POWER PERFORMANCE ASSESSMENT IN A GRID-CONNECTED PV SYSTEM IN THE BRAZILIAN NORTHEAST

Emilly Oliveira[1], Diego Miranda[2], Lucas Emanuel Barboza[3] and Olga de Castro Vilela[4]

[1] Federal University of Pernambuco (UFPE), Recife (Brazil)

[2] German Aerospace Center (DLR), Institute of Networked Energy Systems, Oldenburg (Germany)

[3] Federal University of Pernambuco (UFPE), Recife (Brazil)

[4] Federal University of Pernambuco (UFPE), Recife (Brazil)

Abstract

This paper presents a data-driven framework for performance assessment and anomaly detection in a 0.815 kWp grid-connected photovoltaic (PV) system located in Recife, Brazil. A simplified linear model was implemented to estimate AC power output as a function of plane-of-array irradiance and module temperature, incorporating a calibrated global loss factor (η = 0.90) representative of real operating conditions. The model achieved strong agreement with measured data (R^2 = 0.95, rRMSE = 12.8%), confirming its suitability for low-cost monitoring applications. Two complementary anomaly detection methods were applied: an envelope-based approach using modeled versus measured power binning, and a residual-based approach using boxplot-derived statistical thresholds. Both techniques successfully identified known low-generation events and controlled shading experiments, validating their reliability under real field conditions. The proposed methodology offers a simple, interpretable, and computationally efficient tool for operational PV monitoring and can be extended to real-time fault detection in small and medium-scale distributed systems.

Keywords: PV performace, PV modeling, anomaly detection, power performance assessment

1. Introduction

Ensuring the long-term performance and reliability of photovoltaic (PV) systems remains a major challenge, particularly in outdoor environments where systems are continuously exposed to varying weather and operation conditions. Factors such as high ambient temperatures, rainfall, dust accumulation, partial shading, and electrical degradation can significantly reduce PV power output if not promptly identified and corrected (Madeti & Singh, 2018). Continuous monitoring of system performance is therefore essential to guarantee the expected energy yield and to prevent economic losses for operators and consumers (Firth et al., 2010; Platon et al., 2015). The growing penetration of PV systems in distributed generation, especially in tropical regions such as the Brazilian Northeast, reinforces the need for accurate and low-complexity performance assessment techniques. In these regions, high irradiance levels combined with elevated module temperatures and intermittent cloud cover make the detection of performance anomalies particularly challenging. Traditional key performance indicators, such as the Performance Ratio (PR), provide valuable insight into overall efficiency but are often insufficient to detect short-term deviations or subtle system degradations. Consequently, hybrid-monitoring strategies that combine empirical modeling and statistical fault detection have gained relevance. Building upon this perspective, this work proposes a data-driven approach to assess the operational performance of a grid-connected PV system located at the Renewable Energy Center (CER) of the Federal University of Pernambuco (UFPE), Brazil. The method uses on-site measurements of plane-of-array irradiance, module temperature, and AC power output to model the PV power with an adjusted global loss factor. Deviations between measured and modeled power generation are analyzed using envelope-based and residual-based methods to identify anomalies in real operation. This framework enables the quantification of daily performance deviations and energy losses, providing an efficient anomaly detection tool for PV monitoring under operational conditions.

2. Motivation and Background

Recent literature classifies PV fault detection and performance assessment techniques into two main categories: module-level approaches, often supported by infrared imaging or drone-based thermography, and system-level

10.18086/swc.2025.03.58 Available at http://proceedings.ises.org

approaches, which rely on data acquisition and modeling of normal operational indicators (Qu et al., 2022; Miraftabzadeh et al., 2023). For systems equipped with central or string inverters, common in medium and large-scale installations, module-level inspection can be more challenging. Hence, system-level monitoring approaches has become a key strategy for detecting underperformance using standard field measurements such as irradiance, temperature, and power output (Wang et al., 2018). Within model-based techniques, two broad groups are identified:

- Physical or empirical models, which use simplified analytical relations between irradiance, temperature, and output power; and
- Data-driven or artificial intelligence models, which rely on machine learning algorithms such as neural networks, Random Forest, XGBoost, or k-nearest neighbors (Madeti & Singh, 2018; Petribú, 2022; Miraftabzadeh et al., 2023; Taghezouit et al., 2024).

Physical models remain highly valuable due to their low complexity, interpretability, and ease of implementation. Several studies have shown that even a single-equation model can effectively predict PV power and support fault detection when calibrated with reliable field data (Firth et al., 2010; Platon et al., 2015). These models use irradiance and module temperature to estimate expected power, and faults are detected when the measured generation deviates significantly from model predictions. Platon et al. (2015) demonstrated a detection rate above 90% using a simple AC power model divided by irradiance ranges, while maintaining low computational cost. However, model accuracy strongly depends on good field instrumentation, robust data preprocessing and appropriate identification of normal operating conditions. Wang et al. (2018) emphasized that accurate performance estimation requires models that can withstand noisy or erroneous measurements and adapt to varying weather regimes. Similarly, Qu et al. (2022) introduced an unsupervised weather pattern recognition approach to improve detection under changing hourly conditions. These insights highlight the importance of combining reliable modeling, noise-resilient preprocessing, and persistence analysis for robust PV performance assessment. The present work adopts a simplified physical linear model with a global loss coefficient (η) adjusted through empirical data from a small grid-connected PV system. By coupling this model with residual and envelope-based anomaly detection, the proposed framework achieves reliable fault identification using a minimal set of input variables. This makes the approach particularly suitable for real-world distributed generation applications, where monitoring infrastructure is often limited but the need for dependable performance evaluation continues to grow.

