Relating Aerial Infrared Thermography Defects to Photovoltaic Performance

Preprint

Kirsten Perry,[1] Quyen Nguyen,[1] Dirk Jordan,[1]
Chris Deline,[1] and Burton Putrah[2]

1 National Renewable Energy Laboratory
2 Zeitview

Presented at the 52nd IEEE Photovoltaic Specialists Conference (PVSC52)
Seattle, Washington
June 9-14, 2024

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5K00-90316
June 2024

Relating Aerial Infrared Thermography Defects to Photovoltaic Performance

Preprint

Kirsten Perry,[1] Quyen Nguyen,[1] Dirk Jordan,[1]
Chris Deline,[1] and Burton Putrah[2]

1 National Renewable Energy Laboratory
2 Zeitview

Suggested Citation

Perry, Kirsten, Quyen Nguyen, Dirk Jordan, Chris Deline, and Burton Putrah. 2024. *Relating Aerial Infrared Thermography Defects to Photovoltaic Performance: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5K00-90316 https://www.nrel.gov/docs/fy24osti/90316.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5K00-90316
June 2024

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by Zeitview under agreement CRD-22-22941, and by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number 38258. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

Relating Aerial Infrared Thermography Defects to Photovoltaic Performance

Kirsten Perry[1], Quyen Nguyen[1], Dirk Jordan[1], Chris Deline[1], Burton Putrah[2]
[1]NREL, Golden, CO, 80401, USA
[2]Zeitview, Santa Monica, CA, 90404, USA

Abstract—**In this paper, we examine the relationship between aerial infrared (IR) defect analysis and photovoltaic (PV) performance data for twelve utility- and commercial-scale solar sites in the United States. To do this, we fuse the site diagram GeoJSON's, aerial infrared thermography (aIRT) defect analyses, and associated inverter time series, allowing for a direct comparison between site defects and time series data. Defect analyses were provided by Zeitview, under its Solar Insights platform. Following the data fusion process, we look at the relationship between system performance and aIRT defects. We investigate the relationship between degradation and hotspot defects, as well as the relationship between AC power data and offline strings and misaligned modules. In general, system degradation was not affected by long-term or balance-of-system (BoS) defects as they occurred infrequently in the data set. However, for one system, a near statistically significant relationship (p-value=0.0571) was found when comparing the degradation of inverter blocks with several multi-hotspot defects to all other inverter blocks without this particular defect. There was strong alignment when comparing short-term recoverable module defects such as stuck trackers and offline strings to time series data. In general, we found that when an inverter block has more than 80% of modules flagged for one of these defects, its AC power time data is flat-lined and the inverter block is not producing.**

Index Terms—**photovoltaic, aerial infrared thermography (aIRT), hotspots, potential-induced degradation (PID), bypass diode, module degradation, system degradation**

I. Introduction

The number of deployed photovoltaic (PV) installations has rapidly proliferated over the past 15 years; in the last decade alone, solar experienced a 24% annual growth rate [1]. As of 2023, there are over 162 gigawatts of solar installed in the United States, with the vast majority of solar capacity attributed to utility-scale sites [1]. This rapid growth in development, coupled with a highly competitive energy market demanding maximum PV plant efficiency, has led to an increased need for low-cost site monitoring to ensure system health [2]. Consequently, the solar industry has increasingly turned to aerial infrared thermography (aIRT) to scan solar sites to detect issues. aIRT has been shown to be a cost-effective and non-destructive monitoring technique to assess the health of utility scale photovoltaic (PV) installations [3]. Furthermore, it only requires visual and thermal cameras which don't contact the solar arrays, so it is easy and safe to deploy at scale [4].

Much research has been dedicated to using aIRT to assess photovoltaic (PV) system faults, particularly using deep learning (DL) approaches. Le et al. [5] used a deep ensemble neural network for the automated detection of solar module anomalies, using a 20,000 aIRT image data set labeled for solar site defects. Using this approach, the authors achieved 94% accuracy at detecting anomalies such as hotspots and offline strings. Similarly, Pierdicca et al. introduced the solAIr system for automated defect detection in aIRT images, also using a deep learning approach [6].

Although there is an abundance of literature focused on the automated detection of anomalies in aIRT imagery, there is limited research focused on analyzing aIRT defects in unison with associated PV performance time series data. This research addresses this by fusing sites' aIRT defects with their associated time series data, so we can relate performance directly to defects. In particular, we investigate the relationships between degradation and hotspots. Hotspots occur when module operating current is greater than a low-current producing solar cell's short-circuit [7], resulting a the cell operating in reverse-bias and dissipating instead of delivering energy [8]. Hotspots cause solar arrays to degrade faster, resulting in decreasing power output of the module [8]. Hotspots can be quickly identified via aIRT in solar arrays due to their temperature differential from the rest of the module.

