Electric Power Systems Research 264 (2027) 113799



Contents lists available at ScienceDirect

Electric Power Systems Research

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



Proposal for a performance evaluation method for grid-connected photovoltaic systems based on machine learning forecasting models ⋆

Flávio V. Leão [a,b], Renato C. Sampaio [a], Loana N. Velasco [a], Alex Reis [a], Ênio C. Resende [b], Luiz C. G. Freitas [b,*]

[a] University of Brasília, Faculty of Science and Technology in Engineering, Brasília, Brazil
[b] Federal University of Uberlândia, Faculty of Electrical Engineering, Uberlândia, Brazil

ARTICLE INFO

Keywords:
Distributed generation
Diagnosis
Energy management
Forecasting
Machine learning
Performance ratio
Photovoltaic energy

ABSTRACT

In this work, the authors present the results of a study conducted at the University of Brasília (UnB), Brazil, which had as its main objective to evaluate and compare daily forecast models for electricity generation in photovoltaic (PV) plants. The proposed solution is based on an infrastructure for measuring electrical quantities and developing supervisory software, enabling continuous operational monitoring of the University's buildings. The model evaluation employs six standardized metrics for solar energy forecasting, following NREL guidelines and best international practices. These metrics enable objective selection of the most suitable algorithms for operational photovoltaic plants applications, providing a comprehensive assessment of both accuracy and reliability. The top three exogenous models provided nMAE (Normalized Mean Absolute Error) around 4.91-5.10%, which significantly outperform the three best temporal models, which provided nMAE around 14.16-14.27%, with an improvement of about 65% in forecast error. Even more critical, the coefficient of determination (R2) rose from 0.12 to 0.90, demonstrating that physical variables encode the essential mechanisms of PV energy generation that temporal variables cannot adequately represent. These models represent the three most effective features of engineering strategies. Only models that presented nMAE below 5.2% and R2 above 0.89 were chosen, ensuring high forecasting accuracy for reliable operational anomaly detection. An anomaly detector based on the Energy Performance Index (EPI), framework defined in IEC TS 61724-3, was developed using a machine learning model and used as the expected performance reference, which normalizes the operational analysis by removing the effect of weather conditions. The exceptional accuracy of the selected models (nMAE 5%, $R^2 > 0.89$) shows that the proposed solution has great potential for application in Energy Management Systems (EMS), as it enables the timely implementation of preventive and corrective maintenance measures, avoiding financial losses.

1. Introduction

In recent decades, electric power systems have undergone major changes in the paradigms of design, operation, and maintenance of facilities, as well as process modernization. These are associated with various technological advances and concerns about climate change, energy efficiency and sustainability, and supply security, among other aspects. In this sense, the accelerated increase in the demand for electricity worldwide, associated with the need to reduce dependence on fossil fuels, has driven the search for renewable energy sources and the implementation of Energy Efficiency Actions (EEA) [1].

Therefore, Energy Management emerges as a continuous and structured process to monitor the consumption of electricity in a facility, always monitoring its variations and observing opportunities for the implementation of EEAs. These encompass a wide range of measures that depend on the end use of energy and can cover, for example, retrofitting of facilities, use of equipment with lower energy losses, monitoring systems, building labeling programs, user awareness campaigns, as well as the installation of electricity generation systems based on renewable energy sources, among which photovoltaic and wind energy stand out [2].

With regard to the use of photovoltaic energy as EEA, it should be noted that it is necessary to monitor the performance of photovoltaic plants to ensure that the investments made can actually positively impact the energy efficiency of a given facility. Thus, energy management platforms must incorporate tools capable of monitoring and evaluating

⋆ This research was funded in part by Universidade de Brasília (UnB), and in part by the Ministério da Educação (MEC)-Grant SIMEC-ted-pt-13285, and Fundação de Amparo à Pesquisa do Distrito Federal (FAPDF)-Grant SEI-GDF: 00193–00002423/2023-11.

* Corresponding author.
E-mail addresses: flavio.vl@ufu.br (F.V. Leão), lcgfreitas@ufu.br (L.C.G. Freitas).

https://doi.org/10.1016/j.epsr.2026.113799
Received 8 October 2025; Received in revised form 13 March 2026; Accepted 6 July 2026
Available online 28 July 2026
0378-7796/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the performance of these systems, ensuring that predictive and corrective maintenance measures can be executed in a timely manner [3,4].

In this context, in this work, the authors present the results of a study carried out at the University of Brasília (UnB), which had as its main objective to evaluate and compare daily forecast models for electricity generation in photovoltaic (PV) plants. For this monitoring to be effective, it is essential to have reliable estimates of expected electricity generation in [5] and [6], the authors present comprehensive reviews of methods for forecasting electricity generation in PV plants, categorizing statistical, machine learning, and physical approaches. In [7], the authors compared 24 ML algorithms for PV energy generation forecasting based on numerical weather prediction, demonstrating that ensemble models such as gradient boosting have superior performance on daily horizons. Despite the advances, most studies focus on the accuracy of the forecast itself, without exploring its direct application as a reference for operational anomaly detection.

Concerning monitoring and fault detection in PV systems, in [8,9], the authors present comprehensive reviews of fault types and detection techniques, including monitoring systems. In [10], the authors demonstrated advances in fault diagnosis based on performance data analysis for grid-connected systems. In [11], neural network applications for anomaly detection and predictive maintenance are analyzed, achieving detection rates exceeding 90%. In [12], the authors reviewed model-based methods for fault detection, while in [13] the authors proposed peer-to-peer monitoring as an alternative when meteorological data is unavailable. The IEA PVPS Task 13 [14] established a statistical performance monitoring framework, and in [15] the authors demonstrated performance ratio classification by ML. Unlike most studies, which focus exclusively on forecasting PV electricity generation, this work also uses forecasting models as a reference to evaluate the operational performance of the system at the end of each day, enabling early fault detection.

To present the results obtained, this article is organized as follows: Section II describes the photovoltaic plant under analysis. Section III presents the selection procedures and performance analysis of forecasting algorithms. Section IV presents the practical results obtained, and to conclude, Section V presents the results obtained during the analysis of a photovoltaic plant installed at the University of Brasília. Finally, Section VI presents the final conclusions of the work and proposals for future work.

2. System description

Fig. 1 shows the Gama Campus of the University of Brasília, Brazil, where the photovoltaic system is installed and connected to the distribution network of Neoenergia Brasília through an existing substation on the campus. The PV system under analysis, installed on the roof of the (1) Teaching Unit (UED-acronym in Portuguese). Photovoltaic systems installed in the (2) Academic Unit (UAC-acronym in Portuguese), (3) Sports and Equipment Services Module (MESP-acronym in Portuguese) and the (4) Laboratory for Transport Development and Alternative Energies (LDTEA-acronym in Portuguese) are beyond the scope of this work.