3. Methodology

3.1 Data Description and Pre-processing

The analysis was carried out using data from a 0.815 kWp grid-connected photovoltaic (PV) system located at the Renewable Energy Center (CER) of the Federal University of Pernambuco (UFPE), in Recife, Brazil. The PV array is oriented toward the north with a tilt angle of 20°, representative of the site's optimal inclination. Plane-of-array (POA) irradiance was measured using a calibrated silicon pyranometer installed with the same tilt and azimuth as the PV modules, ensuring accurate representation of the solar radiation incident on the array surface. Module temperature was monitored with LM35 sensors mounted on the rear side of representative PV modules. The AC power output was obtained from micro-inverter measurements, and to derive comparable modeled values, the estimated DC power from the linear model was multiplied by the inverter efficiency, as specified in the manufacturer's datasheet for typical load conditions. Data corresponding to solar elevation angles below 3° were excluded to remove nighttime and near-sunrise/sunset periods, which are particularly susceptible to sensor noise and shading effects. This filtering approach follows the quality assurance procedures recommended by Miranda et al. (2023). Additional manual quality flags were applied to exclude days affected by irradiance sensor malfunctions, identified by persistently low or flatline POA readings. Because irradiance and temperature data were recorded at 1-minute intervals, the modeled power was computed at the same temporal resolution and subsequently averaged to 5-minute intervals to match the inverter dataset. A time-alignment procedure was then performed to synchronize both datasets according to their timestamps. Measurements showing near-zero power during daytime hours were discarded, as these typically indicate temporary communication errors or inverter downtime. This pre-processing ensured that the final dataset used for model calibration and anomaly detection reflected consistent and physically meaningful operating conditions.

3.2 Proposed Model

In order to estimate the PV system output power, the model from Equation 1 was applied. This model adjusts the standard test condition (STC) power of a PV module based on variations in both irradiance and module temperature. It is commonly used in performance studies due to its simplicity and accuracy, without the need for detailed I-V curve modelling (Narvarte, 2008).

$$P_{AC,model} = \eta \cdot \eta_{inv} \cdot P_{STC} \cdot \frac{G_{POA}}{G_{STC}} [1 + \alpha(T_{cell} - T_{STC})] \quad \text{(eq.1)}$$

Where P_{STC} is the nominal array power at Standard Test Conditions (STC), G_{POA} is the measured plane-of-array irradiance (W/m²), G_{STC} is the irradiance under STC (1000 W/m²), T_{cell} is the operating cell temperature (°C), T_{STC} is the module temperature (°C) under TSC (25°C), α is the temperature coefficient of power, η_{inv} is the inverter efficiency obtained by the manufactures datasheet and η is the global loss factor, estimated by fitting modeled vs. measured power on fault-free reference days.

Here, the model includes an additional coefficient, denoted as η, which represents the average behavior of losses associated with the photovoltaic generator, such as soiling, ohmic losses, module mismatch, and incidence angle modifier (IAM) effects (Barboza, 2023). The cell operating temperature obtained according to Equation 2.

$$T_{cell} = \mathrm{T}_m + \frac{G_{POA}}{G_{STC}} \cdot \Delta T \quad \text{(eq. 2)}$$

Where T_{cell} is the predicted operating cell temperature (°C), T_m is the measured back-surface module temperature (°C) and ΔT denotes the temperature difference between the cell and the module back surface at an irradiance level of 1000 W/m² (°C). For flat-plate modules in an open-rack mount, this difference is typically 3 °C (King and Boyson, 2004). To prevent overestimation of the AC power during periods of inverter saturation, the modeled output was clipped at 600 W, corresponding to the rated AC power limit of the inverter. This ensures that simulated values remain consistent with the physical operating constraints of the system and the actual performance observations.

3.2.1 Weather-Corrected Performance Ratio (CPR)

The Performance Ratio (PR) is a standardized metric defined by IEC 61724 to quantify the ratio between the actual energy yield and the theoretical energy yield under standard test conditions. However, PR values are inherently influenced by environmental factors such as cell temperature and irradiance variability, which can mask true performance deviations (Dierauf et al., 2013). To account for these effects, a weather-corrected performance ratio (CPR) was calculated as shown in Equation 3.

$$PR_{corr} = \frac{\sum E_{AC,i}}{\sum P_{STC} \cdot \frac{G_{POA}}{G_{STC}} [1 + \alpha(T_{cell} - T_{STC})]} \quad \text{(eq.3)}$$

Where $E_{AC,i}$ is the is the measured AC energy at time, G_{POA} is the measured plane-of-array irradiance (W/m²), G_{STC} is the irradiance under STC (1000 W/m²), T_{cell} is the cell temperature (°C) and α is the temperature coefficient of power. The CPR normalizes power yield to a constant reference temperature, effectively removing the influence of short-term meteorological fluctuations (Dierauf et al., 2013). Figure 1 illustrates the daily evolution of CPR and highlights the selected reference days used for model calibration. Days within the interquartile range (10th–90th percentiles) were considered representative of normal performance and were used to estimate the global loss factor (η) through least-squares optimization, minimizing the residual between measured and modeled AC power. Significantly low daily PR values often arise from partial shading, inverter malfunction, or excessive thermal and conduction losses. In contrast, CPR values exceeding 1.0 should be interpreted cautiously, as they may result from measurement errors, sensor calibration issues, or atypical environmental conditions that momentarily exaggerate system performance (Vasisht et al, 2016). These extreme cases are valuable for testing the sensitivity and robustness of the detection framework, as they represent atypical operating conditions of the system.

