below the uncertainty level of around 5 °C associated with physical models for PV module temperature estimation [20]. One of the widely used approaches aiming for field applicability and accuracy is to consider the PV module as a single block of materials in thermal equilibrium. Since the PV module thickness is nearly 5 mm, it can be modeled as a concentrated component with the same temperature, T_m, for its entire extension.

We present a semiempirical 1-D transient model based on the energy balance of the PV system. The term transient, often

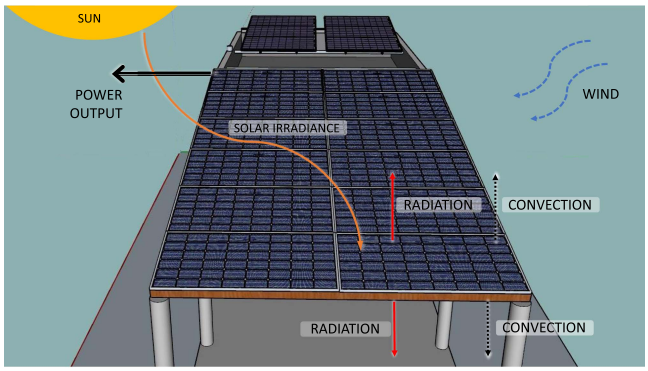

Fig. 2. Energy balance of a free-standing PV system.

overlooked in earlier studies, is incorporated in the energy balance equation, providing a better response to solar irradiance variations. Fig. 2 visually illustrates the PV energy balance. The proposed model is designed specifically for poly-Si modules and incorporates the following considerations.

1) The PV module operates under normal conditions, devoid of shading, dust, or any other agents affecting its absorptivity.
2) The PV module is treated as a block of materials with homogeneous temperature and average thermal characteristics.
3) The system operates under a time-dependent regime.
4) The ambient temperature (T_a) around the PV module is uniform and equal to the ground temperature.
5) The fraction of solar irradiance fraction absorbed by the PV cell, not converted into electricity is transferred through the following.

a) Radiation heat transfer between the PV module front and backside and the environment.
b) Natural and forced convection heat transfer between the PV module front and backside and the environment.
c) Conduction in the PV module layers is neglected, with layers properties considered in the thermal capacity (C_m).

The energy balance in the module, outlined in (1), leads to a differential equation with T_m on both sides, as the cooling because of convection and thermal emission, and the output power depend on T_m itself

$$C_m \frac{dT_m}{dt} = q_s - q_{\text{conv}} - q_{\text{rad}} - P_{\text{out}}. \tag{1}$$

Here, C_m represents the heat capacity of the PV module, $\frac{dT_m}{dt}$ denotes the time derivative of T_m, and q_s symbolizes the solar irradiance absorbed by the module, as expressed in the following:

$$q_s = GA_m\alpha_s. \tag{2}$$

Here, G denotes the incident solar irradiance, A_m represents the PV module area, and α_s stands for the solar absorptance. While some studies use a transmittance-absorptance product ($\tau\alpha$) to represent the module absorptance, it represents the effective absorptance of a thermal absorber in a cover system with multiple reflections [21]. However, within a PV module, the cover is part of the absorber, and the heat absorbed by all layers contributes to the module's heating. Therefore, employing only the absorptance nomenclature is deemed more appropriate [20].

The term q_{conv} in (1), as defined by (3), represents the convective heat transfer between the PV module and the surrounding air. Its primary tendency is to increase as T_m rises above T_a

$$q_{\text{conv}} = hA_m(T_m - T_a). \tag{3}$$

Here, h denotes the module convective heat loss coefficient, which encompasses both natural convection (h_{nat}) and forced convection (h_{forc}), the latter being linked to the local W_S. We assume the representation of h as a linear function of W_S (4), where the terms a and b are empirically determined [22]. The h_{nat} coefficient is assumed constant and is incorporated in b

$$h = aW_S + b. \tag{4}$$

The term q_{rad} in (1) represents the heat loss by radiation at the PV module's top and back-side, as expressed in the following:

$$q_{\text{rad}} = \sigma A_m(\epsilon_p(T_m + 273.15)^4 - F_{t,\text{sky}}\epsilon_{\text{sky}}T_{\text{sky}}^4 - F_{b,\text{roof}}\epsilon_r(T_a + 273.15)^4). \tag{5}$$

Here, σ represents the Stefan–Boltzmann constant, ϵ_p is the emissivity of the PV module, considering both front and back-side emissivities. Additionally, ϵ_{sky} and ϵ_r denote the emissivities of the sky and roof, respectively. $F_{t,\text{sky}}$ and $F_{b,\text{roof}}$ stand for view factors from the top of the PV module to the sky and from the module back-side to the roof, as expressed in (6) and (7) [21]. T_{sky} is the sky temperature (K), estimated according to (8) [23]

$$F_{t,\text{sky}} = \frac{1 + \cos\theta}{2} \tag{6}$$

$$F_{b,\text{roof}} = \frac{1 - \cos(\pi - \theta)}{2} \tag{7}$$

$$T_{\text{sky}} = 0.0552(T_a + 273.15)^{1.5}. \tag{8}$$

The angle θ in (6) and (7) represents the inclination of the module to the horizontal. Finally, P_{out} in (1) denotes the PV module output power, expressed by the following:

$$P_{\text{out}} = GA_m\eta. \tag{9}$$

Here, η represents the module efficiency, dependent on β (the temperature coefficient of the PV module), and module efficiency at standard test conditions (STC), as indicated in (10). T_{stc} is equivalent to 25 °C

$$\eta = \eta_{\text{stc}}[1 - \beta(T_m - T_{\text{stc}})]. \tag{10}$$

In this work, we adjust η based on the linear annual reduction specified by the manufacturer, as this parameter was determined for older PV modules operating at temperatures well above 25 °C.