In addition to analyzing the relationship between degradation and long-term defects such as hotspots, we examine the relationship between soiling trends in time series data and soiling defects, as well as short-term performance trends for defects such as offline strings and stuck trackers around the aerial site scan dates.

II. Methods

A. Data Sets

For this research, sites available from the NREL PV Fleet Data Initiative were analyzed. PV Fleets is a US Department of Energy-funded project focused on the collection of fielded PV performance data to perform large-scale degradation analysis on the US fleet [9]. The associated PV Fleets database contains time series data for over 6500 sites, ranging from residential to utility-scale systems. The results presented in this research include data from 12 sites, located in California and Colorado. Both commercial and utility-scale sites were analysed. A map of the sites is provided in Figure 1.

The aIRT defect analysis used in this research was provided through Zeitview's Solar Insights platform, in GeoJSON format. Using drones and airplanes, Zeitview has scanned over 6100 solar sites across the US, ranging from commercial to

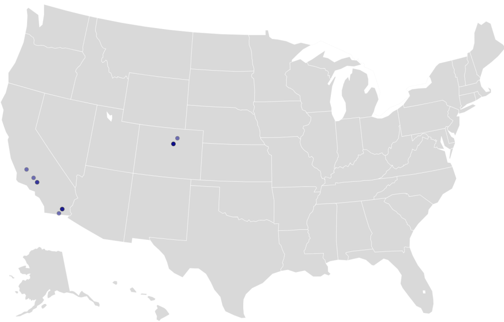

Fig. 1. Map of sites analyzed in this research. Several systems were co-located, and are represented via darker dots.

utility-scale. Following each scan, the associated site is analyzed for anomalies using machine learning (ML) techniques. Flagged defects include the following:

- Hotspots
- Stuck trackers/misaligned modules
- Potential-induced degradation (PID)
- Soiling
- Offline strings
- Bypass diode issues
- Physical damage or anomalies, such as broken glass, damaged modules, and missing modules

For utility-scale sites with multiple inverters, it is important to accurately map the inverter data streams to their associated geographic location, so they can be correctly compared to aIRT results. To do this, PV Fleets data partners provided site diagrams for the sites used in this analysis.

B. Automatically Converting Site Diagrams to GeoJSON Format

Converting site diagram information for solar installations into machine-readable format is challenging. In particular, mapping equipment to its associated geographic location can be difficult and time-consuming, especially when performed manually. Additionally, as-built drawing format can vary significantly across sites. In spite of this, for the site diagrams provided, we were able to identify particular sections within as-built drawings that mapped inverter blocks to their associated locations. Although this process will vary depending on site diagram format, we provide a general framework for automation for this particular scenario. We plan to eventually expand upon this work by identifying and automating for additional site diagram configurations, and releasing these processes publicly via Github.

Figure 2 illustrates the automated process for converting an as-built PDF into GeoJSON format with mapped system blocks. If block tables like the one shown in Figure 2 are available, they are processed into GeoJSON format via the following set of steps:

- Step 1: Extracting the PDF image containing tables for the northing-easting coordinates of each inverter block to a Python Pandas dataframe object, via the OpenCV and EasyOCR packages [10], [11]. Individual tables and their associated cells are found via an OpenCV contouring algorithm [12]. The algorithm pre-processes the image by converting it to grayscale, applying thresholding, and inverting and dilating to highlight the table structure and remove noise. The individual table cells are then cropped and assessed via the EasyOCR package. The text for each cell is extracted and fed into a Pandas dataframe, in accordance with the determined table structure.
- Step 2: Manually verifying the EasyOCR outputs to ensure that all coordinates were successfully transferred to the dataframe object and that any output errors are corrected.
- Step 3: Converting northing-easting coordinates to WGS84 format via the Python PyPROJ package [13].
- Step 4: Generating a polygon based on the sets of coordinates extracted for each inverter block. To do this, the alphashape package was used [14]; in particular, alpha shapes draw a bounding box around a set of points. Each polygon generated was manually reviewed to ensure accuracy.
- Step 5: Feeding generated inverter block lists and their associated polygon information into GeoJSON format. This GeoJSON will be fused with aIRT defect GeoJSON data during the main analysis.

