IMPROVING SOLAR DATA QUALITY IN RIO GRANDE DO NORTE STATIONS THROUGH NIGHTTIME ANOMALY DETECTION TESTS

João Vítor Fonseca de Mendonça[1], Samira de Azevedo Santos Emiliavaca [1], Ana Cleide Bezerra Amorim[1], Alan Rodrigues de Sousa[1], Nicholas Medeiros Lopes[1], André Luiz de Oliveira[1], Gabriela Rocha Alves da Silva[1]

[1] SENAI Innovation Institute for Renewable Energy, Natal (Brazil)

Abstract

This work presents a data validation strategy to improve the quality of solar radiation measurements from stations located in Rio Grande do Norte, Brazil. We focus on nighttime anomaly detection using physical limit and rare extreme tests. A real-world case in the Pau dos Ferros station is discussed, where ants damaged the datalogger input channel, introducing a permanent offset that corrupted both nighttime and daytime readings. By placing the physical and statistical anomaly tests at the beginning of the validation workflow, the proposed methodology enabled early fault detection and prevented its influence on downstream analyses. The study also introduces a modification to the original test structure, improving its sensitivity to rare events. During the presentation, we briefly outline the validation workflow and highlight how this adjustment enhances detection efficiency. Overall, this procedure improves the consistency and reliability of solar monitoring networks, providing a stronger foundation for solar energy modeling, forecasting, and performance assessment in Rio Grande do Norte and other regions.

Keywords: solar radiation, data quality control, solar resource assessment, measurement validation

Introduction

In Brazil, solar resources are abundant throughout the entire national territory, with solar irradiation levels exceeding those observed in most European Union countries such as Germany, France and Spain, nations where solar energy utilization is already well established (Pereira *et al*., 2006). The northeastern region of Brazil stands out with the highest irradiation values per square meter of area, and the state of Rio Grande do Norte, in particular, presents an estimated energy production potential of approximately 213 TWh per year (DO NORTE, RIO GRANDE et al., 2022).

The variability and potential of the solar resource must be reliably known in order to minimize uncertainties and investment risks associated with small- and large-scale photovoltaic projects (Da Silva *et al*., 2024). Obtaining this knowledge depends directly on the installation and operation of solarimetric stations dedicated to acquiring solar irradiation data. However, beyond the quality of the instruments employed, it is essential to ensure the proper installation of equipment as well as the performance of regular maintenance. At the same time, the implementation of a rigorous data quality control process is fundamental to guarantee the representativeness of the observations under local meteorological conditions and to enable comparative analyses with different datasets referring to the same location.

The application of quality control tests to solarimetric data makes it possible to identify suspicious, anomalous, or erroneous values, ensuring that the measurements remain within physical limits and predefined quality criteria. This practice is essential for the reliable operation of solarimetric stations, as it helps detect sensor failures, equipment degradation, or external interferences that could compromise the integrity of the collected data. Continuous application of quality tests also enables maintenance teams to be promptly deployed for corrective actions, reducing the duration of inconsistencies and minimizing data gaps (WMO, 1992; Shafer *et*

10.18086/swc.2025.04.28 Available at http://proceedings.ises.org

al., 2000; Graybeal *et al*., 2004; Estévez *et al.*, 2011).

The growing availability of automated and quality-assured datasets has reinforced the importance of implementing standardized, reproducible, and scalable quality control routines (Quan et al., 2024). Recent advances in data processing have enabled the automation of these tests, allowing real-time detection of anomalies and significant reductions in manual intervention. Such approaches not only enhance the temporal consistency of large-scale solar databases but also facilitate the integration of quality-controlled data into forecasting models and energy planning tools.

In this context, this study employs an automated quality control pipeline for solarimetric data that integrates physical limit tests, consistency checks, and rare outlier detection. These tests help identify sensor failures at an early stage, ensuring higher data quality and preventing long-term distortions in solar datasets.

2. Materials and methods

2.1 Data Quality x Quality Assurance

Reliable observational data are fundamental for parameterization, calibration, and validation processes in solar energy modeling. Ensuring such reliability requires not only the use of high-precision instruments but also the implementation of systematic data quality assessment routines. Within this context, quality assurance and data quality control play complementary roles in maintaining the integrity of the measurements and the consistency of the long-term datasets used for analysis and forecasting.