III. Bayesian Inference of the Thermal Model Parameters

Knowing the values of the physical parameters is a crucial step in solving the thermal model. Some of these parameters, such as A_m, β, η_{stc}, and the view factors depend on θ, are well known and determined by module specifications, or installation configuration. On the other hand, values for α_s, ϵ_p, ϵ_{sky}, and ϵ_r are typically assumed from scientific literature, referencing, for instance, the physical properties of common glass for the module top. Consequently, uncertainties may arise in the model because of potential variations in the physical properties of the PV module materials compared with literature values.

Bayesian inference is a learning procedure that integrates data and explicit prior knowledge to infer the parameters of a model. Based on Bayes' rule, the learning process involves estimating the distribution of the unknown parameters (posterior) by considering the discrete thermal model, the prior probabilities reflecting the uncertainty of the unknown parameters, and the data

$$\text{posterior} \propto \text{likelihood} \times \text{prior}. \tag{11}$$

The fitting of the unknown parameters in the model, derived from (1), is accomplished through Bayesian Inference. The primary objective is to infer the optimal values for the unknown physical parameters: C_m, α_s, ϵ_p, ϵ_{sky}, ϵ_r, and the terms a and b from the convection coefficient, as expressed in (4). These depend on both the characteristics of the PV system and the prevailing local meteorological conditions. The likelihood is formulated based on (1), and the set of priors for the unknown parameters is derived from expert knowledge and the solar panel manufacturer's datasheet.

A. Discrete Model

To numerically solve the dynamic model outlined in (1), incorporating instantaneous values of G, T_a, and W_S, we employ the Euler scheme to discretize the time derivative. The current state is updated from the previous state based on dynamic equations using the data to account the inertia effect, with the initial condition of $T_m(t_0) = T_{m,0}$. For discrete time steps $t_n = n\Delta t + t_0$, where $dt \to \Delta t$, this results in the simplest discretization (explicit Euler scheme: approximation of y by a piecewise linear curve)

$$T_{m,n+1} = T_{m,n} + C_m^{-1}\Delta t(q_s - q_{\text{conv}} - q_{\text{rad}} - P_{\text{out}}). \tag{12}$$

It is important to note that the Euler method has limitations regarding the maximum allowable time step for convergence of the numerical solution. Conversely, the time step size should not be too small to prevent computational burden but must ensure good agreement with the training data. Applying the discrete Euler model requires the module temperature at the initial instant of operation. However, this temperature is not always a parameter measured in PV plants. Therefore, the sunrise condition is assumed as the initial condition, a period during which the PV module is in thermal equilibrium with the environment. Consequently, the ambient temperature becomes the initial value for T_m. The likelihood is then defined as $T_{m,n+1} \sim \text{normal}(T_{m,n} + C_m^{-1}\Delta t(q_s - q_{\text{conv}} - q_{\text{rad}} - P_{\text{out}}, \delta)$, where δ denotes the scale parameter of the normal distribution.

B. Bayesian Estimator and Priors

To estimate these uncertain parameters from the likelihood and priors, we employ Bayesian Inference with the No-U-Turn Sampler, a type of Markov chain Monte Carlo (MCMC) algorithm extension that requires no tuning parameters and can sample from distributions with correlated parameters. Thus, it is an example of a Bayesian inference application that enables working at the model level without implementing the MCMC algorithm itself [24]. This technique requires defining a prior probability distribution to represent the uncertainty of the parameter values being estimated. Two types of distribution are adopted for the priors: Gaussian (normal) distribution and uniform distribution. Reference values from the literature are used to represent the uncertainty of each unknown parameter, as defined as follows.

a) C_m*:* This parameter values typically falls within the range of 15 000 and 30 000 J/K. An optimum value of 22 280 J/K is proposed for crystalline silicon (cSi), polymer laminate and glass modules in [25]. Other studies point to best results with values between C_m 20 and 22 kJ/K [26], [27]. Considering these variations, we assume C_m to follow a Gaussian distribution around the optimum values proposed in previous research, $C_m \sim \text{normal}(25\,000, 10\,000)$.

b) α_s*:* The module absorptance, varies according to the spectral content and module materials [28]; however, α_s is usually assumed to be constant with an optimal value of 0.92 for a polycrystalline module [29]. Common values in literature range from 0.7 to 0.97 [27], [30], [31], [32], [33]. As there is no clear value for α_s, and these values can slightly vary according to the module material, we consider α_s to follow a Uniform distribution $\alpha_s \sim \text{uniform}(0.7, 0.97)$.

c) ϵ_p*:* The radiative heat capacity of the PV module is commonly assumed to be close to 0.9 [34]. Similarly to the parameter α_s, we found no clearly predominating value for ϵ_p and it varies according to the module material. Therefore, we consider ϵ_p to follow a uniform distribution $\epsilon_p \sim \text{uniform}(0.85, 0.98)$.

d) ϵ_{sky}*:* Common values for ϵ_{sky} range between 0.95 and 1, with variations depending on the sky condition [25]. We consider ϵ_{sky} to follow a uniform distribution $\epsilon_{\text{sky}} \sim \text{uniform}(0.85, 1)$.

e) ϵ_r*:* The ϵ_r is influenced by the roof cover material; in this case concrete. We consider ϵ_r to follow a uniform distribution $\epsilon_r \sim \text{uniform}(0.60, 0.90)$.

f) a, b*:* Parameters of the linear model representing wind influence (a, b) vary widely in the literature. Furthermore, wind direction incidence influences convective heat loss, even with the same wind velocity, making these parameters difficult to define. As a result, we define a vague prior for these parameters, following a half-normal distribution with a criterion greater than zero: $a, b \sim \text{half-normal}(0, 1)$.