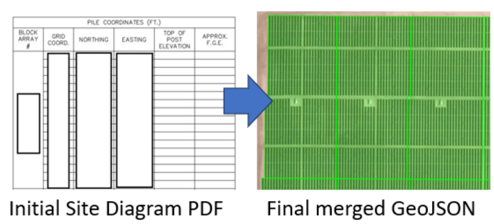

Fig. 2. Diagram illustrating the initial site diagram PDF and final mapped site GeoJSON.

C. Data Fusion

The data fusion process that allows for accurate comparison of solar time series data against aIRT defects is shown in Figure 3. The data fusion process is as follows:

- Step 1: The GeoJSON representation of the site diagram is merged with the associated aerial IR GeoJSON, yielding defects by inverter block. To do this, the Python Shapely package is used to generate polygon objects which allow for direct comparison [15].
- Step 2: The time series data streams associated with the site are mapped to their associated inverter block, using sensor naming conventions provided by site owners.
- Step 3: Inverter-level degradation rates are calculated using the time series data. To calculate inverter-level degradation, the RdTools methodology outlined in [9] was used, using AC power inverter-level data and ambient temperature and irradiance data from the site meteorological station. The Python PVAnalytics package

[16] was used to pre-process each of the time series data streams. Furthermore, each degradation analysis was manually reviewed for data issues and egregiously inaccurate analyses were removed. In total, 54 out of a total 446 analyses were manually flagged for data quality issues and removed. For cases where multiple inverters were associated with a single system block, the median degradation value across the inverters was taken as the block's degradation value.

- Step 4: The inverter block data streams and their associated degradation rates were combined with defect information by inverter block, allowing for a direct comparison between aIRT and time series data/degradation.

A simple example illustrating the site defect fusion is available via Github [17]. This example uses a commercial installation on the NREL campus, so inverter block comparisons are not available.

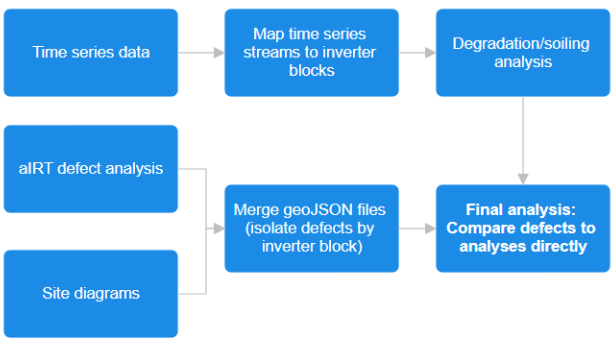

Fig. 3. Diagram illustrating the data fusion process.

In addition to fusing the data sources, the provided code generates several figures, including an interactive satellite image displaying both the site and defect GeoJSON data (generated via the Python Folium package [18]), and short-term performance plots taken around the site inspection date (generated via the Python Plotly package [19]). Examples of these graphics are shown in Figures 4 and 5, respectively.

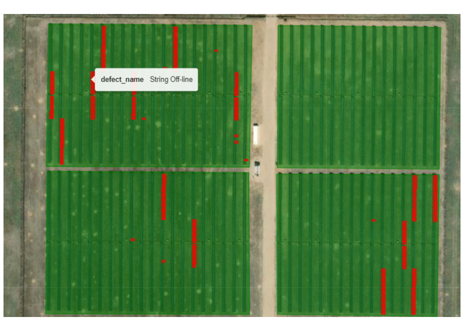

Fig. 4. GeoJSON data for the site (in green) and associated aerial IR defects (in red) displayed in an auto-generated interactive Folium graphic.

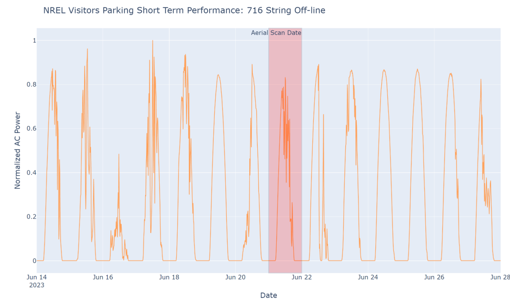

Fig. 5. Short-term system performance around aerial IR scan period (in red) in an auto-generated interactive Plotly graphic.

III. Results

A. Defect Analysis

Figure 6 shows the module defect counts from aIRT, summed across all systems. Short-term performance issues, which we will define as issues where system performance is recoverable with routine site maintenance, are shown in the first subplot and dominate the defect types found. In particular, misaligned modules, which are associated with stuck trackers, and offline strings were the most commonly observed issue across all 12 systems. There were far fewer instances of defects associated with long-term performance issues, which we define as issues causing long-term underperformance of the system. Issues associated with long-term defects or Balance-of-System (BoS) problems are shown in the second subplot of Figure 6. Such defects include hotspots (355 modules across all categories), and potential-induced degradation (PID) (5 modules).