Quality assurance encompasses post-processing procedures conducted after the full acquisition of the dataset. Its objective is to verify the overall consistency, completeness, and physical plausibility of historical time series, identifying possible long-term drifts, calibration errors, or systematic biases (Jiménez *et al.*, 2010). This phase is essential for establishing a trustworthy baseline, particularly in retrospective analyses where sensor degradation or configuration changes may have occurred.

In contrast, data quality control refers to the continuous, real-time monitoring performed during measurement campaigns. It involves detecting and flagging anomalous or inconsistent data as they are recorded, allowing prompt operational responses to issues such as equipment malfunction or sensor contamination (Doraiswamy et al., 2000). By applying automated tests, such as physical limit checks and temporal consistency analyses, it becomes possible to identify outliers and hardware failures early, ensuring that the data feeding predictive models remain within acceptable quality bounds (Graybeal et al., 2004; Estévez et al., 2011).

Together, these two components form a continuous improvement cycle: quality assurance validates historical datasets to establish reference standards, while quality control maintains the accuracy of ongoing measurements. In this study, quality assurance procedures were first applied to evaluate archived data from Rio Grande do Norte stations, serving as a diagnostic step for identifying systematic issues. Subsequently, quality control routines were implemented operationally to detect failures in near real-time, contributing to the robustness of the solar monitoring network and improving the performance of the forecasting system.

2.2 Data

The solar radiation data used in this study were collected from meteorological stations located in the state of Rio Grande do Norte, Brazil. Each station is equipped with a CMP22 pyranometer (Kipp & Zonen, Delft, The Netherlands), a secondary standard instrument compliant with ISO 9060, connected to a datalogger (Campbell Scientific Inc., Logan, Utah, USA). Measurements are collected every second and subsequently integrated into one-minute averages to reduce noise and facilitate processing and storage. Although the infrastructure is designed for reliable operation, environmental and physical interference can still affect data integrity.

The stations are geographically distributed across five cities representing a range of climatic and geographic conditions in Rio Grande do Norte: Jandaíra, Lajes, Pau dos Ferros, Mossoró, and Natal. These locations encompass both coastal and semi-arid regions, providing a representative dataset of the state's solar resource variability.

A notable incident occurred at the Pau dos Ferros station, where ants infiltrated the equipment enclosure and damaged the datalogger's input circuitry. This caused a permanent offset in the irradiance readings, compromising the reliability of the data from that location. The extent of the damage and the presence of the ants can be observed in Figure 1, which shows the affected datalogger and the insect infestation.

Following this incident, a new quality process was established, which represented an important step toward improving the overall performance of solarimetric data.

Figure 1: Ant infestation inside the datalogger at the Pau dos Ferros station.

The meteorological station installed in Pau dos Ferros (DENOCS) is equipped with a comprehensive set of sensors and instruments designed to monitor local atmospheric and solar conditions (see figure 2). Among these devices are a pyranometer for measuring global horizontal irradiance, a thermohygrometer for temperature and humidity, an anemometer for wind speed, and additional components for power and data acquisition. The figure 3 provides a detailed overview of the instruments and their specifications