The UED PV system consists of 298 photovoltaic modules from the manufacturer Canadian Solar, HiKu CS3W-420P model, with a peak power of 420 Wp per module, positioned with a 16º inclination and oriented toward geographic north. The photovoltaic modules are connected to two three-phase string inverters from the manufacturer Canadian Solar, one being the CSI-75K-T400 model with 75 kW and the other the CSI-50KTL-GI model with 50 kW.

2.1. Data collection about the electricity generated

Each inverter is connected to a dedicated Kron Konect IoT energy meter for monitoring electrical parameters. The meters interface with the University's Open Platform Energy Monitoring (MEPA - acronym in Portuguese) system using the Modbus TCP/IP protocol. Data capture occurs via the MEPA API, providing a temporal resolution of 15 minutes. The monitoring system logs active energy generation (kWh), reactive power components (kVArh), and accurate timestamps in ISO 8601 format. The dataset spans 40 months (March 2022 to June 2025). Data from 2022 were reserved exclusively from the operational validation case study, while the remaining 30 months were used for model development and testing. Raw measurements are subjected to systematic preprocessing, which includes data validation, outlier detection by statistical thresholds, and verification of temporal consistency to ensure data quality.



Fig. 1. Gama campus of the university of Brasília: (a) (1) Teaching Unit (UED); (2) Academic Unit (UAC); (3) Sports and Equipment Services Module (MESP); (4) Laboratory for Transport Development and Alternative Energies (LDTEA).



Fig. 2. Solarimetric stations installed nearby (1) Teaching Unit (UED).

2.2. Data collection about meteorological and solar irradiation

The monitoring infrastructure includes solarimetric stations (Fig. 2), currently undergoing recalibration for future ground-truth validation. For this study, the solar irradiance and meteorological data were obtained from Solcast's satellite-derived "estimated actuals" via its API [16], which combines images from geostationary weather satellites and data from meteorological models to provide irradiance data at any point on the planet. For this work, the information is specific to the location of Faculty of Science and Technology in Engineering (FCTE - acronym in Portuguese), based on its latitude and longitude. The dataset used has a temporal resolution of 5 minutes, and the spatial resolution changes depending on the parameter, with irradiance measured at 90 m and cloud data at 2 km. The main parameters extracted were GTI (inclined global irradiation) with an azimuth of 0^{circ} and inclination of 16^{circ}, GHI (horizontal global irradiation), air temperature, cloud opacity, wind speed, and GHI and GTI with a clear sky. Two different ways of processing data are used depending on what the analysis needs: Models for ma-

chine learning: To make the calculations easier while keeping the important information, high-frequency data are combined into daily values using irradiance integration (kWh/m²/day), temperature statistical measures (mean, min, max), and atmospheric parameter averaging. NBR 16274:2014: Meteorological data are preserved at their original 5-minute precision during regulatory calculations, with instantaneous irradiance (kW) values immediately utilized in power computations prior to final aggregation into daily energy totals, so ensuring adherence to NBR.

3. Proposed approach

Forecasting models based on mathematics and statistics were implemented with the objective of analyzing historical data about the electricity generated by photovoltaic plants installed at the University of Brasilia. Although machine learning is a powerful tool, its application is not always the ideal solution to real-world problems, and it is necessary to carefully evaluate the objectives of the solution and the context of the problem. In this sense, in order to predict the energy yield of photovoltaic plants, different Machine Learning Models were implemented to evaluate the performance of the photovoltaic generators installed at UnB.

3.1. Machine learning models

This research employs an experimental methodology divided into four sequential stages. Each stage uses different machine learning models to estimate the generation of electricity from photovoltaic plants, as well as detect anomalies, with the results of one stage working as a reference for the next. The methodology is based on the concept that accurate forecast models enable the detection of anomalies through the analysis of residuals (differences between predicted and observed values). This progressive approach measures the advantages of each phase and points out the constraints of each model. To achieve the objectives proposed in this work, 10 Machine Learning algorithms representing different approaches were tested, namely:

- Bagging: Random Forest and Extra Trees;
- Boosting: XGBoost, LightGBM, Gradient Boost, and HGradient Boost;
- Linear: Linear Regression and Lasso Regression;
- Non-linear: Support Vector Regression (SVR); and
- Tree-based: Decision Tree.

The chosen algorithms were trained using electricity generation data recorded from January 2023 to June 2025 (a total of 30 months). As shown in Fig. 3, the photovoltaic energy dataset was partitioned into training, validation, and test sets. These were obtained by means of direct measurement at the common coupling point (PCC). Initially, for the purpose of parametric adjustments, data recorded between January/23 and September/2024 were used to train the algorithms. These data are shown in Fig. 3 in green. Following the work, the data provided by the algorithms were then compared with the data already recorded between the months of October 2024 and December 2024, illustrated in orange in Fig. 3. Finally, for the performance evaluation of the analyzed algorithms, the expected electricity generation profile was generated for the period between January and June 2025. Therefore, it is observed that the dataset covers a period of 30 months (from January 2023 to June 2025), including full seasonal cycles, to ensure that annual variations in weather patterns and solar irradiance are represented. It should be noted that the time division strategy in question adopts best practices for the evaluation of the time series models.

3.2. Training methodology

The evaluation of the models follows a protocol to avoid data leakage, an essential practice in photovoltaic energy forecasting to ensure unbiased results.

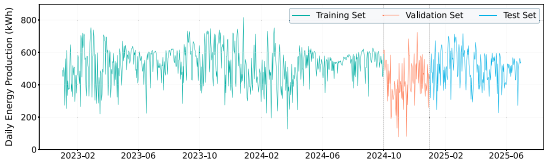


Fig. 3. Dataset partitioning energy production: Division of training, validation and testing sets (January/2023 - June/2025).

- Development phase: Models are trained in the training set (21 months) and validated in the validation set (3 months)
- Final evaluation phase: After defining the optimal configuration, the models are retrained using the expanded set (training + validation = 24 months) and evaluated exclusively in the test set (6 months). which remains untouched throughout the development process.

This approach ensures an evaluation of the actual performance of the models on entirely new data, simulating practical application scenarios in which only historical data is accessible for training. External meteorological data (global horizontal solar irradiance, inclined plane solar irradiance, temperature, among others) were extracted from the historical satellite archive provided by Solcast. Since these records are consulted retrospectively, the validation conducted in Steps I-III, presented below, is characterized as an offline retrospective simulation. The results express the predictive capacity of the model under observed real meteorological conditions and should be understood as the maximum performance limit when compared to a rigorous operational scenario of forecasting for the following day, in which meteorological forecasts with a defined time horizon and their respective uncertainties would be employed. This methodology is aligned with the standard procedure adopted in photovoltaic forecasting investigations [7,11]. The evaluation of performance using meteorological forecast data one day in advance is proposed as a subsequent step in the research.