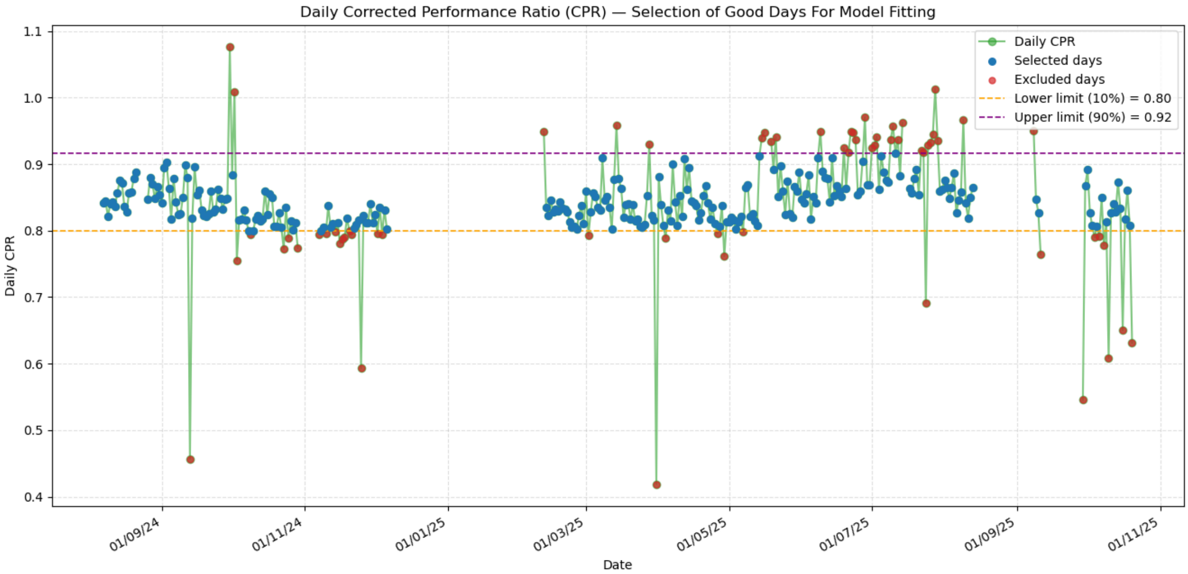

Figure 1: Daily Corrected Performance Ratio (CPR) for the analyzed period. Blue points indicate the selected reference days (within the 10th–90th percentiles) used for model calibration, while red points correspond to excluded days. Orange and purple dashed lines represent the lower and upper percentile thresholds, respectively.

3.3 Anomaly Detection Mechanism

Two complementary techniques were implemented to identify abnormal operational behavior in the PV system:

- a proposed envelope-based method using modeled versus measured power binning, which defines adaptive envelopes representing normal performance behavior;
- a residual-based method, a classical statistical approach that applies boxplot-derived thresholds to identify deviations based on residual analysis.

Both methods aim to detect periods where the measured AC power significantly deviates from the modeled behavior, ensuring anomaly identification under varying irradiance and temperature conditions.

3.3.1 Envelope (Bins) Method

An adaptation of the envelope method proposed by Petribú (2022) was applied to the measured curve. The procedure divides the modeled power axis into uniform irradiance-dependent ranges (bins), each representing a specific operating region of the system. For each bin, the median of the measured AC power is computed, establishing a reference value representative of normal performance within that range. The interquartile range (IQR) then characterizes the dispersion of points around this median. For every measurement, the normalized Euclidean distance from the bin median is calculated based on both the deviation in power and the relative irradiance width of the bin. This process yields a distribution of distances within each bin, allowing an adaptive detection threshold to be defined from the empirical data. Points whose normalized distance exceeds the bin-specific percentile threshold (e.g., 97.5th percentile) are flagged as outliers. In addition, a persistence criterion is applied to avoid spurious detections, and an anomaly is confirmed only when it persists for at least three consecutive time steps. The detection threshold is defined separately for each bin, allowing the envelope method to automatically adapt to variations in data dispersion across different operating conditions. In this work, the independent variable x corresponds to the modeled AC power while the dependent variable y represents the measured AC power, allowing the deviations from the theoretical curve to be interpreted directly as performance losses or anomalies. The resulting envelope delineates the expected power generation region under normal conditions, and any point outside this dynamic boundary is classified as an anomalous measurement.

3.3.2 Residual (Boxplot) Method

A second approach was implemented based on the residuals between measured and modeled AC power. This method assumes that, under normal operation, the residuals should be symmetrically distributed around zero. Outliers are identified when the absolute residual exceeds a global statistical threshold derived from the boxplot rule, as shown in Equation 4.

$$|r_i| > Q_3 + k \cdot (Q_3 - Q_1) \qquad \text{(eq. 4)}$$

Where Q_1 and Q_3 are are the first and third quartiles of the absolute residual distribution, and $k = 1{,}5$ is a scaling factor controlling sensitivity. To increase reliability, an anomaly is only confirmed if the threshold violation persists for at least three consecutive samples, consistent with the temporal resolution of the dataset of 5-minute samples. This persistence filter eliminates transient spikes caused by short-term irradiance fluctuations or sensor noise. The residual-based method provides a global statistical perspective, complementing the local adaptability of the envelope approach. The two methods were compared to evaluate their agreement in identifying abnormal points.