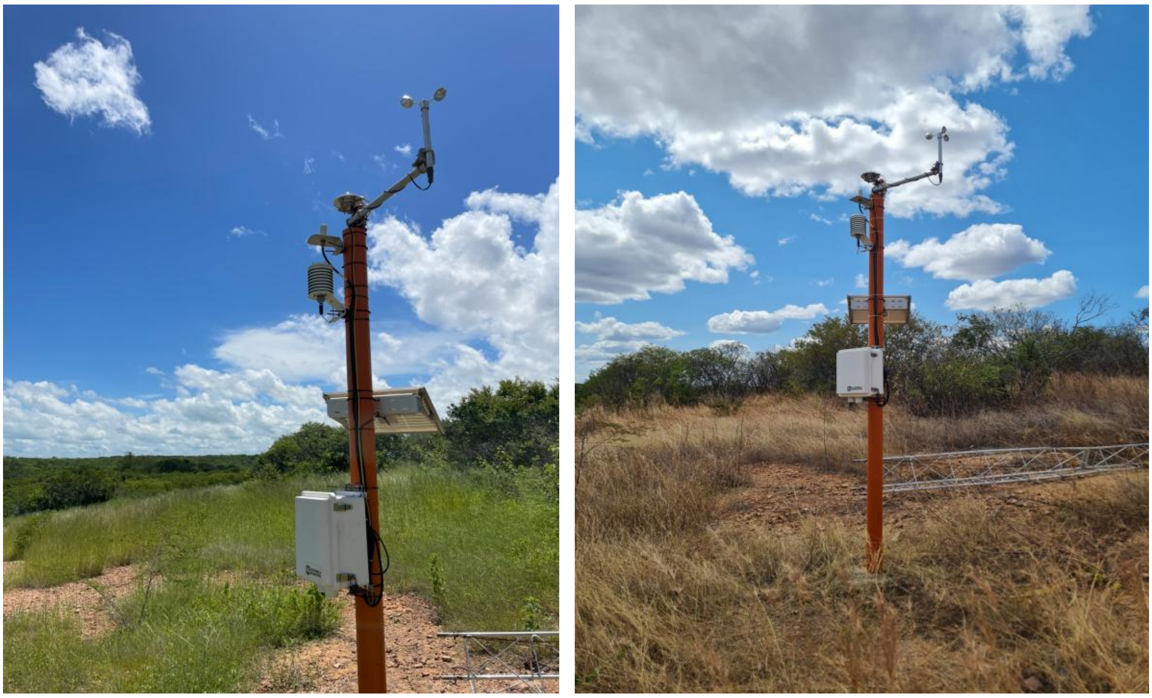

Figure 2: Solarimetric station RNES03 - Pau dos Ferros/RN..

STATION DATA					
Location	Pau dos Ferros – RN (DENOCS)			**Acronym Station**	RNES03
Latitude	6°8'38.4031"S	-6,144001°	9320809.47	**Altitude**	199 MSL
Longitude	38°11'25.5772" O	-38,190438°	589572.90	**Magnetic Declination**	21° West
Installation Date		02/06/2021		**Date of Preparation**	07/11/2024

IDENTIFICATION OF INSTRUMENTS					
Instruments	**Measured variable/Function**	**Acronym**	**Unit**	**Series**	**Note**
Datalogger CR300-CELL215 – Campbell Scientific	Data acquisition	-	-	26824	-
Pyranometer CPM 6 – Kipp & Zonen	Global Horizontal Irradiance	GHI	W/m²	214238	Sensitivity 19,14 µV/W/m² Calibrated on 19/01/2021
Thermohygrometer Hygro VUE10 – Campbell Scitenfic	Air temperature	Temp	°C	E2191	Calibrated on 03/02/2021
	Relative humidity	UR	%		
Anemometer Wind SPD Sensor – Met One Instruments	Wind speed	Vel	m/s	B11083	Calibrated on 02/02/2021
Photovoltaic module 20M-V – Electrical Raiting	Power generation	-	W	002020M190316065127	Power 20W
Unidirecional Antenna TRA6927M3PW-001 - Laird	Amplify 4G Signal	-	-	07202010	-

Figure 3: Pau dos Ferros station instrumentation.

2.3 Overview of Applied Validation Tests

The use of quality indicators, or flags, in data assurance and control procedures is important for defining the confidence level of each record in a time series (Petribu et al., 2017). In most cases, the data are not corrected but flagged according to their reliability. This approach is widely adopted by international meteorological networks such as the World Radiation Monitoring Center (WRMC), the Oklahoma Mesonet (MESONET), and the Meteorological Resource Center (WEBMET), as well as by Brazil's National Organization System for Environmental Data (SONDA), developed by the National Institute for Space Research (INPE).

Unlike binary systems that classify data simply as good or suspect, the use of multiple flags allows a more detailed description of data quality. For example, SONDA employs four indicators; MESONET uses eight, and WEBMET applies seven. The methodology originally adapted by Silva, Emiliavaca et al. (2024) introduced a structured sequence of tests based on these principles. However, the quality control procedure adopted in the present study was updated and refined to address specific requirements and limitations identified in previous applications.

In this revised approach, the results of each test correspond to columns in a flag matrix, following the structure initially proposed by Moraes (2015). Each row represents a timestamp in the time series, and each cell contains a flag value indicating the outcome of the respective test.

Data that pass all tests are labeled as good (flag 1). If a test yields an uncertain result, the data are marked as suspect (flag 2). When a suspect value is not confirmed as erroneous in subsequent tests, predefined criteria may automatically reclassify it tas anomalous (flag 3), depending on the proportion of suspect results. In such cases, users may decide whether to include or discard the data in their analysis. Once a data point is marked as anomalous, it is labeled as previously anomalous (flag 4) in all remaining tests.