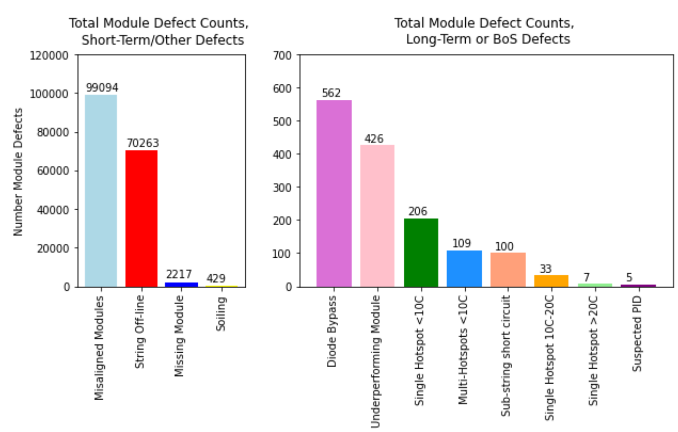

Fig. 6. Number of module defects across all systems, as determined from aIRT.

Total module percentage defects were examined in an effort to better understand the frequency of their occurrence. Total module counts were tabulated via site diagram information. To calculate frequency of occurrence, the following equation was used:

$$module\ defect\ percentage = \frac{module\ defect\ count}{total\ module\ count} \quad (1)$$

TABLE I
PERCENTAGE MODULES BY DEFECT TYPE, ACROSS ALL SYSTEMS

Defect Type	Percentage (%)
Multiple Hotspots (<10 deg C)	0.00625
Single Hotspot (<10 deg C)	0.01182
Single Hotspots 10C-20C	0.00189
Single Hotspot >20C	0.00040
Suspected PID	0.00029
Bypass Diode Issues	0.0322
String Offline	4.03179
Underperforming/Isolated Module	0.02444
Misaligned Module	5.68616
Missing Module	0.12722
Soiling	0.02462
Sub-string Short Circuit	0.00574

Table I shows defect frequency results. Long-term or BoS issues are incredibly infrequent in this data set, with bypass diode issues most commonly occurring at 0.0322% of the time. Aggregated hotspot defects have a frequency of approximately 0.02%. Short-term or recoverable defects are much more frequently occurring, with misaligned module/stuck tracker defects occurring with a frequency of 5.69% and offline string defects occurring with a frequency of 4.032%. Overall, these defect occurrence percentages are promising for solar operators, as it is much easier and cheaper to fix mechanical issues causing short-term performance losses than it is to fix long-term performance issues.

B. Long-Term Performance

1) Degradation: To ensure anonymity and allow for direct comparison across systems, all degradation values presented in this study were min-max normalized on a system level, with 0 as the lowest degradation and -1 as the highest degradation. To reduce noise and focus primarily on long-term or BoS defect relationships, inverter blocks with more than 100 misaligned modules or offline strings were removed from the analysis. Although both of these are short-term defect types, if not resolved, they could be misconstrued as faster degradation and increase noise when evaluating for other types of defects.

The degradation rates of inverter blocks with certain types of defects were compared against inverter blocks where the defect was not present. Table II outlines the degradation rates of systems with and without these defects, as well their statistical T-test p-values. Statistically significant p-values (<0.05) indicate that inverter blocks containing a particular type of defect degrade faster than inverter blocks without this defect present.

The presence of multiple hotspots (<10 degrees Celsius) did appear to lead to higher degradation rates in inverter blocks, with a statistically significant p-value of 0.0093. Because some of the blocks contained only one or two multiple hotspot defects out of hundreds or thousands of modules, the statistically significant relationship is most likely a coincidence.

However, in an effort to better understand if the relationship between multiple hotspots and degradation is real, distributions were examined at a system level. System 4, which had the most modules flagged for multiple hotspot issues in the entire data set, experienced higher degradation rates in inverter blocks with multiple hotspots detected vs. inverter blocks with no hotspots (p-value: 0.0571). The distributions for this system are shown in Figure 7. In particular, one block in system 4 had 28 modules out of a total 7680 modules detected with multiple hotspots present (approx. 0.4%), as well as an additional 3 modules with single hotspots defects. An additional 40 modules were flagged as offline in this block; this is in line with the rest of the inverter blocks being compared, where the median modules marked as offline was 40 and the mean was 53. Because the number of offline string modules is generally constant when comparing across blocks, the main difference between this particular block and all other blocks is the presence of multiple hotspot defects. This could indicate that multi-hotspot defects could increase degradation rates; although, there may be other unknown factors at play here that may be contributing to faster degradation for this particular inverter block. Figure 8 shows the location of these hotspot defects in context of the inverter block. All multi-hotspot defects are located along a single string.