3.4 Model Evaluation Metrics

The performance of the proposed linear model was evaluated using standard statistical metrics that quantify accuracy and bias between the measured and modeled AC power. The following indicators were calculated over all valid samples (N) as shown in Equations (5–8).

$$MAE = \frac{1}{N}\sum_1^N |P_{meas,i} - P_{model,i}| \qquad \text{(eq. 5)}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_1^N (P_{meas,i} - P_{model,i})^2} \qquad \text{(eq. 6)}$$

$$R^2 = \left(\frac{Cov(P_{meas}, P_{model})}{\sigma_{P_{meas}} \sigma_{P_{model}}}\right)^2 \qquad \text{(eq. 7)}$$

$$Bias = \frac{1}{N}\sum_1^N (P_{model,i} - P_{meas,i}) \qquad \text{(eq. 8)}$$

where MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) indicate the overall model accuracy, R^2 (coefficient of determination) represents the proportion of variance in the measured data explained by the model, and Bias indicates systematic over- or under-estimation. Lower MAE and RMSE values, combined with a high R^2 (close to 1), demonstrate good agreement between modeled and measured power. Additionally, the relative RMSE (rRMSE) was computed as RMSE normalized by the mean measured power, providing a dimensionless measure of error suitable for performance comparison across different PV systems.

3.5 Identification of Faulty Days

For daily-level analysis, the classification of fault days was performed based on the cumulative distribution of the daily percentage of anomalous points. The empirical cumulative distribution function (CDF) of outlier percentages was computed to establish a data-driven threshold. The threshold was defined at the 99th percentile, meaning that 99% of all days exhibited lower anomaly rates, while the remaining 1% of days were classified as faulty days. The resulting CDF curve is shown in Figure 2 for the bins method, where the shaded region represents the fault-prone domain. The days identified as faulty were then compared against known low-generation events and supported by performance indicators such as the corrected performance ratio (CPR), providing an additional layer of validation for the anomaly detection framework.

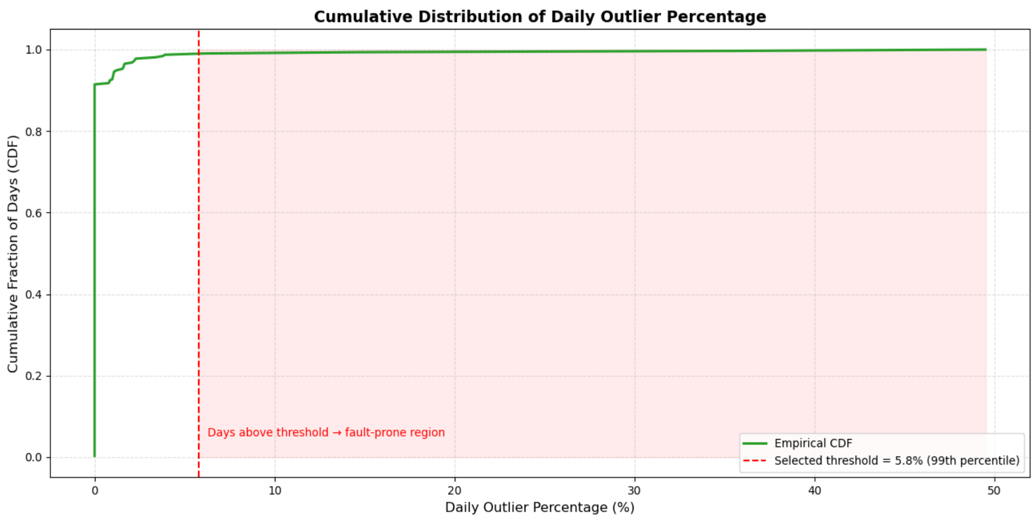

Figure 2: Empirical cumulative distribution function (CDF) of the daily persistent outlier percentage obtained from the envelope (bins) method. The dashed red line marks the 99th percentile threshold (5.8%), used to classify fault days. The shaded region represents the fault-prone domain, where days exhibiting anomalous point percentages above this limit are considered abnormal and subject to further inspection.

To further validate the sensitivity of the detection methods, two controlled shading experiments were conducted under real operating conditions. In one of the tests, one of the PV modules was partially covered with a 50% shading mesh to simulate partial soiling or diffuse obstruction, as shown in Figure 3a. In the second test, approximately four out of six upper cell rows of another module were covered with cardboard, aiming to activate two bypass diodes, as shown in Figure 3b.

Figure 3: Photographic record of the controlled shading experiments conducted on the PV system. (a) October 10, 2025. A 50% mesh was placed over one PV module for an entire day to simulate partial shading conditions and validate the anomaly detection framework. (b) October 16, 2025. A cardboard sheet was applied to the upper section of another PV module to reproduce a localized shading condition.

The results from these controlled tests were included in the analysis and used to assess the responsiveness of both the envelope and residual methods under a known, artificial fault scenario.

4. Results and Discussion

4.1 Data Description and Pre-processing

Before model evaluation, a data quality control procedure was applied to remove periods affected by sensor malfunction or unrealistic irradiance readings. Days presenting inconsistent POA measurements were excluded from the analysis. After this initial filtering, the corrected performance ratio (CPR) was calculated for all valid days, and those within the interquartile range (between the 10th and 90th percentiles) were selected as representative of normal operation. From the original 316 daily samples, 252 days satisfied this criterion and were used for model calibration and evaluation. The global loss factor (η), adjusted using the reference days identified by the CPR criterion, was estimated at 0.9047. This value represents the combined effect of the PV generator's losses. An η value close to 0.9 indicates that the model effectively captures the overall conversion performance of the system under typical operating conditions. It is important to note that η is a dataset-specific calibration parameter rather than a fixed characteristic of the PV system. Therefore, when new data are introduced, such as additional time periods or modified operating conditions, the factor must be re-estimated to maintain model accuracy.