Some data cannot be evaluated by certain tests, for example, solar radiation values during nighttime, and are thus marked as not tested (flag 5). When data are missing for a given timestamp, they are classified as not available (flag 6).

The complete quality assurance process, summarized in the flowchart presented in Figure 2, includes four categories of tests: global, physical, refinement, and comparative respectivaly. Together, these procedures form an updated and consistent framework for evaluating and validating the quality of meteorological data.

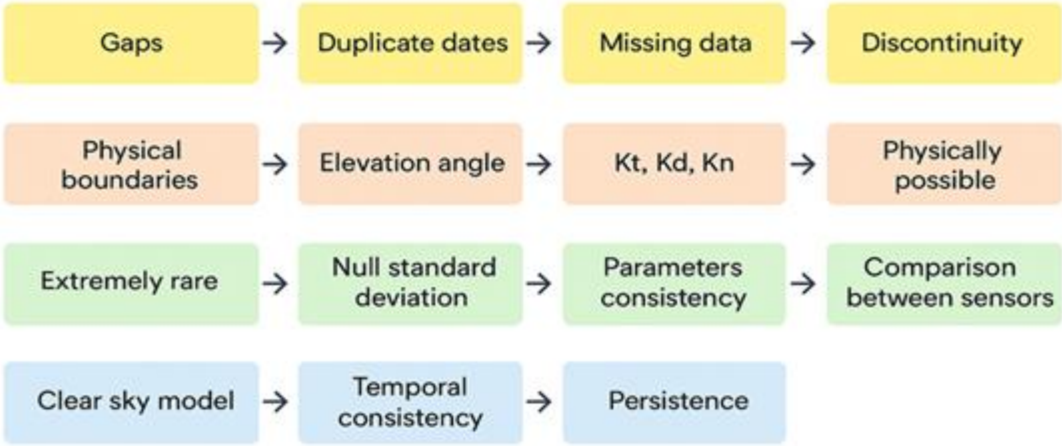

Figure 4: Test flowchart of the quality assurance procedure for solarimetric observational data used in ISI-ER.

2.3.1 Global Tests

The first stage of the quality control procedure corresponds to the global tests, represented by the yellow boxes in Figure 2. These tests are applied to the entire time series to identify general inconsistencies such as discontinuities, missing data, temporal gaps, or duplicated timestamps. Their main goal is to verify the structural integrity and completeness of the dataset before applying more specific local tests. When critical flaws are detected, such as extended gaps or date and time errors, the affected portions of the series may be segmented or excluded from subsequent analyses (EPE, 2021).

Among these tests, the gap and discontinuity analyses detect interruptions that exceed the expected acquisition interval, while the checks for duplicate timestamps and missing values ensure temporal coherence (Petribú et al., 2017). Missing intervals are typically filled with NaN values to preserve the chronological order and prevent distortions in phase or frequency. According to EPE (2021) guidelines, datasets with long or continuous data losses, such as gaps exceeding fifteen days, should be subdivided into independent segments or discarded. Together, these procedures provide an initial assessment of dataset completeness and reliability, ensuring that only consistent and representative time series advance to the next validation stages.

2.3.2 Local Tests

The second stage of the quality control procedure correspond to local tests. These tests, which include the physical, comparative, and refinement groups, are applied to individual samples or subsets of the time series. Their main objective is to verify the internal consistency and physical plausibility of the measurements, either through empirical thresholds derived from sensor behavior or through statistical relationships between correlated variables (EPE, 2021). Unlike the global tests, which focus on the overall structure of the dataset, the local tests assess the reliability of each data point in detail.

The physical boundaries test is the first verification step for solar irradiance data, applied primarily to Global Horizontal Irradiance (GHI) and Direct Normal Irradiance (DNI). It ensures that recorded values fall within physically possible ranges under normal atmospheric conditions. Measurements between approximately −5 and 2000 W/m² are considered plausible, while values outside this range are flagged as anomalous. A complementary test based on the solar elevation angle discards irradiance measurements taken when the Sun's elevation is below 10°, a range where sensor uncertainty is typically high (Maxwell et al., 1993). These criteria prevent nighttime or low-angle errors from propagating through subsequent analyses.