4. Experimental methodology

The experiments were structured in four main steps, aiming to compare different modeling approaches for PV generation forecast and operating system development:

1. **Stage I**: Statistical Models of Time Series: Classical models based exclusively on univariate historical generation data.
2. **Stage II**: Machine Learning temporal univariate: Application of ML algorithms using temporal feature engineering based only on internal historical data.
3. **Stage III**: Operational Model with Regulatory Validation: Development of physically grounded models using high-resolution meteorological variables, including implementation and direct comparison with the ABNT NBR 16274:2014 method.
4. **Stage IV**: Operational Validation and Anomaly Detection: Implementation of an EPI-based anomaly detection system using ML-predicted energy as reference, validated with a real scenario of documented failure.

4.1. Stage I: Statistical time series models

This step evaluates reference models using traditional statistical methods using exclusively the plant's energy generation history, serving as a basis for comparison for later steps. The following algorithms have been implemented:

- ARIMA (Automatic parameter selection)
- SARIMA (Annual seasonality)
- ETS (Exponential Smoothing) with trend and seasonality
- Prophet (Multiple seasonalities)

4.2. Stage II: Temporal machine learning

The second stage applies Machine Learning algorithms using only internal historical data, but with extensive temporal feature engineering. The objective is to evaluate whether greater complexity surpasses traditional statistical models. The following ML algorithms have been implemented:

- **Ensemble:** Random Forest, Extra Trees, XGBoost, LightGBM, and Gradient Boosting;
- **Linear:** Linear Regression and Lasso Regression
- **Non-linear:** Support Vector Regression.

Therefore, new features were created from the historical data and organized into five distinct categories to capture different aspects of the temporal dimension:

- **T (Basic Temporal):** Temporal components extracted directly from dates, such as day of the week, month, quarter, and season
- **S (Solar/Temporal Specific):** Variables that capture solar and astronomical cycles, including days elapsed since the summer and winter solstices, equinoxes, and other specific seasonal variations
- **C (Cyclic):** Trigonometric (sine/cosine) transformations applied to temporal variables to preserve the cyclical nature and continuity of temporal patterns;
- **L (Lag):** Historical values of the time series in different lags (1, 2, 3, 7, 14, 30 days before) and their respective descriptive statistics to capture temporal dependencies
- **W (Window):** Mobile statistics calculated on sliding windows of different sizes, including means, standard deviations, maximum and minimum values to identify trends and local variability.

From the 5 groups of temporal characteristics, 31 different combinations were generated, representing all the relevant possibilities of grouping the features. This systematic approach totaled 310 experiments, allowing to evaluate both the best set of characteristics for each model and the overall combination that offers the best forecasting performance.

4.3. Stage III: Operating model development

In this stage, models were developed that use meteorological and physical variables of the environment, which reduces the need for specific historical data for the construction of more solid and consistent models, analyzing how these climate variables influence the performance of the same machine learning algorithms that were used in stage II. In addition, a hybrid validation is carried out between machine learning models and the regulatory method defined in the ABNT NBR 16274:2014 standard, which enables an evaluation of forecasting performance in relation to Brazilian regulatory compliance. Therefore, the following meteorological variables were integrated:

- **Irradiance:** Global Tilted Irradiance - GTI, Global Horizontal Irradiance - GHI, Clear Sky Irradiance (W/m^2)
- **Atmospheric:** Air temperature (^{circ}C), wind speed (m/s), cloud opacity (0–1)

Physics-Based Feature Engineering: From high-resolution meteorological data, specific features were engineered following NBR 16274:2014 [17] principles:

- **Clear Sky Index:** $CSI = \frac{GTI}{GTI_{clearsky}}$
- **Cell Temperature:** $T_c = T_a + \frac{GTI}{800}(T_{NOCT} - 20)$
- **Net Irradiance:** GTI corrected for thermal efficiency losses
- **Variability Indicators:** Cloud event detection and irradiance fluctuation metrics

Next, it was necessary to aggregate the meteorological data of solar irradiation so that both sets of data were on the same time scale, since the generation data from the solar plant were available daily and the meteorological data every 5 minutes. The high-frequency data (5 minutes) was aggregated to a daily scale, keeping the important statistical information and preserving essential physics:

- **Irradiance:** Energy integration to kWh/m^2/day
- **Atmospheric:** Statistical measures (mean, min, max)
- **Indicators:** Solar duration hours and weather event frequency

Following the Brazilian technical standard NBR 16274:2014 [17] for photovoltaic system performance analysis, the theoretical model was implemented as the validation benchmark, allowing direct comparison between machine learning predictions and established regulatory methodologies.

DC Power Generation:

$$P_{cc}(i) = P_N \frac{G(i)}{1000}\left[1 + \gamma(T_c(i) - 25)\right]\left[1 + c\ln\left(\frac{G(i)}{1000}\right)\right] \quad (1)$$

Where:

- $P_{cc.}$ = Theoretical DC power (W)
- P_N = Nominal system power (W)
- γ = Temperature coefficient (%/^{circ}C)
- c = Irradiance efficiency coefficient

The cell temperature is estimated by the TNOC (Nominal Cell Operating Temperature) model:

$$T_c(i) = T_a(i) + \frac{G(i)}{800}(T_{NOCT} - 20) \quad (2)$$

Where:

- T_c = Cell temperature (^{circ}C)
- T_a = Ambient temperature (^{circ}C)
- G = Solar irradiance (W/m^2)
- T_{NOC} = Nominal Operating Cell Temperature (^{circ}C)

DC Loss Factor:

$$CP_{c.c.} = \frac{\frac{G}{1000}}{\frac{G}{1000} + a_0 + a_1\frac{G}{1000} + a_2\left(\frac{G}{1000}\right)^2} \quad (3)$$

Where:

- $CP_{c.c.}$ = DC loss correction factor
- a_0, a_1, a_2 = Loss model coefficients
- G = Solar irradiance (W/m^2)

AC Power Conversion with Inverter Efficiency Model:

$$P_{c.a.}(i) = CP_{c.a.}P_{NI}\frac{-(k_1+1) + \sqrt{(k_1+1)^2 - 4k_2\left(k_0 - \frac{P_{c.c.}(i)}{P_{NI}}\right)}}{2k_2} \quad (4)$$

Where:

- $P_{c.a.}$ = Theoretical AC power (W)
- P_{NI} = Nominal inverter power (W)
- k_0, k_1, k_2 = Inverter efficiency curve coefficients
- $CP_{c.a.}$ = AC loss correction factor

The Performance Ratio (PR) is calculated according to the standard methodology defined in IEC 61724-1 [18]:

$$PR = \frac{Y_f}{Y_r} = \frac{E_{PV}}{P_{PV}} \cdot \frac{G_I}{G_{I,ref}} \quad (5)$$

where:

- E_{PV} is the total PV system production [kWh]
- P_{PV} is the rated PV capacity [kWp]
- G_I is the solar irradiation in the plane of the array [kWh/m^2]
- $G_{I,ref}$ = 1.0 kW/m^2 is the reference irradiation.