Table I summarizes the constant and unknown parameters adopted for fitting the thermal model.

(a) (b) (c)

Fig. 3. Three PV plants under consideration in this study. (a) Fortaleza PV Plant 1: Free-standing mounted. (b) Fortaleza PV Plant 2: Open rack roof-mounted. (c) Cologne PV Plant: Open rack ground-mounted.

TABLE I
DEFAULT PARAMETER VALUES OF THE THERMAL MODEL

	Type	Reference values	Distribution
A_m	Constant	$0.94 \times 1.9\,\mathrm{m}^2$	–
η_{STC}	Constant	17%	–
β	Constant	–0.42%/°C	–
θ	Constant	10°	–
σ	Constant	$5.67 \times 10^{-8}\,\mathrm{Wm}^{-2}\mathrm{K}^{-4}$	–
C_m	Unknown parameter	$25\,000 \pm 10\,000$ J/K	Gaussian
α_s	Unknown parameter	0.7–0.97	Uniform
ϵ_p	Unknown parameter	0.85–0.98	Uniform
ϵ_{sky}	Unknown parameter	0.85–1	Uniform
ϵ_r	Unknown parameter	0.60–0.90	Uniform
a	Unknown parameter	$a > 0$	Half-Gaussian
b	Unknown parameter	$b > 0$	Half-Gaussian

IV. DATA AND EXPERIMENTAL SETTINGS

A. Training Data

The training data employed for fitting the thermal model originated from a monitoring campaign conducted at a PV plant in the Northeast Region of Brazil (3.74 S, 38.57 W). The intrahour dataset includes solar irradiance (W/m^2) both horizontal and in the plane of the PV array, wind speed (m/s) at the PV array level, and ambient and PV module temperature (°C), recorded at one-minute intervals. This data was obtained by an IoT-embedded system developed in a prior research [35]. The PV plant, referred to as Plant 1[2] of LEA-UFC, consists of 12 polycrystalline silicon 330Wp modules mounted on an open rack inclined at 10° facing north [see Fig. 3(a)]. Considering the transient behavior of the module and the need to account for the prior state, days with missing data exceeding two consecutive hours are removed from the dataset to ensure the continuity of time series, and prevent temperature analysis disruption. Because of the physics of the analyzed problem, all the data was confined to daylight hours.

B. Thermal Model Evaluation

To evaluate the thermal model's accuracy, we utilize new data from the same PV plant used for parameter calibration. Additionally, to gauge the model's generalizability, we incorporated data from two distinct polycrystalline silicon PV plants. The first, LEA-UFC's Plant 2, is situated in the same city as the calibration site but features a different installation type, being roof-mounted. The model is tested using one-year data, encompassing both typical climates in Fortaleza, including a rainy season from February to May and a dry period in the second semester of the year. The second, Cologne PV Plant, is a ground-mounted PV system located in Cologne, Germany, characterized by climate conditions distinct from the model's training environment. Refer to Fig. 3 for visual representations of the site installations. The collected data per case study used in this work, for training and testing purposes, is summarized in Table II.

C. State of Art Models

The thermal model results are compared with two well-established models in the literature: the Sandia model [36] and the Faiman model [37]. These models are widely utilized for T_m estimation in both industry and scientific applications. The Sandia model defines T_m as

$$T_m^{\text{Sandia}} = G(e^{a+bW_s}) + T_a \tag{13}$$

where $a = -3.56$ and $b = -.0750$ for Glass/cell/polymer sheet modules with open rack mounting.

The Faiman model describes T_m as

$$T_m^{\text{Faiman}} = T_a + \frac{G}{U_o + U_1 W_s} \tag{14}$$

where $U_0 = 25\,W/m^2K$ and $U_1 = 6.84\,W/m^3sK$.

The evaluation metrics used to assess the models' performance include the mean absolute error (MAE), rooted-mean-square error (RMSE), and the R-squared score.

V. EXPERIMENTAL RESULTS

In this section, we analyze the proposed model for T_m estimation, considering the estimation of model parameters and its applicability to different PV installation types across different sites.

[2]The experiments took place at the *Laboratório de Energias Alternativas* (LEA) of the Federal University of Ceará (UFC).