The transmissivity index test evaluates the ratio between global irradiance (GHI) and extraterrestrial irradiance (I_0), normalized by the cosine of the solar zenith angle (θz). This dimensionless index, commonly referred to as *kt*, is used to identify outliers and unrealistic measurements influenced by cloud cover, reflection, or calibration issues. Studies such as Younes et al. (2005) and Raichijk (2012) have proposed different threshold ranges for this parameter depending on sensor type and regional conditions. For the present work, the limits were adjusted to account for the higher irradiance variability typically observed in northeastern Brazil.

$$k_t = \frac{GHI}{I_0 \cos \Theta z} \qquad \text{(eq. 1)}$$

The Physically Possible test sets upper and lower bounds for Global Horizontal Irradiance (GHI), Direct Normal Irradiance (BNI), Diffuse Horizontal Irradiance (DHI), and Ground-Reflected Irradiance (GRI), all derived from the extraterrestrial irradiance (I_0) and the solar zenith angle (Θ_z). The equation applied in this procedure to GHI are presented in (eq. 2).

$$-4 \leq GHI \leq I_0 \times 1.5 \times \cos^{1.2}(\Theta z) + 100 \qquad \text{(eq. 2)}$$

In addition, the Extremely Rare test applies similar upper-bound expressions but with more restrictive coefficients, used to flag values that, while physically possible, occur infrequently under normal atmospheric conditions. These limits in GHI are shown in (eq. 3).

$$-2 \leq GHI \leq I_0 \times 1.2 \times \cos^{1.2}(\Theta z) + 50 \qquad \text{(eq. 3)}$$

The next group, known as comparative and statistical consistency tests, examines the coherence among measurements from different sensors and derived parameters. For example, irradiance data obtained from two pyranometers should exhibit a correlation above 0.95, ensuring sensor agreement. In addition, the standard deviation and parameter consistency tests verify that the mean value lies between the recorded minimum and maximum, and that the standard deviation is nonzero. These conditions confirm that the datalogger correctly captured short-term variability within each averaging period.

And the clear-sky model, temporal consistency, and persistence tests refine the dataset by evaluating the temporal behavior of the series. The clear-sky model establishes an upper boundary for irradiance under cloud-free conditions, and values exceeding approximately 10–14 % above the theoretical limit are flagged as overirradiance events (Petribú et al., 2017). The temporal consistency test identifies unrealistic variations between consecutive samples, such as irradiance changes greater than 1000 W/m² within one minute. The persistence test detects sequences of constant readings lasting more than 20 minutes, which may indicate sensor malfunction, shading, or data acquisition failure. Together, these procedures ensure that the dataset is physically realistic, internally consistent, and suitable for subsequent statistical or modeling applications.

3. Results and Discussion

3.1 New process of quality

Initially, consistency checks based on solar position and empirical models, such as the elevation angle, clearness index (Kt), and diffuse index (Kd), were applied to the data. However, this sequence proved ineffective in specific cases, particularly at the Pau dos Ferros station, where physical damage caused by ants introduced a persistent sensor offset. Due to the reliance on empirical thresholds, this anomaly went undetected, leading to erroneous readings being classified as valid, especially during overirradiance periods.

To address this, we prioritized detecting physical anomalies, such as sensor offsets, allowing for earlier identification of faulty data, especially during nighttime when irradiance should be near zero. In the Pau dos Ferros case, this approach flagged the sensor offset, removing erroneous values before they impacted analysis. Figure 4 shows how the validation tests detected the offset on 06/17/2024, preventing its effect on the daily analysis.

The validation strategy in this study includes two main quality checks: physical plausibility tests, which ensure irradiance values remain within acceptable limits (e.g., rejecting nonzero values during nighttime), and rare event tests, which detect abrupt deviations in irradiance that are highly unlikely under normal conditions. Figure 5 shows the revised validation pipeline, with these tests now placed at the beginning of the process.