Daily Energy Integration: The NBR 16274:2014 model calculates instantaneous AC power at 5-minute intervals corresponding to the meteorological data frequency. However, for compatibility with the target variable and ML models, these instantaneous values require temporal integration to daily energy totals (kWh/day). The daily AC energy output is computed as:

$$E_{AC} = \sum_{i=1}^{N} P_{AC}(i) \times \frac{\Delta t}{1000} \quad (6)$$

This aggregation process ensures direct comparability between the regulatory benchmark predictions and both the observed generation data and ML model outputs, all expressed in consistent daily energy units (kWh/day). The temporal integration preserves the physical accuracy of the NBR model while enabling fair performance evaluation against machine learning approaches operating on the same daily prediction horizon.

4.4. Operational validation and anomaly detection

In accordance with the guidelines of international standards IEC 61724-1 [18] and IEC TS 61724-3 [19], in order to evaluate the performance of the PV plant, an anomaly detector was implemented. This indicator, called the Energy Performance Index (EPI), is obtained by comparing the observed daily electricity generation with the expected energy calculated by the performance model, as shown in (7).

$$\mathrm{EPI}(t) = \frac{E_{\mathrm{obs}}(t)}{E_{\mathrm{exp}}(t)}, \quad (7)$$

where $E_{\mathrm{obs}}(t)$ the daily energy is measured and $E_{\mathrm{ext}}(t)$ is the energy is expected by the selected model under actual weather conditions. Therefore, in normal operation, $EPI \approx 1{,}0$. Thus, persistent drops in this indicator signal problems that climatic issues alone do not explain. Unlike PR, which normalizes the performance of the PV system by the reference irradiance (Y_r), EPI uses the electrical energy predicted by a detailed model as the denominator [18]. As the ML models were trained with real injected energy data, the normal system losses are already incorporated [8,10]. Since the calculation $E_{\mathrm{exp}}(t)$ uses the observed irradiance of the day itself, available via Solcast a few hours after sunset, the EPI can be applied daily in real-world operations, without relying on weather forecasts.

4.5. Evaluation metrics

The model's performance is evaluated using six standardized solar forecasting metrics. These metrics, which are mostly in line with the National Renewable Energy Laboratory's (NREL) [20,22] and are widely used in international reviews of forecasting methods [5], make it easier to choose an algorithm and do a full evaluation based on operational deployment needs.

4.5.1. Absolute error metrics

The Mean Absolute Error (MAE) quantifies the average magnitude of prediction errors, providing an intuitive measure of forecasting accuracy:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |p_i - \hat{p}_i| \quad (8)$$

The normalized version (nMAE) expresses the error relative to the mean observed daily generation over the evaluation window:

$$nMAE = \frac{MAE}{\bar{E}_{obs}} \times 100\% \quad (9)$$

4.5.2. Quadratic error metrics

The Root Mean Squared Error (RMSE) emphasizes larger deviations, making it sensitive to outliers and peak prediction errors:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (p_i - \hat{p}_i)^2} \quad (10)$$

Its normalized counterpart (nRMSE) follows the same normalization:

$$nRMSE = \frac{RMSE}{\bar{E}_{obs}} \times 100\% \quad (11)$$

4.5.3. Correlation and variability metrics

The coefficient of determination (R^2) measures the proportion of variance in observed values explained by the model, indicating the goodness of fit:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (p_i - \hat{p}_i)^2}{\sum_{i=1}^{N} (p_i - \bar{p})^2} \quad (12)$$

4.5.4. Bias metrics

The Normalized Bias Error (NBE) quantifies the magnitude of the systematic deviation between predicted and observed values, normalized by the mean observed generation:

$$NBE = \frac{\left| \frac{1}{N} \sum_{i=1}^{N} (\hat{p}_i - p_i) \right|}{\bar{E}_{\mathrm{obs}}} \times 100\% \quad (13)$$

These metrics collectively provide a robust framework for assessing model performance across different operational scenarios, from real-time monitoring to long-term planning applications.

5. Results and discussions

The results achieved in the four experimental stages are presented in this chapter, evidencing the gradual evolution of forecasting performance until the creation of an operational system for anomaly detection. The results are presented in order to show how each phase overcomes the limitations of the previous phase, resulting in an effective solution for the monitoring of photovoltaic plants.

5.1. Stage I - statistical models of time series

ARIMA (Auto-Regressive Integrated Moving Average) is a statistical model used to analyze time series. The model has three main parameters:

- **p (autoregressive)**: how many previous values influence the current value
- **d (integration):** degree of differencing needed to make the series stationary
- **q (moving average):** how past forecast errors affect the current model

The best parameters (p, d, q) were selected using the AIC (Akaike Information Criterion), which seeks the most efficient combination [21]. The optimal values found were: p = 2, d = 1, and q = 2. This means that the model considers the two previous values, applies a differentiation to remove trends, and uses two moving average terms to relate the observations to past errors.

Fig. 4 shows the forecasting performance of the ARIMA model using energy generation data from January 2023 to December 2024 as

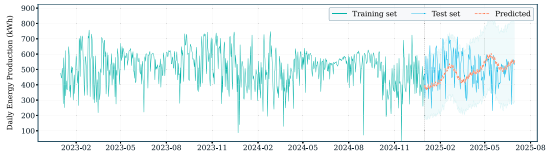


Fig. 4. Forecasting performance of the ARIMA model: comparison between observed and forecasted values (January/2025 - June/2025).