TABLE II
SUMMARY OF THE DATA

PV Plant	Period	G (W/m^2)	W_s (m/s)	T_a (°C)	T_m (°C)	Samples
Fortaleza 1	1 Dec. 2022–8 Nov. 2023	0–1 036.75	0–11.5	22–38.9	21.12–70.1	103 543 (30 000 train, 73 543 test)
Fortaleza 2	1 Jan. 2019–1 Jan. 2020	0–1 561.70	0–9.70	22–45.0	25.30–71.0	172 108
Cologne	1 Jul. 2018–31 Dec. 2018	0–1092.0	0–44.0	9.2–39.4	10.9–66.1	80 201

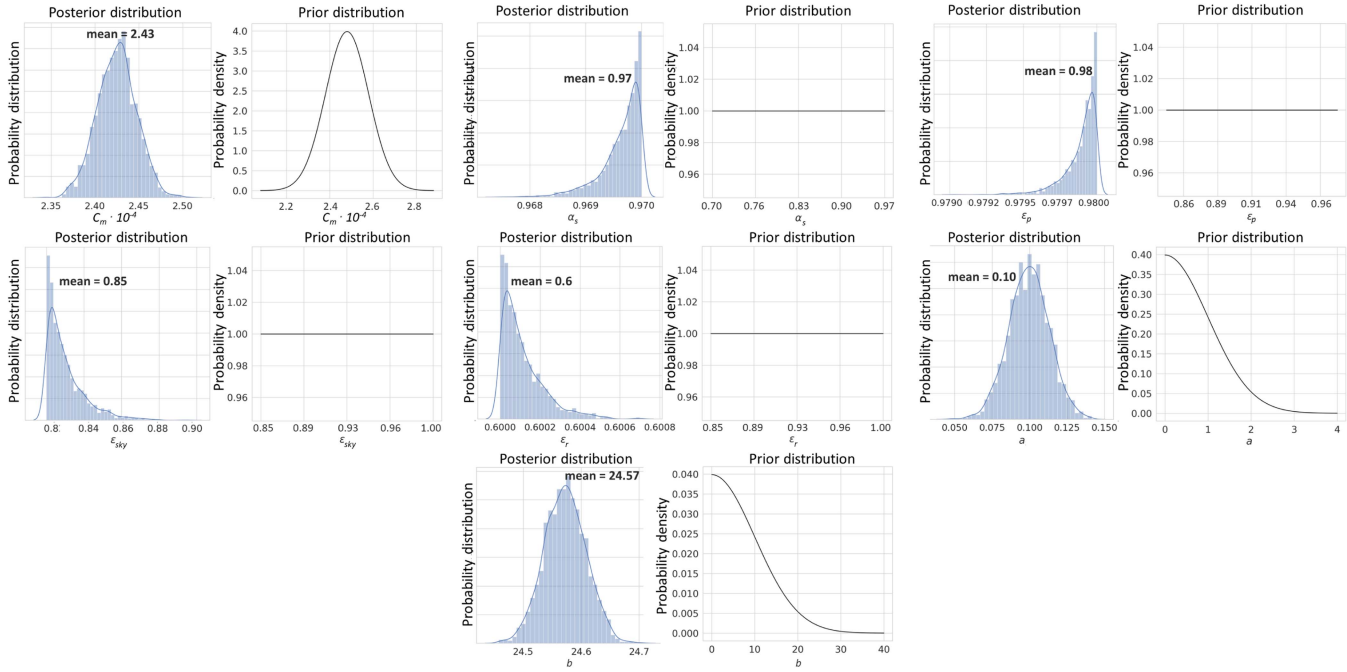

Fig. 4. Posterior and prior distributions of the thermal model optimized parameters.

A. Evaluation of the Optimized Parameters

After fitting the thermal model using Bayesian inference with the training data described in Section IV, the posterior of the unknown parameters, including α_s, C_m, the emissivities ϵ_p, ϵ_{sky} and ϵ_r, as well as the h_{forc} terms a and b, are identified. The summary and the associated distributions are depicted in Fig. 4. The mean values obtained by Bayesian optimization are as follows: α_s value of 0.97; C_m equal to 24 250.98 J/K; ϵ_p, ϵ_{sky}, and ϵ_r of 0.98, 0.85, and 0.60, respectively; a and b values of 0.10 and 24.57, respectively.

From the analysis of the energy balance optimized parameters, the following can be concluded.

1) Except for a slightly higher value of α_s compared with common thermal models values, all physical parameters (α_s, C_m, ϵ_p, ϵ_{sky}, and ϵ_r) are within ranges found in literature and summarized in Table I.
2) a and b values lead to $h_{\text{forc}} = 0.1W_s + 24.57$. Given that W_s values in PV Plant 1 surroundings are usually lower than 3 m/s, it indicates a low influence of the wind compared with b, which remains constant.

B. PV Module Temperature Estimation

To assess the performance of the thermal models, the mean parameter values are used to compose (1). The model is initially evaluated on the test data of LEA 1 Plant, and the results are presented in Fig. 5(a). This figure displays a snapshot of seven days of T_m model estimation alongside the reference values measured by the PV module temperature sensor. The model demonstrates its capability to estimate T_m in minute scales, even in the case of fast irradiation fluctuations caused by clouds. In comparison with reference models in Fig. 5(b), our proposed model exhibits a superior response to swift fluctuations of T_m, suggesting its potential suitability for hourly and intrahour forecasting.

A snapshot of seven days' results for the second test site, the LEA 2 Plant, is depicted in Fig. 6(a). A behavior similar to that of LEA 1 is observed, where the thermal model captures the variance and tendency of the reference data at minute scales, despite the training data being from a PV plant with a different installation type. Upon comparison with the reference models in Fig. 6(b), it becomes evident that the proposed model responds better to rapid fluctuations in T_m.