4.2 Proposed Model

Once fitted using the selected reference days, the model was applied to the entire dataset to estimate the expected power for all valid period. Performance metrics were then computed over all days with POA $\geq$ 50 W/m², allowing the assessment of model accuracy under general operating conditions. The statistical indicators, presented in Table 1, were calculated from data averaged at a 5-minute temporal resolution, consistent with the acquisition interval of the micro-inverter measurements.

Table 1: Performance metrics of the linear power model with adjusted global loss factor (η).

Metric	Value
RMSE	40.59 W
MAE	27.52 W
Bias	-1.98 W
R^2	0.947
rRMSE	12.83 %

The performance indicators in Table 1 demonstrate that the proposed linear model provides a good representation of the PV system's behavior. The RMSE and MAE values of 40.59 W and 27.52 W, respectively, correspond to approximately 6.8% and 4.6% of the micro-inverter's nominal AC power (600 W). The relative RMSE of 12.83% indicates that the model can reproduce the system's expected power output with satisfactory precision, considering the variability of environmental conditions and sensor uncertainties. The coefficient of determination (R^2 = 0.947) confirms a strong linear correlation between the modeled and measured values, indicating that the model effectively captures the main physical dependencies of the system on irradiance and temperature. The slight negative bias suggests a minor systematic underestimation of the measured power, possibly caused by small deviations in the assumed inverter efficiency or global losses. Nevertheless, this bias is negligible when compared to the overall magnitude of the system output and does not compromise the accuracy of the model for performance assessment purposes.

4.3 Anomaly Detection Results

After fitting the linear model and computing the expected AC power for all valid time steps, anomaly detection was performed using the envelope (bins) and residual (boxplot) methods described in Section 3.3. Both approaches were applied to the same preprocessed dataset. Figure 4 presents the scatter plot of measured versus modeled power, where the bins method dynamically identifies outliers based on the distribution of data within each bin. For the envelope method, the modeled power range was divided into 20 bins, each representing a distinct irradiance-dependent operating region. To ensure statistical robustness, only bins containing at least 50 valid samples were retained. The red markers represent the bin medians corresponding to normal operating behavior, while purple and orange crosses indicate outliers below and above the expected envelope, respectively.

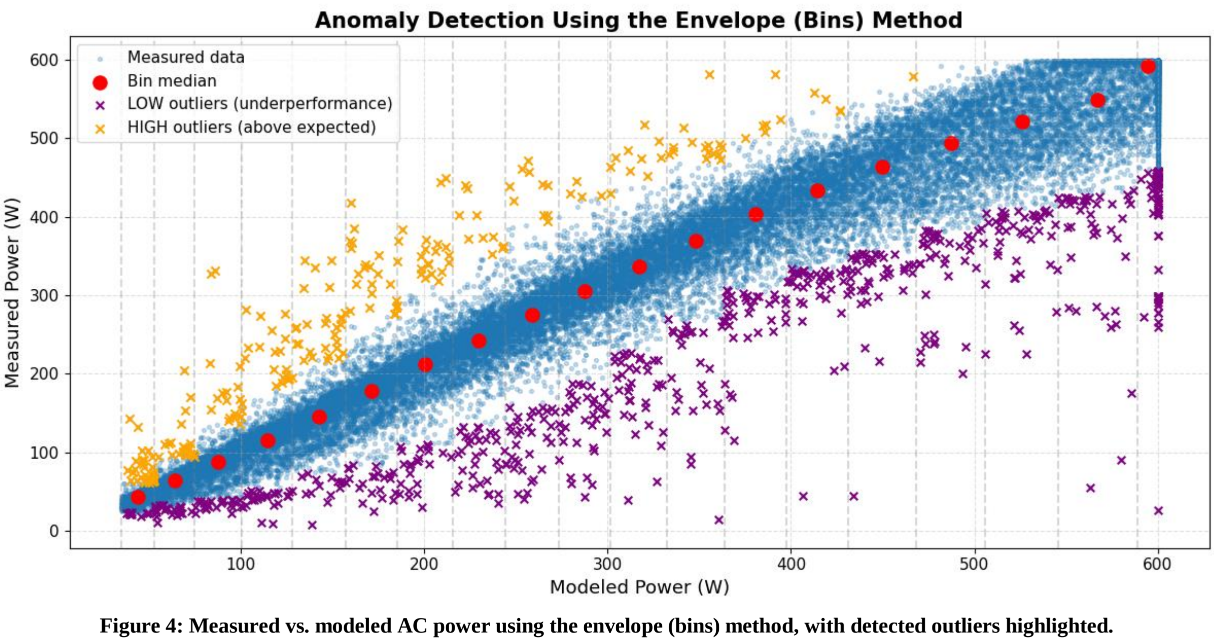

Figure 4: Measured vs. modeled AC power using the envelope (bins) method, with detected outliers highlighted.

To quantify the temporal distribution of anomalies, the percentage of persistent anomalous points was computed for each day. This ranking highlights days with persistent deviations from expected performance, which may indicate sustained faults or operational disturbances. Based on the CDF-derived threshold of 5.8% (corresponding to the 99th percentile), four days were identified as faulty, exceeding the established limit. These results are summarized in

Table 2, which presents the days most affected by persistent deviations from the modeled power curve.

Table 2: Faulty days as detected by bins method.