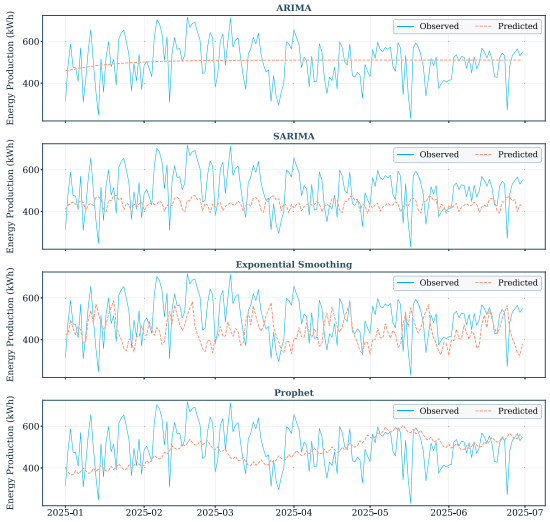


Fig. 5. Individual performance assessment of statistical forecasting models in the test period (Jan/2025 - Jun/2025). Each subplot compares observed daily energy production (solid blue line) with predictions from a single statistical model (orange dashed line): (1) ARIMA, (2) SARIMA, (3) Exponential Smoothing, and (4) Prophet.

Table 1
Performance comparison of statical time-series models.

Model	MAE	nMAE	R^2	NBE	RMSE	nRMSE
ARIMA	78.82	15.62	0.0	1.06	98.08	19.43
Prophet	90.57	17.95	-0.37	4.38	114.7	22.73
SARIMA	103.5	20.51	-0.56	13.91	122.65	24.3
ETS	113.95	25.03	-0.29	5.38	138.11	30.33

the training period (shown in green). The model's predictions for the test period (January to June 2025) are displayed in orange and compared against the actual observed values (shown in blue). The detailed comparison reveals that the ARIMA model struggles to accurately capture the inherent variability characteristic of photovoltaic energy generation, demonstrating limited effectiveness in predicting the fluctuations typical of solar power systems.

The complete evaluation of statistical forecasting models (SARIMA, Holt-Winters, Prophet, and ARIMA) was conducted under standardized conditions with identical parameter optimization, training periods, and evaluation datasets. The comparative results, presented visually in Fig. 5 and quantitatively in Table 1, reveal a consistent pattern of inadequate performance across all traditional time series methods. The statistical metrics (MAE, nMAE, RMSE, nRMSE, R^2, and NBE) confirm the visual observations, showing that conventional forecasting approaches struggle to model the stochastic nature of solar irradiance and its impact on photovoltaic generation patterns.

The performance measures, shown in Table 1, back up what could be seen as wrong with the forecasting plots. All statistical models aren't very good at making predictions. Their normalized mean absolute errors (nMAE) range from 15.62% to 25.03%, and their R^2 values are either zero or null, which means they don't do any better or worse than simple

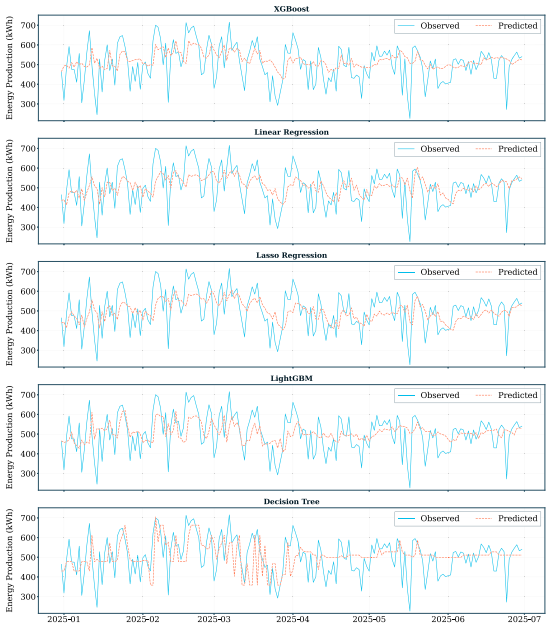


Fig. 6. Comparison of the five best-performing machine learning models for daily photovoltaic energy forecasting using only temporal features. Solid lines represent observed energy production, and dashed lines represent model predictions over the same test period (January-July 2025).

Table 2
Performance comparison of statical time-series models with temporal features only.

Model	MAE	nMAE	R^2	NBE	RMSE	nRMSE
XGBoost	71.43	14.16	0.12	1.48	92.1	18.26
Linear Regression	71.67	14.21	0.11	1.19	92.45	18.32
Lasso Regression	71.98	14.27	0.13	0.54	91.32	18.1
Gradient Boost	72.16	14.3	0.07	1.74	94.27	18.68
LightGBM	72.39	14.35	0.1	0.34	92.8	18.39
Decision Tree	73.58	14.58	0.05	2.88	95.73	18.97
Random Forest	74.01	14.67	0.05	0.22	95.61	18.95
SVR	76.0	15.06	0.02	4.48	96.78	19.18
Extra Tree	76.43	15.15	-0.02	0.17	99.17	19.66
HGradient Boosting	81.61	16.18	-0.11	0.83	103.3	20.47

historical averaging. The ARIMA model that works the best only gets $R^2 = 0.0$ with 15.62% nMAE. Other methods have negative correlation values. The results clearly show that traditional time series methods are inadequate for predicting solar generation under highly variable weather conditions.

5.2. Stage II - temporal machine learning

The second phase of this study analyzed Machine Learning algorithms, but only with historical internal data, obtained through temporal characteristic engineering. The aim was to investigate whether machine learning models with temporal features would be able to outperform conventional statistical models of the First Stage. (Fig. 6) presents the comparison between actual and forecasted values for the main machine learning models tested, demonstrating the superior ability of these methods to capture the temporal patterns of solar generation.

Graphs of actual and forecasted values show a great improvement in machine learning models compared to traditional statistical methods,

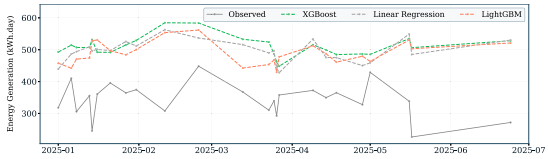


Fig. 7. Performance of the three best temporal ML models (XGBoost, Linear Regression and LightGBM) exclusively on low-efficiency days. (test period: January-July 2025).

Table 3
Model performance comparison with advantage over NBR baseline.

Model Type	Best	nMAE (%)	R^2	Advantage over NBR
Physics ML	XGBoost	4.91	0.90	+30.7% accuracy
NBR 16274:2014	Regulatory	7.09	0.79	Baseline
Temporal ML	XGBoost	14.16	0.12	-99.7% accuracy
Statistical	ARIMA	15.62	0.00	-120.3% accuracy

as they are better aligned with the data and can follow the temporal variations of solar energy. Despite the progress, the graphs show limitations in forecasting, with discrepancies in periods of low irradiance and inconsistencies that indicate a lack of understanding of the physical causes. Visual analytics are corroborated by the performance metrics presented in Table 2. The models showed normalized mean errors (nMAE) between 14% and 16%, indicating significant limitations for critical operational applications. Some models, such as XGBoost and Linear

Regression, show that their coefficients of determination (R^2) are greater than 0.10. But most algorithms have an R^2 close to zero, showing that they do not explain the variations in the data satisfactorily.