In both plants in Fortaleza, the estimated values from the proposed model closely align with the measured ones when compared with benchmark models. As previously shown in [38, Ch. 2], some well-established PV temperature models do not perform well for T_m estimations in Fortaleza, given the city's proximity to the equator line and its location in the intertropical convergence zone (ITCZ) with characteristics such as clouds,

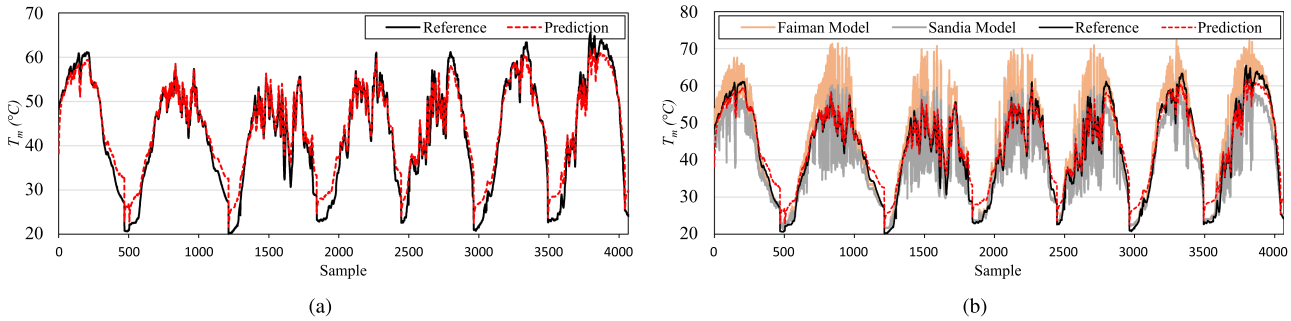

Fig. 5. Performance evaluation of the proposed hybrid model in LEA 1 PV plant: (a) against empirical data; (b) benchmarked with the reference models.

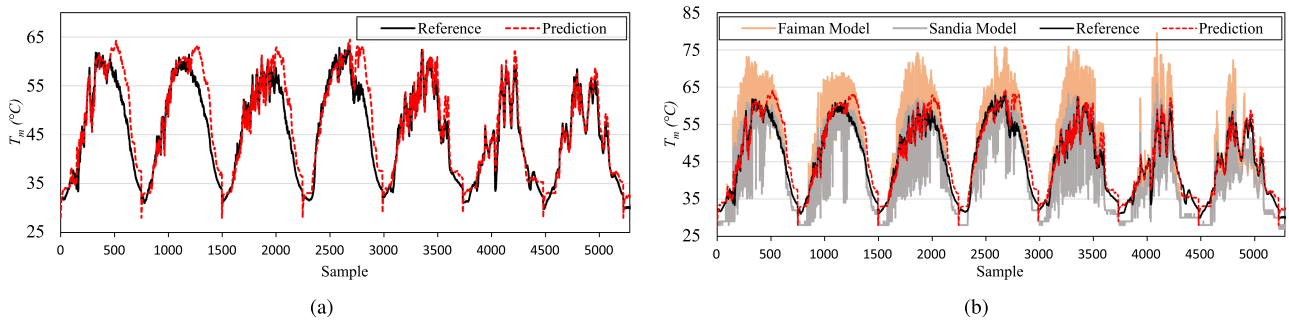

Fig. 6. Performance evaluation of the proposed hybrid model in LEA 2 PV plant: (a) against empirical data; (b) benchmarked with the reference models.

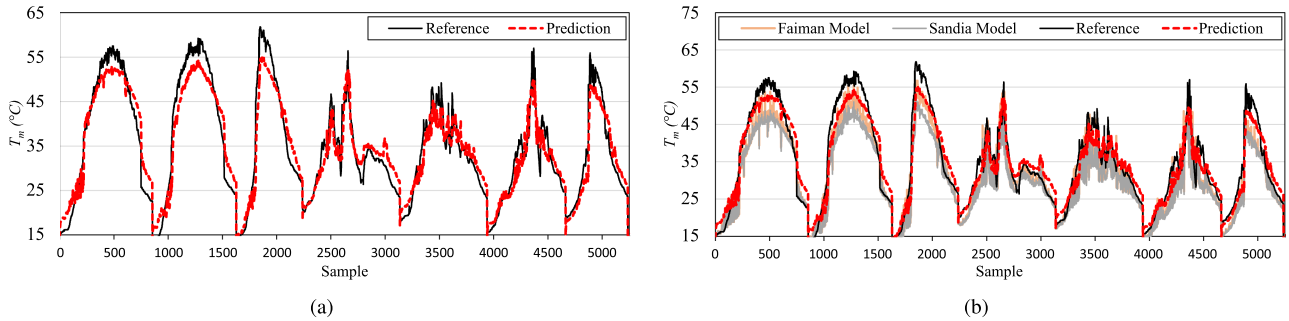

Fig. 7. Performance evaluation of the proposed hybrid model in Cologne PV plant: (a) against empirical data; (b) benchmarked with the reference models.

trade winds, and rapid fluctuations in environmental parameters. The proposed model, as demonstrated by Figs. 5 and 6, effectively captures the internal behaviors under fluctuating environments, with only minor deviations observed for short periods. Furthermore, in Figs. 5(b) and 6(b), it is apparent that, when compared with the reference values, the estimations from the Faiman model apparently overestimate T_m, while those from the Sandia model seem to underestimate it in Fortaleza. This behavior may be attributed to the model expressions, as W_s appears in the exponential that multiplies G in the Sandia model, whereas in the Faiman model, it appears in the denominator of G, opposite algebraic effects for the same parameter.

