Fault Detection In Solar Power Plants Using Energy Production Data

Dominykas Vilčinskas Lukas Voveris Jolita Bernatavičienė

Institute of Data Science and Digital Technologies, Vilnius University

Introduction

Solar power is becoming more common due to growing concerns about climate change and sustainability. However, solar power plants face various challenges such as adverse weather conditions, component failures and shading. Timely fault detection is crucial, as even minor issues can lead to significant energy losses. By applying statistical and machine learning methods, anomalies in a solar power plant's power generation can be identified, aiding in detection of malfunctioning component early. This allows for timely repairs, optimizing plant performance and improving energy efficiency.

Methodology

During data preprocessing, 16 key features were extracted from each string's time series data to represent the global structure of the data. The extraction process resulted in a transformed dataset, where each time series is represented as an object with 16 features, enabling for more effective analysis. Statistical and machine learning techniques — including PCA + α-HULL [1], Isolation Forest (iForest) [2], and Local Outlier Factor (LOF) [3] — were employed to identify systems exhibiting abnormal behavior. The results demonstrate that a combination of these methods can help effectively identify outliers, with a combined anomaly score providing a comprehensive assessment of string performance. Additionally, RANSAC and DBSCAN methodology [4] was used to construct fault profiles, which enabled a more in-depth analysis of each system's performance and provided further confirmation of previously identified systems exhibiting abnormal behavior.