Date	% Outliers	CPR
31-03-2025	49.5	0.42
10-10-2025	34.1	0.61
24-07-2025	14.85	0.69
30-09-2024	6.12	1.095

A closer inspection of the faulty days reveals distinct operational issues. March 31, 2025, which exhibited the highest percentage of anomalies, clearly corresponds to a fault event where only one PV module was operating correctly. October 10, 2025, as expected, matches the controlled shading experiment, during which one of the PV modules was partially covered with a 50% mesh, resulting in reduced generation. July 24, 2025 showed a localized underperformance period between 08:00 and 11:00 a.m., likely caused by temporary inverter instability, during which only one module remained connected. In contrast, September 30, 2024 presented an atypical pattern with a short interval of power overestimation (CPR = 1.09) and an anomaly rate of 6.12%. This behavior is most likely associated with an irradiance sensor misreading, possibly caused by partial shading of the pyranometer, rather than a true system fault, and can therefore be interpreted as a measurement-related anomaly. The three most representative faulty cases are illustrated in Figures 5, 6, and 7, showing the measured and modeled power curves with the detected outliers highlighted.

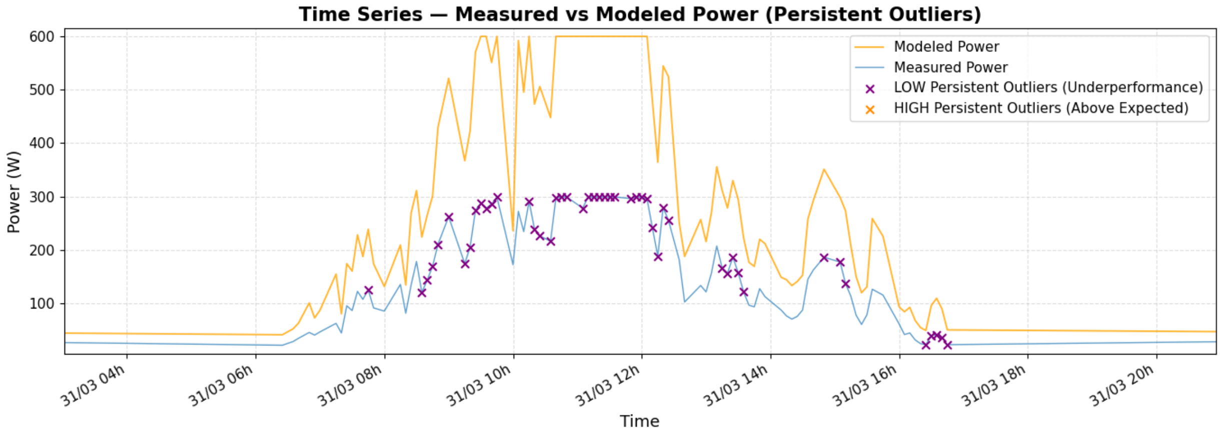

Figure 5: Time series of measured and modeled power for March 31, 2025 with outliers highlighted.

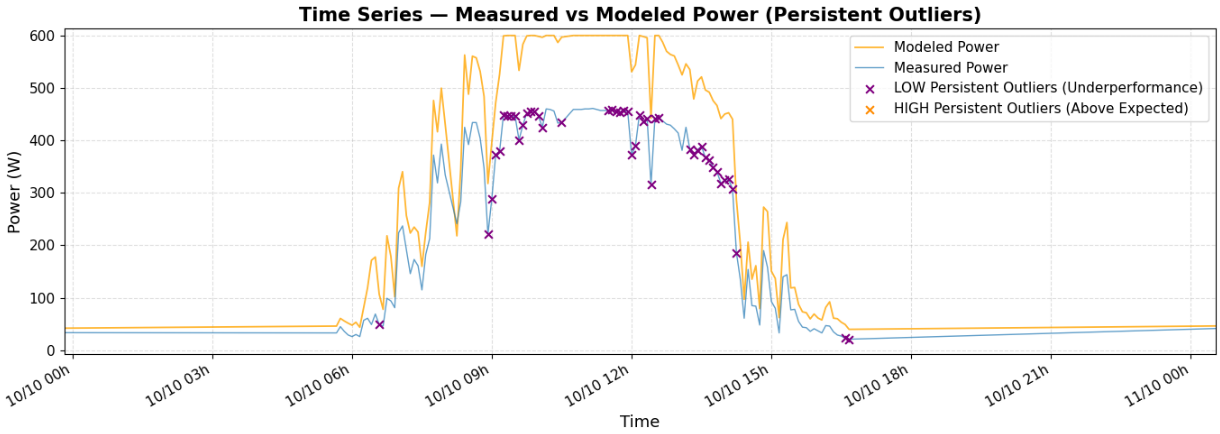

Figure 6: Time series of measured and modeled power for October 10, 2025 with outliers highlighted.