To assess the model's generalization ability for a new PV plant, we test it with data from the Cologne PV Plant, which features distinct climate conditions. In Fig. 7(b), seven days of the T_m models estimation are compared with reference values. By comparing the proposed model with reference models in Fig. 7(b), it is evident that the Sandia and Faiman models in Cologne present less variance around the reference values on days with no rapid irradiance variation. Consequently, steady-state model results are closer to the curve of the proposed model.

The error metrics between estimated and measured T_m values in the three analyzed PV plants are presented in Table III. The proposed model exhibits the smallest MAE when compared with the average temperature measurements of the three PV plants: 2.44 °C, 2.66 °C, and 2.85 °C for PV plants 1 and 2 in Fortaleza and Cologne PV plant, respectively. Compared with the Faiman model, our model shows an MAE improvement of 41.1% at Fortaleza PV plant 1, 49.7% at Fortaleza PV plant 2, and 39.1% in Cologne, respectively. When compared with the

TABLE III
EVALUATION OF ACTUAL AND ESTIMATED T_m IN TERMS OF MAE, RMSE, AND R^2 ON DATA OF PLANTS 1 AND 2 IN FORTALEZA AND COLOGNE

Model	MAE (°C)			RMSE (°C)			R^2		
	PV Plant 1 Fortaleza	PV Plant 2 Fortaleza	Cologne	PV Plant 1 Fortaleza	PV Plant 2 Fortaleza	Cologne	PV Plant 1 Fortaleza	PV Plant 2 Fortaleza	Cologne
Proposed	**2.44**	**2.66**	**2.85**	**3.17**	**3.35**	**3.53**	**0.97**	**0.91**	0.93
Faiman	3.82	5.78	5.31	5.07	7.38	6.00	0.89	0.78	**0.94**
Sandia	4.14	5.29	4.68	5.78	7.06	5.13	0.86	0.74	**0.94**

The best metrics in each location are in bold.

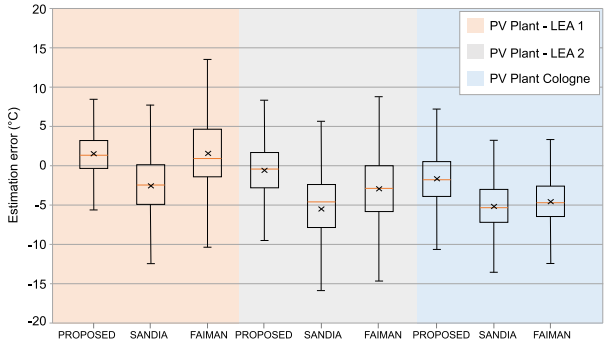

Fig. 8. Estimation error of each model for PV plant.

Sandia model, our model demonstrates an MAE improvement of 36.1% at Fortaleza PV plant 1, 54.0% at Fortaleza PV plant 2, and 46.3% in Cologne. The same pattern is observed in terms of RMSE for the three PV plants, with values of 3.17 °C, 3.35 °C, and 3.53 °C for PV plants 1 and 2 in Fortaleza, and Cologne's PV plant, respectively. In terms of R^2, the proposed model archieves a value of 0.97, 0.91, and 0.93 for PV plants 1 and 2 in Fortaleza, and Cologne's PV plant, respectively. In Fig. 8 is shown a boxplot of the estimation error in the mentioned PV plants: In all three locations, the error metric using the proposed model is closest to zero compared with Faiman and Sandia models. Moreover, the proposed model demonstrates a reduction in variance, especially in PV Plant 1 in Fortaleza, where the data was used to train the model. The majority of the model temperature estimations fall within 5 °C. Considering the uncertainty of the benchmark models, these results are within a satisfactory range to be trusted in energy performance calculations.

Utilizing the improvements in MAE of the proposed model, as shown in Table III, and assuming a temperature coefficient for power production of $-0.4\%/°C$, leads to estimated energy accuracy improvements ranging from 0.55% for a module in PV plant 1 in Fortaleza, 0.73% for the Cologne PV plant, to 1.05% in PV Plant 2 in Fortaleza. This result is an improvement compared with the hybrid model provided by Prilliman et al. [4]. Performance gains of this scale can have a significant impact on the design of PV systems.

This analysis underscores the usefulness of the proposed model, enhancing the accuracy of existing steady-state models without requiring complex empirical analysis or black box models. It is important to note that the proposed model was tested using a large dataset compared with the training data (i.e., 30 000 instances of only one plant) and demonstrated good performance and ability to estimate T_m under varying environmental and operating conditions, affirming the model's effectiveness in characterizing the PV module's internal behavior. This outcome highlights the benefits of employing hybrid models. By leveraging the universality of physics, these models demonstrate good generalization ability and require a low amount of training data. On the other hand, the data-driven approach enables the model to incorporate effects, such as weather patterns, that purely physics models might not account for.

VI. CONCLUSION

This article introduces a hybrid model for long-term T_m estimation, which considers the PV module dynamics and enhances the accuracy of thermal steady-state models. The model relies on assessing PV temperature through a physical model that considers both module characteristics and heat exchange. In this model, environmental parameters and the PV module installation configuration serve as inputs for T_m estimation. The proposed hybrid model relies on a data-driven approach to capture weather patterns and mitigate uncertainty associated with physical parameter values, which are determined using Bayesian inference.