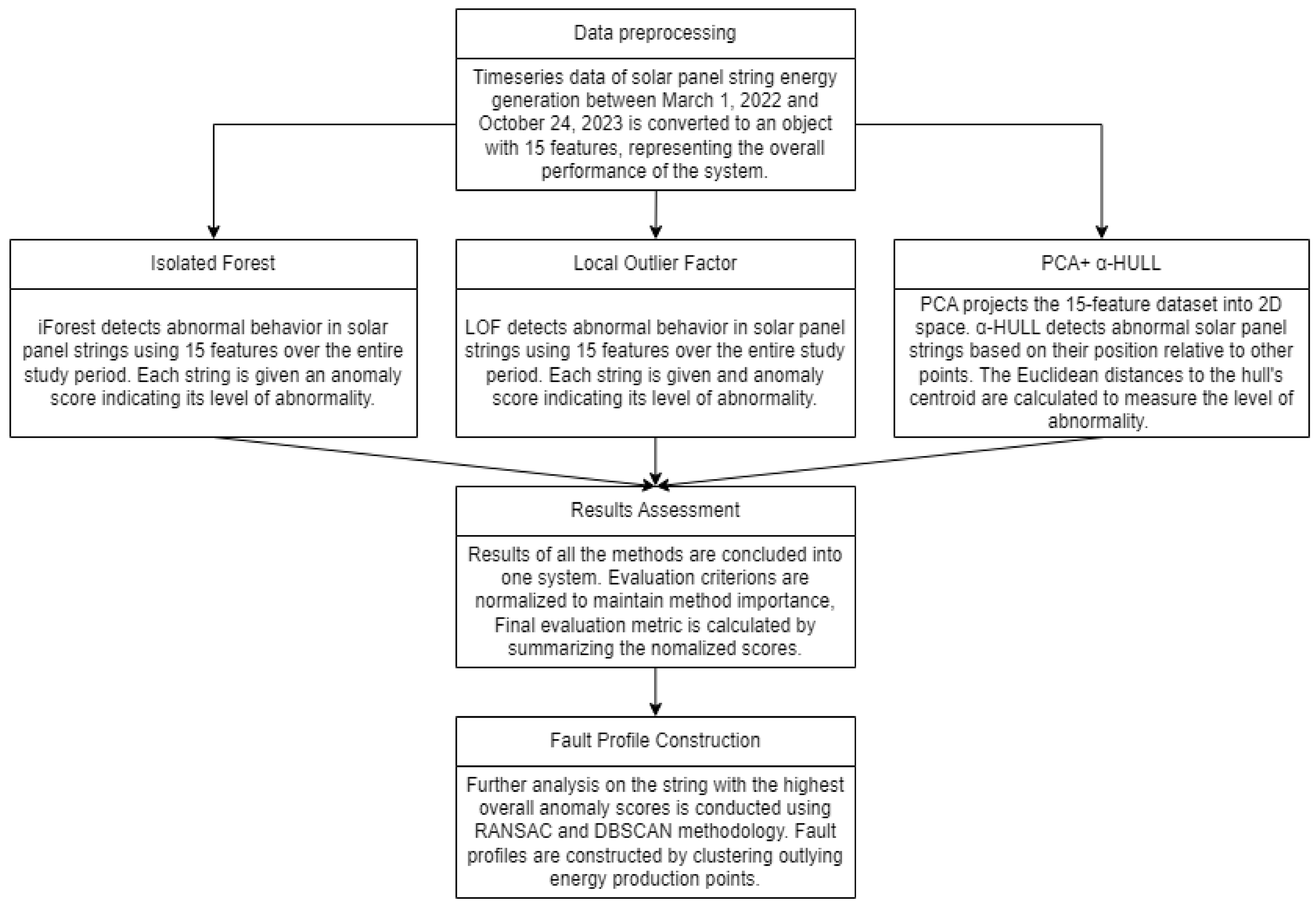

Figure 1: Abstract workflow chart.

Experiments

In case of PCA+α-HULL methodology, various α parameter values were tested. Although α-HULL does not directly yield any anomaly score, a naive scoring system based on the Euclidean distance from the centroid C of the cluster enclosed by the hull was implemented. For iForest and LOF methods the decision was made to check the results of outliers using two different *contamination* parameter values: 0.05 and 0.10. For the results assessment iForest with a contamination parameter set to 0.10, the LOF method with a contamination value of 0.10, and the PCA + -HULL method with an value set to 0.5 were used. Figure 3 presents both the combined normalized anomaly scores and the separate anomaly scores.