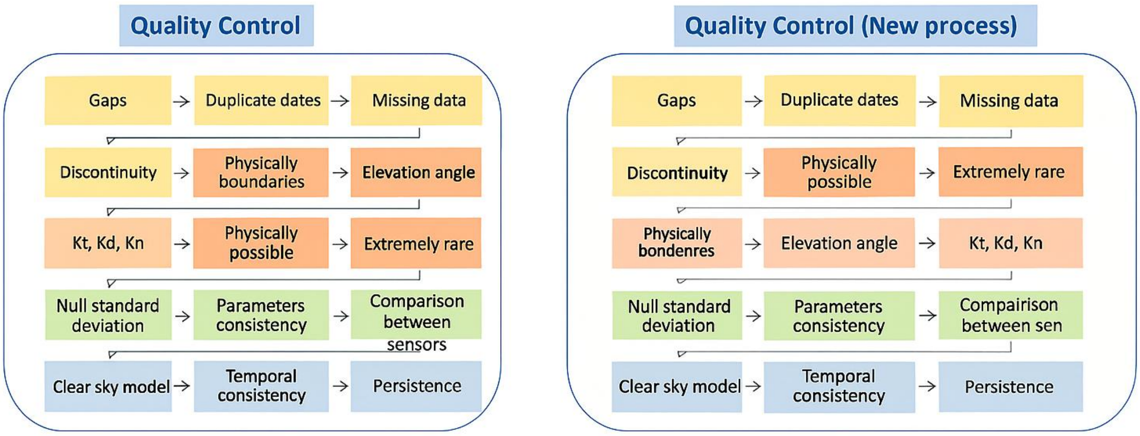

Figure 5: Validation pipeline before and after workflow adjustment.

To address this, we prioritized detecting physical anomalies, such as sensor offsets, allowing for earlier identification of faulty data, especially during nighttime when irradiance should be near zero. In the Pau dos Ferros case, this approach flagged the sensor offset, removing erroneous values before they impacted analysis. Figure 5 shows how the validation tests detected the offset on 06/17/2024, preventing its effect on the daily analysis.

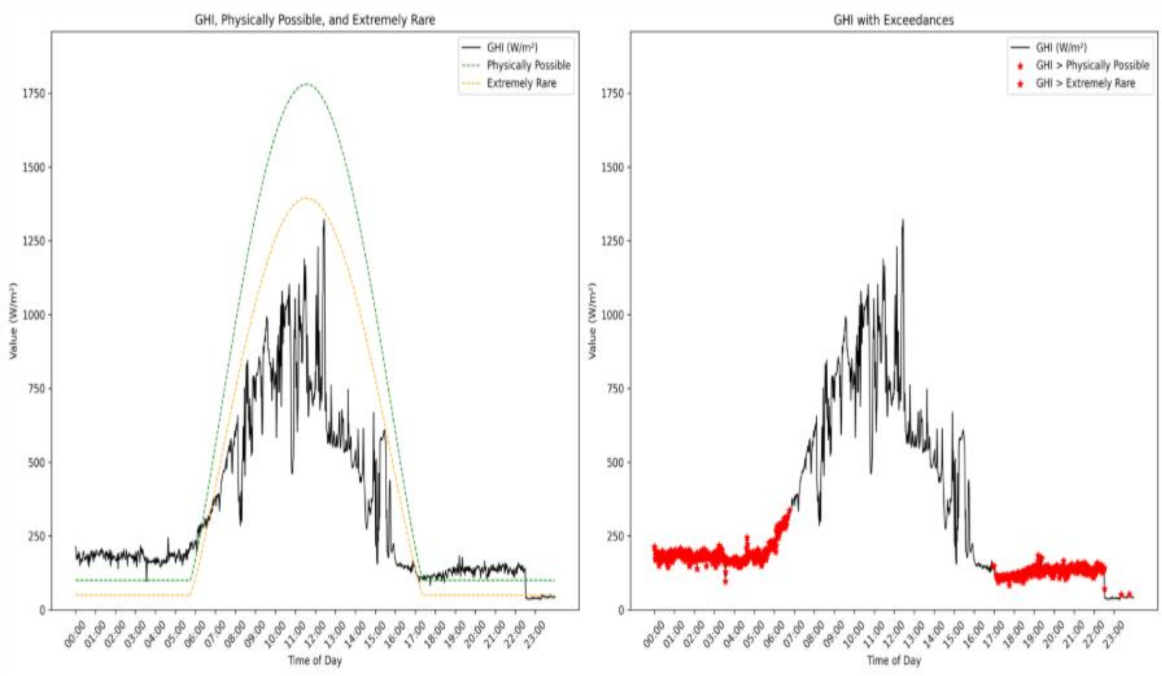

Figure 6: Anomaly Detection at the Pau dos Ferros Station (06/17/2024)

To overcome this limitation, we revised the validation workflow to prioritize the physical and statistical rarity tests. This new sequence enables early detection of clear physical faults—such as the constant offset in Pau dos Ferros—before applying model-based consistency checks. As a result, rare but physically plausible events are preserved, while genuinely erroneous values are more effectively filtered.