The adaptivity and robustness of the proposed approach are validated using data from three PV plants: the same PV plant (LEA 1 plant) used for model training, a PV plant at the same site but with a different installation type, and a PV plant with a different weather conditions. The proposed model demonstrates higher accuracy compared with well-known steady-state models, commonly applied for long-term estimations, even under different weather conditions and installation configurations. For the LEA 1 plant, the model shows an MAE of 2.44 °C, outperforming the Faiman and Sandia models with MAEs of 3.82 °C and 4.14 °C, respectively. Despite employing a data-driven approach, the model remains fundamentally physical and is suitable for estimating PV generation throughout a project's lifetime. The proposed model is robust, grounded in a solid physical theoretical basis, and offers high accuracy and applicability.

Future work will focus on refining the physical-based model by exploring different considerations, while elaborating the energy balance and its mathematical solution. The Bayesian optimization adopted in this work was trained on a dataset from one PV plant and one PV technology, then applied to different plants with the same technology but different types of installation and weather conditions. Although the model showed a good regional generalization ability in the obtained results, incorporating data from various PV plants and different PV technologies into the training set could further enhance its generalization. Furthermore, as the model parameters differ according

to the module's material, future works may explore providing different coefficients per module and/or installation type. Despite its focus on long-term estimations, the model has shown good performance in short time intervals, warranting further studies to explore its application across different time horizons. Lastly, the proposed model can be coupled with other models in physical model chains to calculate the power output, providing an assessment of the temperature impact on PV power output.

REFERENCES

[1] X.-J. Dong, J.-N. Shen, Z.-F. Ma, and Y.-J. He, "Simultaneous operating temperature and output power prediction method for photovoltaic modules," *Energy*, vol. 260, 2022, Art. no. 124909.

[2] D. P. N. Nguyen, K. Neyts, and J. Lauwaert, "Proposed models to improve predicting the operating temperature of different photovoltaic module technologies under various climatic conditions," *Appl. Sci.*, vol. 11, no. 15, 2021, Art. no. 7064.

[3] J. Barry et al., "Dynamic model of photovoltaic module temperature as a function of atmospheric conditions," *Adv. Sci. Res.*, vol. 17, pp. 165–173, 2020.

[4] M. Prilliman, J. S. Stein, D. Riley, and G. Tamizhmani, "Transient weighted moving-average model of photovoltaic module back-surface temperature," *IEEE J. Photovolt.*, vol. 10, no. 4, pp. 1053–1060, Jul. 2020.

[5] D. D. Silva et al., "A new predictive model for a photovoltaic module's surface temperature," *Energy Rep.*, vol. 8, pp. 15206–15220, 2022.

[6] C. Li, S. V. Spataru, K. Zhang, Y. Yang, and H. Wei, "A multi-state dynamic thermal model for accurate photovoltaic cell temperature estimation," *IEEE J. Photovolt.*, vol. 10, no. 5, pp. 1465–1473, Sep. 2020.

[7] G. Osma-Pinto and G. Ordóñez-Plata, "Dynamic thermal modelling for the prediction of the operating temperature of a PV panel with an integrated cooling system," *Renewable Energy*, vol. 152, pp. 1041–1054, 2020.

[8] E. Kaplani and S. Kaplanis, "Dynamic electro-thermal PV temperature and power output prediction model for any PV geometries in free-standing and BIPV systems operating under any environmental conditions," *Energies*, vol. 13, no. 18, 2020, Art. no. 4743.

[9] H. Zhu et al., "Online modelling and calculation for operating temperature of silicon-based PV modules based on BP-ANN," *Int. J. Photoenergy*, vol. 2017, 2017, Art. no. 6759295. [Online]. Available: https://doi.org/10.1155/2017/6759295

[10] H. Amiry et al., "Assessment of improved models for predicting PV module temperature and their electrical performance in a semi-arid coastal region," *Int. J. Green Energy*, vol. 20, pp. 1–13, 2023.

[11] S. I. Sulaiman, N. Z. Zainol, Z. Othman, and H. Zainuddin, "Modeling of operating photovoltaic module temperature using hybrid cuckoo and artificial neural network," in *Proc. Pacific Rim Knowl. Acquisition Workshop*, 2014, pp. 29–37.

[12] Y. Sun et al., "Research on short-term module temperature prediction model based on BP neural network for photovoltaic power forecasting," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2015, pp. 1–5.

[13] O. M. Tzuc, A. Bassam, P. Mendez-Monroy, and I. S. Dominguez, "Estimation of the operating temperature of photovoltaic modules using artificial intelligence techniques and global sensitivity analysis: A comparative approach," *J. Renewable Sustain. Energy*, vol. 10, no. 3, 2018, Art. no. 033503.

[14] N. Bailek, K. Bouchouicha, M. A. Hassan, A. Slimani, and B. Jamil, "Implicit regression-based correlations to predict the back temperature of PV modules in the arid region of South Algeria," *Renewable Energy*, vol. 156, pp. 57–67, 2020.

[15] X.-J. Dong, J.-N. Shen, G.-X. He, Z.-F. Ma, and Y.-J. He, "A general radial basis function neural network assisted hybrid modeling method for photovoltaic cell operating temperature prediction," *Energy*, vol. 234, 2021, Art. no. 121212.