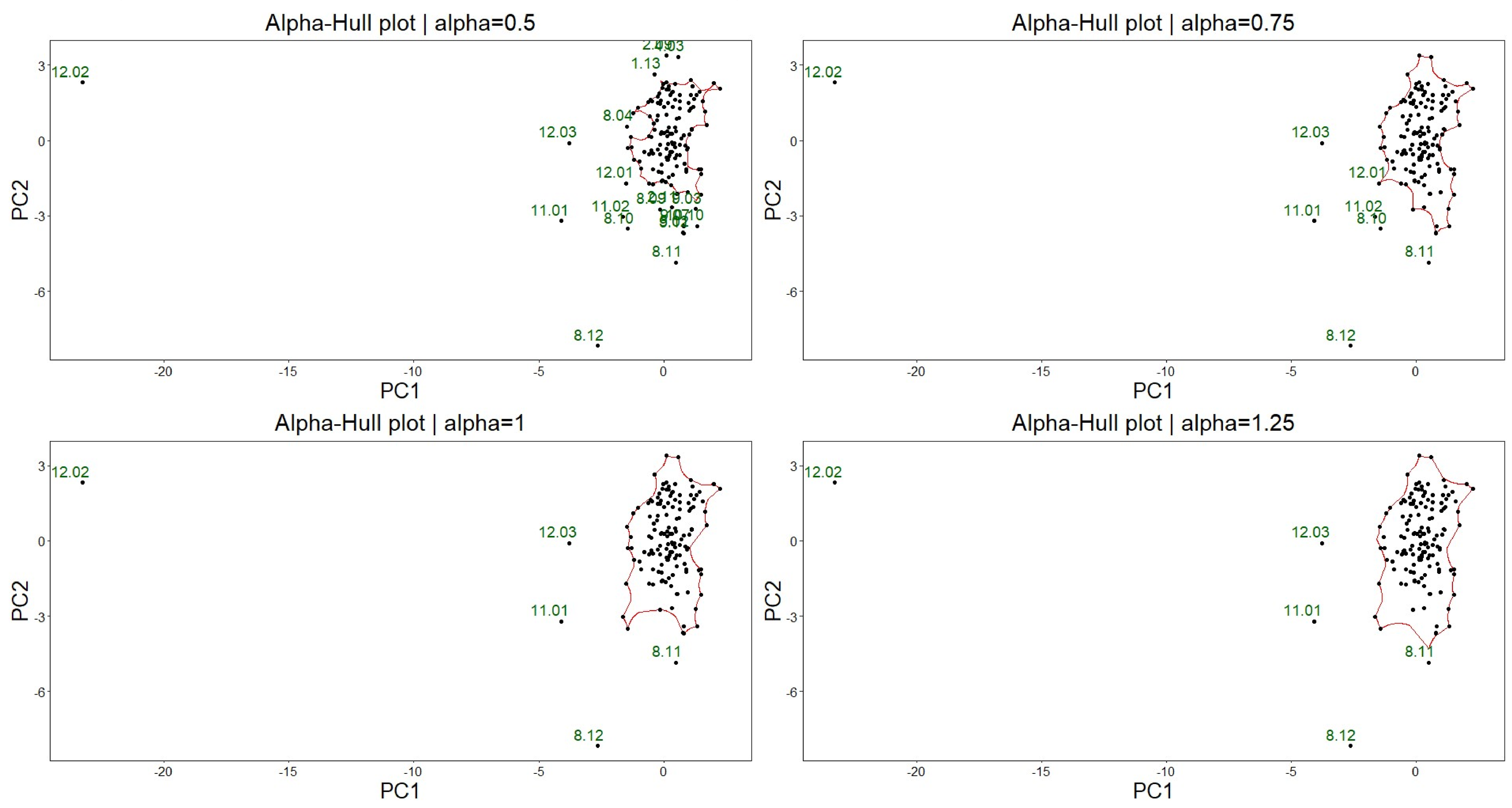

Figure 2: α-Hull results.

α-HULL outlier distances ($\alpha = 0.50$)

String	Distance
12.02	23.63597
8.12	8.871502
11.01	5.533009
8.11	5.072597
8.10	4.097642
12.03	4.052169
9.11	3.939188
5.02	3.908802
...	...

iForest results ($contamination = 0.10$)

String	Anomaly score
12.01	0.267126
8.11	0.157312
12.02	0.083125
10.13	0.077421
8.10	0.04342
12.04	0.033654
8.09	0.024318
8.08	0.019742
...	...

LOF results ($contamination = 0.10, neighborhood\_size = 10$)

String	Anomaly score
12.01	2.955928
12.02	2.086247
8.11	2.049132
10.13	1.740679
1.02	1.52342
12.10	1.453929
8.10	1.431485
11.01	1.393056
...	...

Total sum of normalized outlier scores

Strings	iForest	LOF	α-HULL	Sum
12.01	1	1	0.037889	2.037889
12.02	0.311113	0.472538	1	1.783651
8.11	0.588864	0.450027	0.150722	1.189613
10.13	0.289758	0.26295	-	0.552708
8.10	0.16246	0.075424	0.106117	0.344001
8.12	-	-	0.324522	0.324522
11.01	-	0.052117	0.171786	0.223902
1.02	0.020543	0.131183	-	0.151725
...	...	...	...	...

Figure 3: Anomaly scores.

Previously used methods identified solar panel string outliers based on 15 performance features, yet they lacked explanatory insight into the underlying causes of this detection. As a result, the five systems exhibiting the highest anomaly scores (12.01, 12.02, 8.11, 10.13, 8.10) were investigated by using the RANSAC + DBSCAN methodology.

The first step clusters points into inliers and outliers. Firstly, it is assumed that two closely situated solar systems with identical specifications exhibit a nearly linear relationship under normal conditions. For each solar panel string, an optimal energy generation reference is built using the closest neighboring strings from the same inverter, which are expected to behave similarly. An optimal energy generation benchmark is created by selecting the maximum energy output from these neighboring strings at each timestamp. Linear regression and RANSAC estimate this relationship, marking points that deviate as outliers.

In the second step, energy generation points are visualized on a scatter plot with the hour on the x-axis and the date on the y-axis. Nighttime points are removed for clarity. Patterns emerge as outliers cluster around specific hours, and consistent deviations along the x-axis suggest potential performance issues with the string on those days.

In the third step, the DBSCAN algorithm clusters points, identifying dense regions of anomalous energy generation. Noise points, which do not belong to any cluster, are filtered out, effectively establishing a consistent fault profile.