4. Conclusions

Ensuring the quality of solar radiation data is fundamental for accurate modeling and optimization of solar energy systems. This study demonstrates that applying physical limit and rare value tests to nighttime irradiance data can effectively detect anomalies and equipment failures. The case of a solarimetric station located in the city of Pau dos Ferros, in northeastern Brazil, highlights the importance of automated quality control, allowing for the early identification of suspicious data and improving the overall integrity of the data. This methodology increases the reliability of measured data and contributes to more efficient maintenance and a more accurate assessment of solar resources in Rio Grande do Norte and other regions. Furthermore, it provides valuable insights for the solar resource prospecting process and for evaluating the performance of operating photovoltaic systems.

5. Acknowledgments

This research was carried out as part of the activities of the National Institute for Technological Development of Photovoltaic Solar Energy Applications (INCT-DAESF), funded by the Brazilian agencies CNPq and MCTI under process number 408486/2024-4. The authors would like to thank the SENAI Institute for Innovation in Renewable Energy for its support and infrastructure; CTG Brasil, for its financial contribution to the ANEEL PD-10381-0620/2020 project; and TotalEnergies, for its current financial contributions to the ANP 23980-6 and ANP 24118-2 projects. The authors would like to thank ISI-ER for their support in the development of this work and TotalEnergies for the current financial support under the ANP PD-24118-2 project.

6. References

Pereira, E.B., et al., 2006. Atlas brasileiro de energia solar. São José dos Campos: INPE, 60 p. ISBN 978-85-17-00030-0.

WMO, 1992. Guide on the global data-processing system. WMO-No. 485, Geneva, 305 p.

Shafer, M.A., Fiebrich, C.A., Arndt, D.S., 2000. Quality assurance procedures in the Oklahoma mesonetwork. Journal of Atmospheric and Oceanic Technology 17, 474–494

Graybeal, D.Y., Degaetano, A.T., Eggleston, K.L., 2004. Improved quality assurance for historical hourly temperature and humidity: development and application to environmental analysis. Journal of Applied Meteorology 43, 1722–1735.

Estévez, J., Gavilán, P., Giráldez, J.V., 2011. Guidelines on validation procedures for meteorological data from automatic weather stations. Journal of Hydrology 402, 144–154.

Do Norte, Rio Grande, et al., 2022. Atlas eólico e solar do Estado do Rio Grande do Norte.

Da Silva, L.S., et al., 2024. Procedimento para qualidade de dados solarimétricos em estações padrão EPE com medição de albedo. In: X Congresso Brasileiro de Energia Solar (CBENS), 2024, Natal. Anais, Natal: CBENS.

Gueymard, C.A., 2012. Clear-sky irradiance predictions for solar resource mapping and large-scale applications: improved validation methodology and detailed performance analysis of 18 broadband radiative models. Solar Energy 86(8), 2145–2169. https://doi.org/10.1016/j.solener.2011.11.011

Costa, M.E.N.R. da, 2024. Comparative analysis of overirradiance estimation at different temporal scales for cities in Rio Grande do Norte. Undergraduate Thesis (Meteorology) – Federal University of Rio Grande do Norte, Natal, Brazil. Available at: https://repositorio.ufrn.br/handle/123456789/58900

Kipp & Zonen, 2015. CMP22 Pyranometer Instruction Manual. Delft, The Netherlands: Kipp & Zonen B.V. Available at: https://www.kippzonen.com

Mierzwiak, M., Kroszczyński, K., Araszkiewicz, A., 2023. WRF parameterizations of short-term solar radiation forecasts for cold fronts in Central and Eastern Europe. Energies 16(13), 5136. https://doi.org/10.3390/en16135136

Quan, W., et al., 2024. A quality-assured dataset of nine radiation components. Earth System Science Data 16, 961–983.

Da Silva, L.S., Emiliavaca, S.D.A.S., de Sousa, A.R., de Souza, V.R.S., de Oliveira, J.V., do Nascimento, B., de Lira, A.L.O., de Araujo, P.A.A., 2024. Procedimento para qualidade de dados solarimétricos em estações padrão EPE com medição de albedo. In: Proceedings of the X Congresso Brasileiro de Energia Solar (CBENS), Natal, Brazil, 2024.

Shafer, M.A., Fiebrich, C.A., Arndt, D.S., 2000. Quality assurance procedures in the Oklahoma Mesonetwork. Journal of Atmospheric and Oceanic Technology 17, 474–494.