Finally, to evaluate the capacity of the models in adverse situations, a segmented analysis was made on days of low performance, when the weather conditions were unfavorable, such as a lot of cloud and low solar irradiance. These days, forecasting accuracy is essential for anomaly detection and operational management of photovoltaic systems. As illustrated in Fig. 7, segmented analysis shows a recurring failure in temporal machine learning models, especially in operations-critical situations. The overall performance shows moderate adequacy, but some days showed low efficiency, with an estimation error of 25 to 50%, since the models predicted about 500 kWh daily, while the actual generation was around 200 and 450 kWh.

5.3. Stage III - operating model development

The third stage sought to build models from meteorological and external variables, minimizing the use of internal historical data to generate more consistent and stable models and implementing the NBR 16274:2014 regulatory model as benchmark. The intention was to verify whether the addition of climatic variables and physical-meteorological characteristics could improve performance in relation to the models analyzed in Stage II, which were based only on time. Fig. 8 presents the comparison between actual and forecasted values for the main machine learning models tested, demonstrating the superior ability of these methods to capture the temporal patterns of photovoltaic energy.

NBR 16274:2014 Regulatory Baseline: The Brazilian technical standard NBR 16274:2014 was used as the regulatory standard. It gave a physics-based point of reference for comparison.

Implementation Parameters:

- DC loss coefficients: $a_0 = 0.075$, $a_1 = -0.08$, $a_2 = 0.20$
- Inverter efficiency curve: $k_0 = 0.015$, $k_1 = 0.045$, $k_2 = 0.050$
- AC system losses: 2.5% total distributed components

Physics-Enhanced Machine Learning Models:

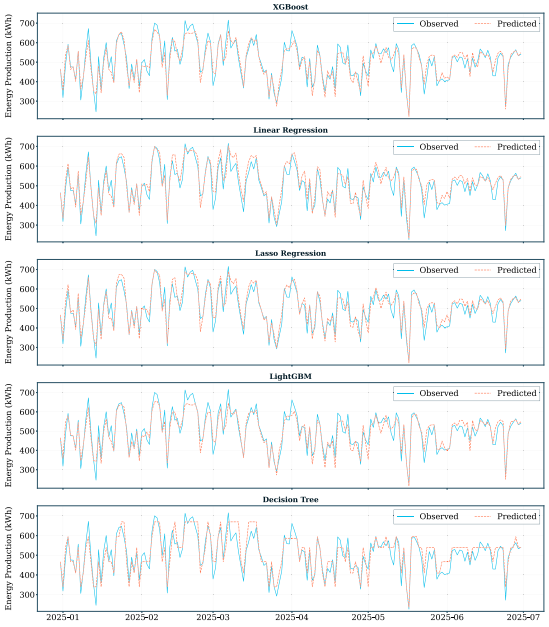


Fig. 8. Comparison of the five best-performing machine learning models for daily photovoltaic energy forecasting with exogenous variables. The subplots show the observed and predicted energy series over the same test period (January-July 2025), illustrating the contribution of external inputs to model performance.

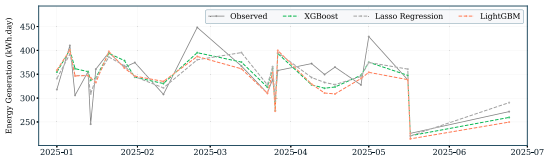


Fig. 9. Performance of the three best physics-enhanced ML models (XGBoost, Lasso Regression and LightGBM) exclusively on low-efficiency days. (test period: January-July 2025).

To ensure a direct and fair comparison with the results of Stage II, the same segmentation methodology was applied, using exactly the same days of low performance used in Stage II. This analysis focused on periods when weather conditions were unfavorable, such as high cloud cover and low solar irradiance, maintaining methodological consistency between the stages.

Applying the physical models to the same subset of critical days as Step II, the results show a big improvement (Fig. 9):

- Temporal ML Models (Step II): 25–50% errors, predicting 500 kWh when actual generation was between 200–450 kWh.
- Exogenous ML Models (Stage III): Exceptional adherence to observed values, with minimal discrepancies even in the harshest conditions.

Models with external variables have normalized mean errors (nMAE) of 4.9% to 9.5%, evidencing greater accuracy for critical operational applications. Models such as XGBoost, Lasso Regression, and LightGBM have R^2 between 0.89 and 0.90, showing excellent explanatory capacity for variations in the data. The model with the lowest performance (SVR) has an nMAE of 9.5%, lower than the best results.

The comprehensive performance metrics are presented in Table 4. Models with external variables demonstrated normalized mean errors (nMAE) of 4.9% to 9.5%, evidencing greater accuracy for critical operational applications. Models such as XGBoost, Lasso Regression, and LightGBM achieved R^2 between 0.89 and 0.90, showing excellent explanatory capacity for variations in the data. Even the model with the lowest performance (SVR) maintained an nMAE of 9.5%, still superior to the best temporal models.

Table 4
Performance of physics-enhanced ML models with exogenous variables.

Model	Features	MAE	nMAE	R^2	NBE	RMSE	nRMSE
XGBoost	TI	24.77	4.91	0.90	0.15	31.09	6.16
Lasso Regression	IC	24.89	4.94	0.90	1.03	32.12	6.37
LightGBM	CC	25.76	5.10	0.89	0.15	33.40	6.62
Linear Regression	IC	26.12	5.18	0.88	2.45	33.34	6.61
Extra Tree	TI	27.32	5.41	0.88	1.30	35.59	7.05
HGradient Boosting	TI	27.53	5.46	0.87	0.68	34.25	6.79
Random Forest	TI	27.75	5.50	0.88	1.25	35.76	7.09
Gradient Boost	IC	30.63	6.07	0.87	2.30	39.47	7.82
Decision Tree	TI	32.58	6.46	0.84	1.94	41.74	8.27
NBR 16274:2014	–	35.73	7.09	0.79	4.64	44.02	8.74
SVR	IC	47.95	9.50	0.60	2.88	61.86	12.26

Feature Sets: • **TI**: gti_net, csi, cell_temp, day_of_year_cos, season, day_since • **IC**: gti, ghi, csi_mean, day_of_year_cos, month_cos • **CC**: csi_mean, day_of_year_cos, season, day_since, gti

Table 5
Performance summary by stage and approach.