[16] A. Gholami et al., "Impact of harsh weather conditions on solar photovoltaic cell temperature: Experimental analysis and thermal-optical modeling," *Sol. Energy*, vol. 252, pp. 176–194, 2023.

[17] C. Babu and P. Pathipooranam, "PV module temperature estimation by using ANFIS," in *Proc. Soft Comput. Problem Solving*, 2020, pp. 311–318.

[18] L. d. O. Santos, P. C. M. d. Carvalho, and C. d. O. C. Filho, "Photovoltaic cell operating temperature models: A review of correlations and parameters," *IEEE J. Photovolt.*, vol. 12, no. 1, pp. 179–190, Jan. 2022.

[19] Y. Du, W. Tao, Y. Liu, J. Jiang, and H. Huang, "Heat transfer modeling and temperature experiments of crystalline silicon photovoltaic modules," *Sol. Energy*, vol. 146, pp. 257–263, 2017.

[20] A. Driesse, M. Theristis, and J. S. Stein, "PV module operating temperature model equivalence and parameter translation," in *Proc. IEEE 49th Photovoltaic Specialists Conf.*, 2022, pp. 0172–0177.

[21] J. A. Duffie and W. A. Beckman, *Solar Engineering of Thermal Processes*. Hoboken, NJ, USA: Wiley, 2013.

[22] W. Hu et al., "Experimental research on the convective heat transfer coefficient of photovoltaic panel," *Renewable Energy*, vol. 185, pp. 820–826, 2022.

[23] M. Mohanraj, P. Chandramohan, M. Sakthivel, and S. Kamaruzzaman, "Performance of photovoltaic water pumping systems under the influence of panel cooling," *Renewable Energy Focus*, vol. 31, pp. 31–44, 2019.

[24] J. Annis, B. J. Miller, and T. J. Palmeri, "Bayesian inference with Stan: A tutorial on adding custom distributions," *Behav. Res. Methods*, vol. 49, pp. 863–886, 2017.

[25] A. Luketa-Hanlin and J. Stein, "Improvement and validation of a transient model to predict photovoltaic module temperature," Sandia Nat. Lab., Albuquerque, NM, USA, Tech. Rep. SAND2012-4307C, 2012.

[26] X. C. Ngo, N. Y. Do, and Q. V. Dang, "Modeling and experimental studies on water spray cooler for commercial photovoltaic modules," *Int. J. Renewable Energy Develop.*, vol. 11, no. 4, pp. 926–935, 2022.

[27] D. T. Lobera and S. Valkealahti, "Dynamic thermal model of solar PV systems under varying climatic conditions," *Sol. Energy*, vol. 93, pp. 183–194, 2013.

[28] J. L. Bryan, T. J. Silverman, M. G. Deceglie, and Z. C. Holman, "Thermal model to quantify the impact of sub-bandgap reflectance on operating temperature of fielded PV modules," *Sol. Energy*, vol. 220, pp. 246–250, 2021.

[29] Y. Yin, H. Zhang, Z. Qin, and S. Wang, "Mathematical model establishment and in-situ experimental verification about the effects of meteorological environmental factors on the temperature of photovoltaic module," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 186, no. 3, 2018, Art. no. 012038.

[30] K. J. Ali, A. H. Mohammad, and G. T. Hasan, "An empirical correlation of ambient temperature impact on PV module considering natural convection," *Indonesian J. Elect. Eng. Comput. Sci.*, vol. 19, no. 2, pp. 627–634, 2020.

[31] F. Rahmaniah, W. Zhang, and S. E. Tay, "State space transient model for photovoltaic module temperature estimation," in *Proc. 47th IEEE Photovoltaic Specialists Conf.*, 2020, pp. 2505–2508.

[32] N. Dabaghzadeh and M. Eslami, "Temperature distribution in a photovoltaic module at various mounting and wind conditions: A complete CFD modeling," *J. Renewable Sustain. Energy*, vol. 11, no. 5, 2019, Art. no. 053503.

[33] S. Jacques, A. Caldeira, Z. Ren, A. Schellmanns, and N. Batut, "Impact of the cell temperature on the energy efficiency of a single glass PV module: Thermal modeling in steady-state and validation by experimental data," *Renewable Energy Power Qual. J.*, vol. 11, pp. 1–4, 2013.

[34] A. Jones and C. Underwood, "A thermal model for photovoltaic systems," *Sol. Energy*, vol. 70, no. 4, pp. 349–359, 2001.

[35] R. I. Pereira, S. C. Jucá, and P. C. Carvalho, "IoT embedded systems network and sensors signal conditioning applied to decentralized photovoltaic plants," *Measurement*, vol. 142, pp. 195–212, 2019.

[36] D. L. King, W. E. Boyson, and J. A. Kratochvil, "Photovoltaic array performance model," Sandia Nat. Lab. (SNL), Albuquerque, NM, Livermore, CA, USA, Rep. SAND2004-3535, Aug. 2004. [Online]. Available: https://doi.org/10.2172/919131

[37] D. Faiman, "Assessing the outdoor operating temperature of photovoltaic modules," *Prog. Photovolt.: Res. Appl.*, vol. 16, no. 4, pp. 307–315, 2008.

[38] P. J. P. Abdala, *Energia Solar e Eólica*, Ponta Grossa, Brazil: Atena Editora, 2019, vol. 1.