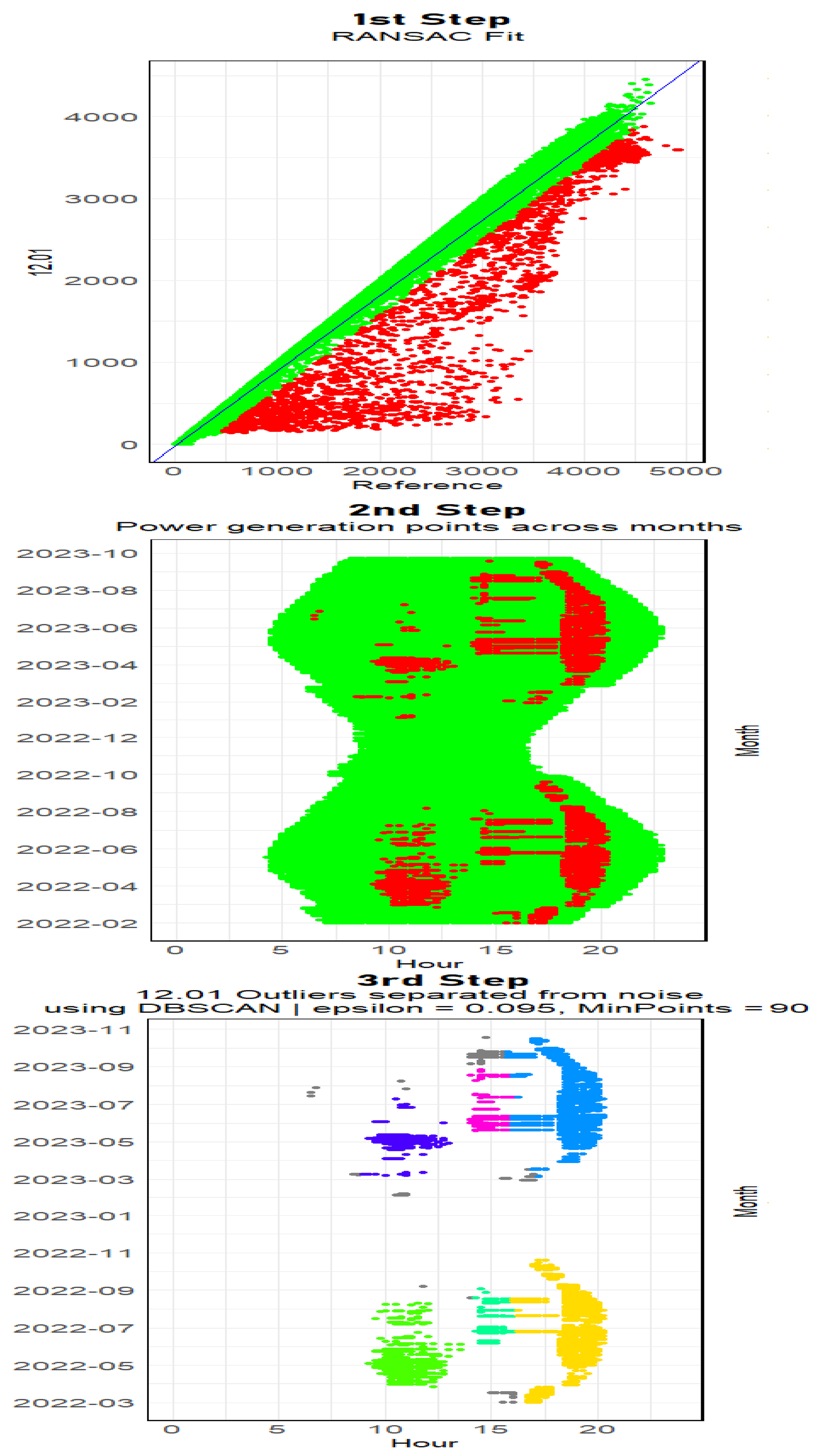

Figure 4: 12.01 string fault profile.

Results

- Identified 34 potentially faulty energy strings.
- Created a system to evaluate a combined anomaly score based on metrics from the applied methodologies.
- Constructed fault profiles revealing recurring patterns of reduced energy generation and further validating previous results.
- Identified consistent occurrence of reduced energy generation during specific hours, indicating to potential issues such as shadowing.

Acknowledgements: The conference participation is funded by EPAM.

References

[1] Beatriz Pateiro-Lopez and Alberto Rodriguez-Casal. alphahull: Generalization of the Convex Hull of a Sample of Points in the Plane.

[2] Liu, Fei Tony and Ting, Kai and Zhou, Zhi-Hua. Isolation Forest.

[3] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. LOF: Identifying density-based local outliers.

[4] Orestis Tsafarakis and Wim G.J.H.M. van Sark. A density-based time-series data analysis methodology for shadow detection in rooftop photovoltaic systems.