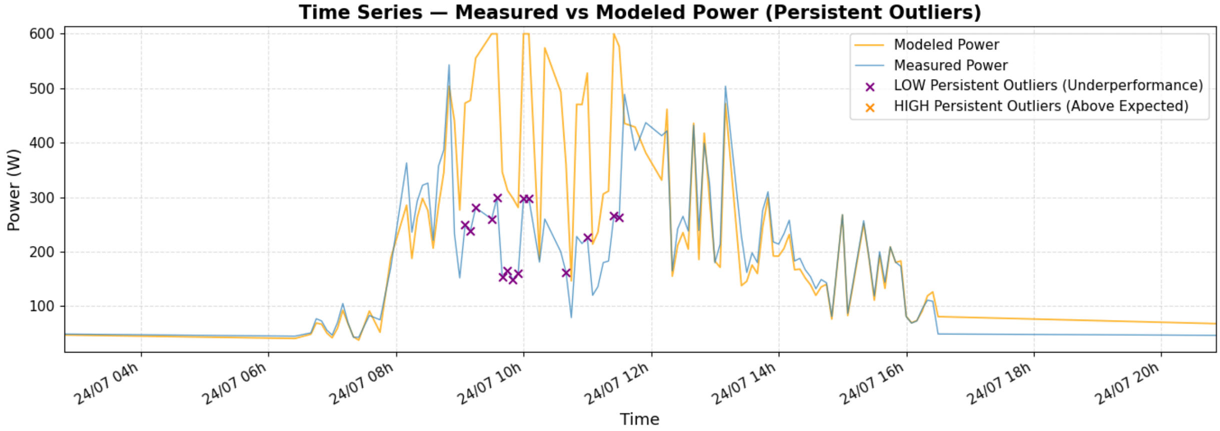

Figure 7: Time series of measured and modeled power for July 24, 2025 with outliers highlighted.

As for the residual-based detection method, the global statistical threshold calculated from the boxplot rule was estimated at 86.3 W, as illustrated in Figure 8.

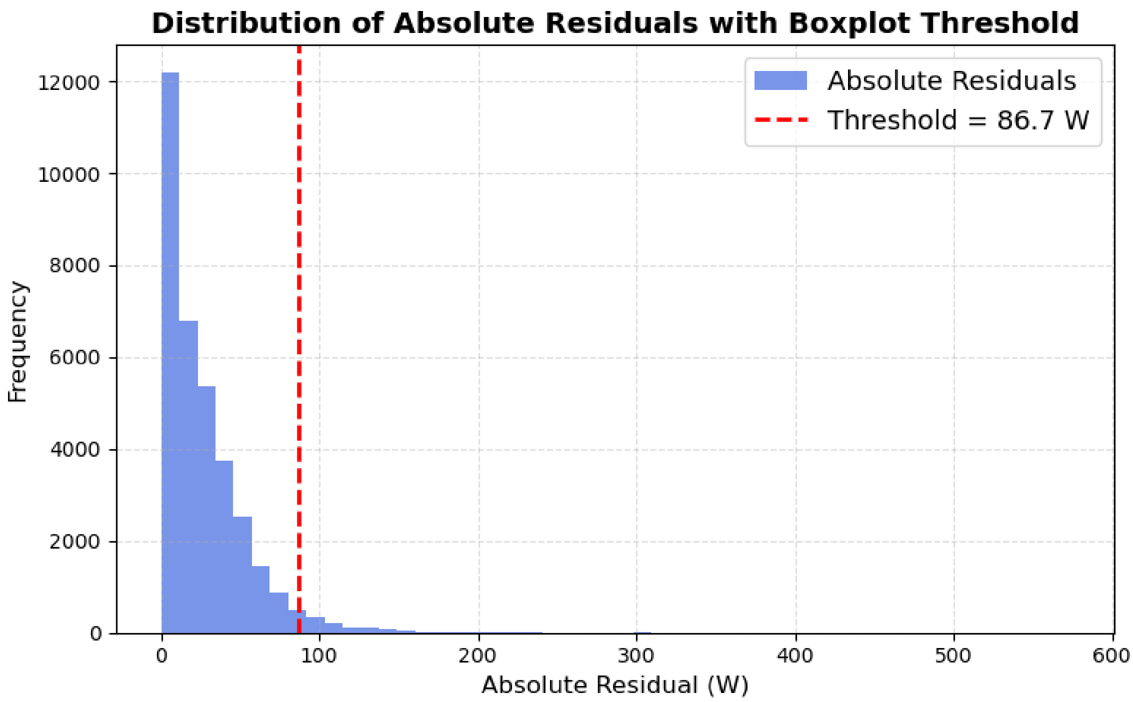

Figure 8: Distribution of absolute residuals and detection threshold derived from the boxplot rule.

To extend the analysis to the daily scale, the cumulative distribution function (CDF) of the daily percentage of persistent anomalies was evaluated. The 99th percentile of this CDF corresponded to 21.4%, which was adopted as the fault-day threshold. This value is higher than the one obtained from the bins method because the latter defines its limits locally within each bin cluster rather than using a global statistical threshold. According to this criterion, four days exceeded the defined limit and were classified as faulty, as shown in Table 3.

Table 3: Faulty days as detected by residual method.

Date	% Outliers	CPR
10-10-2025	56.58	0.42
31-03-2025	54.64	0.61
24-07-2025	24.75	0.69
16-10-2025	21.875	0.65

Among the four faulty days identified, the three highest-ranking days coincided with those detected by the envelope (bins) method. The residual-based approach additionally flagged October 16, 2025, corresponding to the controlled shading experiment in which one PV module was partially covered with a cardboard, simulating a localized shading fault. This event was not captured by the envelope (bins) method because the power deviations occurred intermittently throughout the day, failing to meet the persistence requirement of three consecutive samples for most of the day. Moreover, the adaptive nature of the envelope method, which adjusts its threshold according to the variability within each bin, makes it less responsive to brief or low-intensity anomalies. Conversely, the residual method relies on a global statistical threshold and is therefore more sensitive to subtle deviations distributed over time. The corresponding time series is shown in Figure 9, where the detected anomalies are highlighted over the modeled and measured power curves by the residual method.