Stage	Approach	Best model	nMAE (%)	R^2
I	Statistical time-series	ARIMA	15.62	0.00
II	Temporal ML	XGBoost	14.16	0.12
—	ABNT NBR 16274:2014	Regulatory	7.09	0.79
III	Physics-enhanced ML	XGBoost	4.91	0.90
IV	Anomaly detection	EPI-based	—	—

The top three exogenous models (4.91-5.10% nMAE) significantly outperform the three best temporal models (14.16-14.27% nMAE) with an improvement of about 65% in forecast error. Even more critical, the explanatory power rose from R^2 0.12 to R^2 0.90, demonstrating that physical variables encode the essential mechanisms of solar generation that temporal variables cannot adequately represent.

These models represent the three most effective feature engineering strategies that were discovered in the previous step. Only models that presented nMAE below 5.2% and R^2 above 0.89 were chosen, ensuring high forecasting accuracy for reliable operational anomaly detection.

An impressive discovery emerges when examining the performance of NBR 16274:2014 (Table 3): This physics-based regulatory model (nMAE 7.09%) dramatically outperformed all statistical models (15.62-25.03% nMAE) and temporal ML algorithms (14.16-16.18% nMAE) that required 24 months of training data. The NBR model, using only equipment/design specifications and meteorological data, achieved superior accuracy without any reliance on historical data. Although the physics-enhanced ML models achieved the best performance (4.91% nMAE), their modest advantage over the NBR demonstrates that physical understanding offers exceptional value when operational history is limited. This makes this approach ideal for new installations or systems without established data records.

6. Case study

To simulate realistic operating conditions, a specific time division was implemented that reflects the practical challenges of implementing monitoring systems with limited data. To this end, unlike the development stages that used data from 2023 to 2025, in this case study data obtained in 2022 was used to simulate a practical implementation of the proposed methodology. This methodology allows us to analyze the stability of the models and their ability to generalize to periods not used in creation.



Fig. 10. Time division of data for operational validation: limited training period and validation phases (test period: jan/2022-jan/2023).

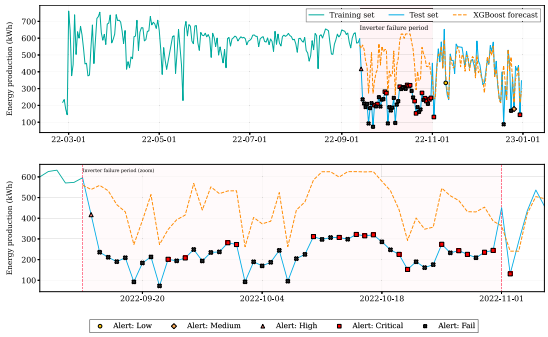


Fig. 11. Time series of energy production comparing observed values with those estimated by the XGBoost model. The shaded region highlights the inverter failure period, while the markers indicate the alert levels defined by the Energy Performance Index (EPI).

For this case study, it should be noted that between September and October 2022, the UED PV system, depicted in Fig. 1, had a failure in the 75 kW inverter, which was inactive for approximately 48 days. This event provided ideal conditions to evaluate the effectiveness of the anomaly detection mechanism and to estimate the impact of underperformance over a long period of time. Fig. 10 illustrates time division used for this validation.

Based on this split, three aspects are evaluated: (i) model robustness with only 6.5 months of training data (March-September 2022); (ii) anomaly detection during the inverter failure period (September-October 2022); (iii) operational stability under normal post-repair conditions (November 2022-February 2023).

Stability is defined as the consistency of Energy Performance Index (EPI) in normal operation (low variance around 1.0), assessed by the EPI standard deviation and trend verification in the residuals.

The EPI detector given by (7) was applied to the three periods. The thresholds were calibrated based on the system behavior during the normal operating period (March-August/2022). According to IEC TS 61724-3, the evaluation criteria must be customized for each system [19], therefore, the thresholds were defined empirically based on the distribution of EPI during the normal operating period, being specific to this system and model.

- EPI $\geq$ 0.90: Normal
- 0.85 $\leq$ EPI $<$ 0.90: Low
- 0.80 $\leq$ EPI $<$ 0.85: Medium
- 0.75 $\leq$ EPI $<$ 0.80: High
- EPI $<$ 0.75: Critical
- EPI $<$ 0.50: Fail

Throughout the 48-day period in which the failure occurred (between September 14, and October 31, 2022), the observed generation was 10.9 MWh against the 22.5 MWh expected by the XGBoost model (Fig. 11), representing an estimated loss of 11.6 MWh (51.7%). Resulting in an estimated loss with the isolated operation of the remaining 50 kW inverter, confirming that the 75 kW PV system remained inactive throughout the period. Had the EPI system been operational, the failure would have been detected on the first day, at which point the EPI would

have reached approximately 0.40, a value significantly lower than the critical limit of 0.50, which could have reduced losses to the minimum period necessary for repairs.

The inclusion of meteorological variables and physical characteristics completely eliminated all critical limitations that had been pointed out in Step II. Days of low efficiency that previously generated operationally unacceptable errors are now within acceptable ranges for practical applications, with ML models using exogenous variables demonstrating superior accuracy. This improvement reinforces that the physical approach is crucial to ensure that the monitoring systems work reliably, especially in critical situations where it is vital to detect anomalies accurately.

7. Conclusions

This work evaluated and compared progressive strategies for daily forecasting of electricity generation in photovoltaic plants and their application in performance monitoring and anomaly detection systems. The models were analyzed using standardized error and fit metrics, allowing the selection of the most suitable approaches for operational use in grid-connected PV systems. Analyzing the synthesis of the results obtained and presented in Table 5, it is possible to conclude that physical/exogenous variables are the main accuracy factor, with the best model with meteorological variables achieving nMAE = 4.91% e $R^2 = 0.90$, while the best purely temporal model achieved nMAE = 14.16% e $R^2 = 0.12$.

Therefore, the results indicate that the use of forecasting models can support the early detection of relevant performance deviations, reducing the risk of prolonged energy losses and favoring preventive maintenance interventions, and contributing to achieving the expected financial return. With nMAE = 7.09%, $R^2 = 0.79$, the ABNT NBR 16274:2014 baseline outperformed purely temporal models, indicating that this can be an acceptable procedure when there is little historical operating data.

It is important to highlight that the results presented in Stages I-III were achieved based on real historical meteorological data, which was used as input parameters. In a strictly day-ahead operational application, the uncertainties inherent in meteorological forecasts would add a margin of error; therefore, the results presented constitute the maximum possible performance level in this context.