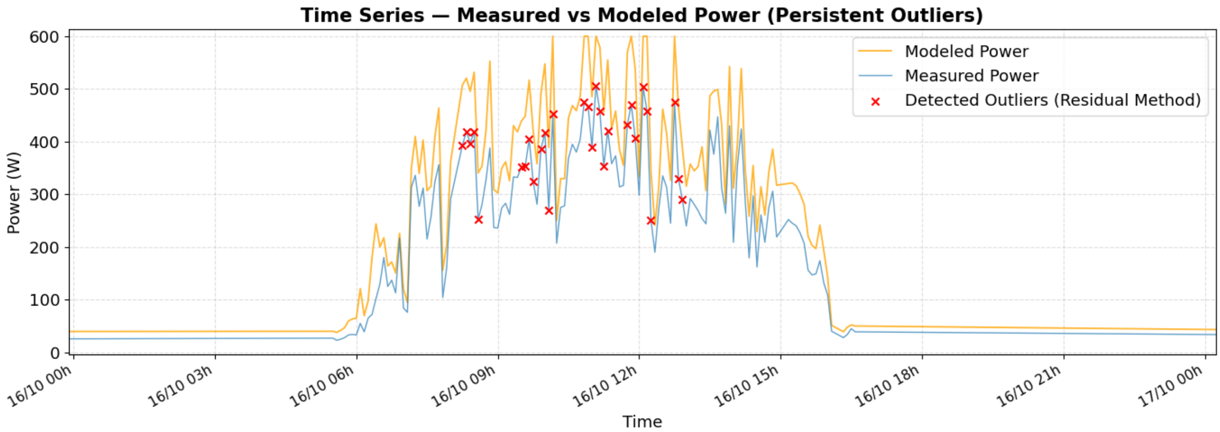

Figure 9: Time series of measured and modeled power for October 16, 2025 with outliers highlighted.

5. Conclusions

The proposed framework provides a simple and robust approach for detecting generation anomalies in grid-connected photovoltaic systems using standard monitoring data. Overall PV performance was satisfactory, with faulty days representing a small fraction of the total analyzed period (≈1.9%), confirming that the system operated normally for most of the time. Both implemented methods were effective in identifying periods of abnormal operation, including known low-generation events and controlled shading experiments. The envelope method demonstrated a conservative detection profile, focusing on consistent and persistent deviations across irradiance ranges. This conservative nature effectively minimized false alarms but led to the omission of one verified shading fault, indicating less sensitivity to short-duration anomalies. In complement, the residual-based method exhibited higher sensitivity to subtle and intermittent deviations. A key advantage of the proposed methodology lies in its low computational cost, interpretability, and independence from large historical datasets. These characteristics make it particularly suitable for small and medium-scale distributed systems. Future work should focus on integrating this detection framework into an automated real-time monitoring pipeline, enabling continuous system supervision and early fault alerts. Expanding the approach with machine learning algorithms or adaptive statistical thresholds could improve its robustness under varying climatic and operational conditions. Additionally, the methodology could be integrated with satellite-derived irradiance data to extend its applicability to PV systems that operate without on-site measurement equipment.

Acknowledgments

Authors would like to thank the PRH 48.1 program for supporting this research, and the FAE-UFPE laboratory and technician Rinaldo for their assistance with sensor installation, data acquisition, and experimental setup.

References

BARBOZA, Lucas Emanuel Almeida. Elaboração de um modelo de identificação de falhas e análise de confiabilidade para sistemas de rastreamento solar em usinas fotovoltaicas. 2023.

BRENNAND, Leonardo José de Petribú. Detecção e isolamento de falhas em aerogeradores utilizando dados do sistema SCADA. 2022. Dissertação (Mestrado em Tecnologias Energéticas e Nucleares) – Universidade Federal de

Pernambuco, Recife, 2022.

DIERAUF, Timothy et al. Weather-corrected performance ratio. National Renewable Energy Laboratory (NREL), Golden, CO (United States), 2013.

FIRTH, Steven K.; LOMAS, Kevin J.; REES, Simon J. A simple model of PV system performance and its use in fault detection. Solar Energy, v. 84, n. 4, p. 624–635, 2010.

KING, D. L.; BOYSON, W. E.; KRATOCHVIL, J. A. Photovoltaic Array Performance Model. SAND2004-3535. Albuquerque, NM: Sandia National Laboratories, August 2004.

MADETI, Siva Ramakrishna; SINGH, S. N. Modeling of PV system based on experimental data for fault detection using kNN method. Solar Energy, v. 173, p. 139–151, 2018.

MIRAFTABZADEH, Seyed Mahdi et al. Anomaly detection in photovoltaic systems via deep learning autoencoder. In: 2023 International Conference on Smart Energy Systems and Technologies (SEST). IEEE, 2023. p. 1–6.

MIRANDA, Diego et al. Quality Assurance Procedure for Solar Radiation at Minute Resolution. 2023.

NARVARTE, L.; LORENZO, E. Tracking and ground cover ratio. Progress in Photovoltaics: Research and Applications, v. 16, n. 8, p. 703–714, 2008.

PLATON, Radu et al. Online fault detection in PV systems. IEEE Transactions on Sustainable Energy, v. 6, n. 4, p. 1200–1207, 2015.

QU, Jiaqi et al. An unsupervised hourly weather status pattern recognition and blending fitting model for PV system fault detection. Applied Energy, v. 319, p. 119271, 2022.

TAGHEZOUIT, Bilal et al. Model-based fault detection in photovoltaic systems: A comprehensive review and avenues for enhancement. Results in Engineering, v. 21, p. 101835, 2024.

VASISHT, M. Shravanth; SRINIVASAN, J.; RAMASESHA, Sheela K. Performance of solar photovoltaic installations: Effect of seasonal variations. Solar Energy, v. 131, p. 39–46, 2016.

WANG, Jing-Yi et al. Performance assessment of photovoltaic modules based on daily energy generation estimation. Energy, v. 165, p. 1160–1172, 2018.