Finally, the implemented anomaly identification system operates in D + 0 mode, that is, at the end of each day, the EPI compares the actual generation recorded with that predicted by the model from the measured irradiance, without the need for meteorological projections. Continuous reductions that exceed the established limits trigger warnings that support preventive or corrective maintenance actions. As exemplified in the case study presented in section VI, it was observed that the performance drop was detected on the very first day that one of the inverters was inoperative, requiring intervention from the technical team at the installation site.

To conclude, it was shown that the proposed approach can be used to develop energy management systems capable of real-time monitoring PV systems in medium and large buildings, tracking electrical parameters, storing historical data, presenting information in graphs and reports, and generating standardized alerts. It is important to highlight the importance of monitoring and evaluating the Energy Performance Index (EPI) in order to send alerts that are easy for managers to understand, as well as to determine the preventive or corrective actions necessary to ensure the expected payback.

Future work could include evaluations using day-ahead weather forecasting, multi-site validation, sub-hourly forecasts, comparison with deep learning approaches, and full integration into an Energy Management System.

CRediT authorship contribution statement

Flávio V. Leão: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Renato C. Sampaio:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Data curation, Conceptualization; **Loana N. Velasco:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization; **Alex Reis:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation; **Ênio C. Resende:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation; **Luiz C. G. Freitas:** Writing - review & editing, Writing - original draft, Methodology, Investigation, Formal analysis, Data curation.

Acknowledgement

The authors would like to thank Ministério da Educação (MEC), Fundação de Amparo à Pesquisa do Distrito Federal (FAPDF), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Universidade de Brasília (UnB), and Universidade Federal de Uberlândia, for supporting this research project.

Data availability

Data will be made available on request.

References

[1] P. Zhao, S. Suryanarayanan, M.G. Simões, An energy management system for building structures using a multi-agent decision-making control methodology, in: 2010 IEEE Industry Applications Society Annual Meeting, Houston, TX, USA, 2010, pp. 1–8. https://doi.org/10.1109/IAS.2010.5615412

[2] G.K. Venayagamoorthy, R.K. Sharma, P.K. Gautam, A. Ahmadi, Dynamic energy management system for a smart microgrid, IEEE Trans. Neural Netw. Learn. Syst. 27 (8) (2016) 1643–1656. https://doi.org/10.1109/TNNLS.2016.2514358

[3] P. Chaudhary, M. Rizwan, Energy management supporting high penetration of solar photovoltaic generation for smart grid using solar forecasts and pumped hydro storage system, Renew. Energy 118 (2018) 928–946. https://doi.org/10.1016/j.renene.2017.10.113

[4] S. Voutsinas, D. Karolidis, I. Voyiatzis, et al., Development of a machine-learning-based method for early fault detection in photovoltaic systems, J. Eng. Appl. Sci. 70 (2023) . https://doi.org/10.1186/s44147-023-00200-0

[5] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.J. Martinez-de Pison, F. Antonanzas-Torres, Review of photovoltaic power forecasting, Sol. Energy 136 (2016) 78–111. https://doi.org/10.1016/j.solener.2016.06.069

[6] S. Sobri, S. Koohi-Kamali, N.A. Rahim, Solar photovoltaic generation forecasting methods: a review, Energy Convers. Manag. 156 (2018) 459–497. https://doi.org/10.1016/j.enconman.2017.11.019

[7] D. Markovics, M.J. Mayer, Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction, Renew. Sustain. Energy Rev. 161 (2022) 112364. https://doi.org/10.1016/j.rser.2022.112364

[8] S.R. Madeti, S.N. Singh, A comprehensive study on different types of faults and detection techniques for solar photovoltaic system, Sol. Energy 158 (2017) 161–185. https://doi.org/10.1016/j.solener.2017.08.069

[9] S.R. Madeti, S.N. Singh, Monitoring system for photovoltaic plants: a review, Renew. Sustain. Energy Rev. 67 (2017) 1180–1207. https://doi.org/10.1016/j.rser.2016.09.088

[10] A. Livera, M. Theristis, G.M. akrides, G.E. Georghiou, Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems, Renew. Energy 133 (2019) 126–143. https://doi.org/10.1016/j.renene.2018.09.101

[11] M.D. Benedetti, F. Leonardi, F. Messina, C. Santoro, A. Vasilakos, Anomaly detection and predictive maintenance for photovoltaic systems, Neurocomputing 310 (2018) 59–68. https://doi.org/10.1016/j.neucom.2018.05.017

[12] A.A. Abdallah, et al., Model-based fault detection in photovoltaic systems: a comprehensive review and avenues for enhancement, Heliyon 10 (4) (2024) . https://doi.org/10.1016/j.heliyon.2024.e25475
[13] A. Alcañiz, M.M. Nikam, Y. Snow, O. Isabella, H. Ziar, Photovoltaic system monitoring and fault detection using peer systems, Prog. Photovolt. Res. Appl. 30 (9) (2022) 1072–1086. https://doi.org/10.1002/pip.3558
[14] IEA PVPS Task 13, Improving efficiency of PV systems using statistical performance monitoring, Technical Report IEA-PVPS T13-07:2017, International Energy Agency, 2017.
[15] A.M. del Rio, et al. Performance rate analysis in photovoltaic solar plants by machine learning, Adv. Energy Sustain. Res. 6 (2025) 2500144. https://doi.org/10.1002/aesr.202500144
[16] Solcast, "Solar irradiance data and PV system power output data," (2026), (Solcast Pty Ltd, 2024. [Online]. Available: https://solcast.com).
[17] Grid Connected Photovoltaic Systems, 2014.
[18] Photovoltaic system performance — Part 1: Monitoring, 2021.
[19] Photovoltaic system performance — Part 3: Energy evaluation method, Technical Report IEC TS 61724-3:2016, International Electrotechnical Commission, Geneva, Switzerland, 2016.
[20] J. Zhang, B.-M. Hodge, A. Florita, S. Lu, H.F. Hamann, V. Banunarayanan, A.M. Brockway, Metrics for evaluating the accuracy of solar power forecasting, Technical Report NREL/CP-5500-60142, National Renewable Energy Laboratory, Golden, CO, USA, 2013. https://docs.nrel.gov/docs/fy14osti/60142.pdf.
[21] J.E. Cavanaugh, A.A. Neath, The Akaike information criterion: background, derivation, properties, application, interpretation, and refinements, Wiley Interdiscip. Rev. Comput. Stat. 11 (3) (2019) . https://doi.org/10.1002/wics.1460
[22] J. Zhang, A. Florita, B. Hodge, S. Lu, H. Hamann, V. Banunarayanan, A. Brockway, A suite of metrics for assessing the performance of solar power forecasting, Sol. Energy 111 (2015) 157–175. https://doi.org/10.1016/j.solener.2014.10.016