TABLE II
COMPARING SYSTEM BLOCK AC POWER OUTPUTS TO DEFECTS

Defect Type	Normalized Median Defect Rd	Normalized Median Non-Defect Rd	p-value
Multiple Hotspots (<10 deg C)	-0.75	-0.50	0.0093
Hotspots (all)	-0.51	-0.52	0.519
Suspected PID	-0.65	-0.46	0.4
Bypass diode Issues	-0.51	-0.58	0.544

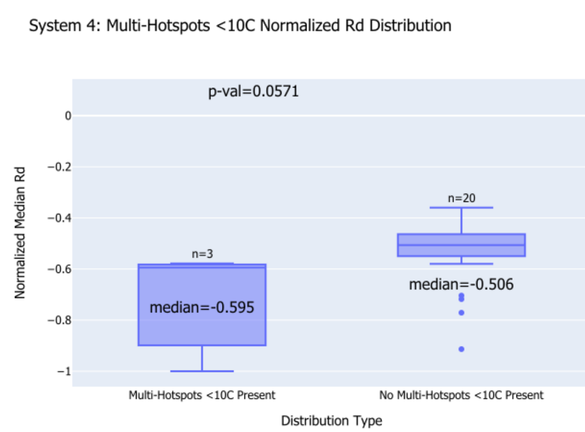

Fig. 7. Box plot showing the normalized degradation distribution for inverter blocks with multiple hotspots, vs. inverter blocks without. These results are for system 4 only.

Many systems experienced similar degradation rates when comparing defect vs. non-defect distributions. Generally, long-term defects only affected a few modules across an inverter block, which may contain hundreds or thousands of modules.

TABLE III
SYSTEM 4 MULTI-HOTSPOT BLOCK DETAILS

Total Number Modules	7680
Number Multi-Hotspot Modules	28
Number Single-Hotspot Modules	3
Number Offline String Modules	40
Avg Number Offline String Modules (all blocks)	53
Median Number Offline String Modules (all blocks)	40

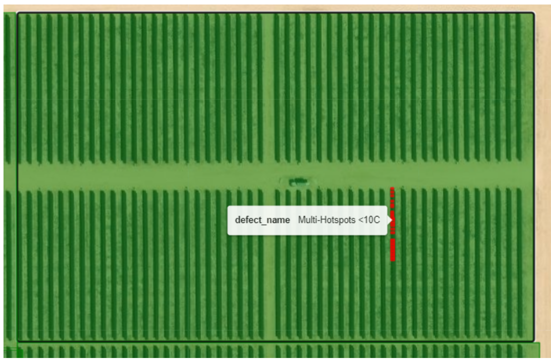

Fig. 8. Screenshot of multi-hotspot defects for the heavily affected block in system 4.

Given this scale, the impacts of these defects will likely not impact overall degradation rates for the entire inverter block. If degradation was calculated at a string level, these relationships would be more prominent, and stronger relationships could be formed.

Furthermore, systems being compared in this analysis are in different geographic locations, with different module types, and have their own unique set of issues. Although we attempt to minimize differences by normalizing degradation rates on a system level so systems can be compared side-by-side, these additional factors may be causing noise in the distribution comparisons.

On a positive note, many of the systems analyzed in this research are relatively robust to certain defect issues at an inverter block level, as overall degradation rates did not increase by a statistically significant margin in the presence of these types of issues. This bodes well for overall system reliability.

2) Soiling: In this section, we examine the relationship between soiling signals present in time series data against detected aIRT soiling defects. To calculate the inverter-level soiling ratio in the time series data, the stochastic rate & recovery algorithm (SRR) in the RdTools package was used [20]. Soiling outputs were manually reviewed after SRR calculation to ensure accuracy; this included checking for a "saw-tooth" pattern representative of a soiling-cleaning event, which is prevalent in heavily soiled systems. Systems were manually reviewed and labeled as no/minimal soiling or soiling based on this pattern and associated SRR outputs. An example output of SRR is shown in Figure 9. Due to the prevalence of saw-tooth behavior, this system was classified with soiling issues.

In total, 5 systems in the data set were labeled for the

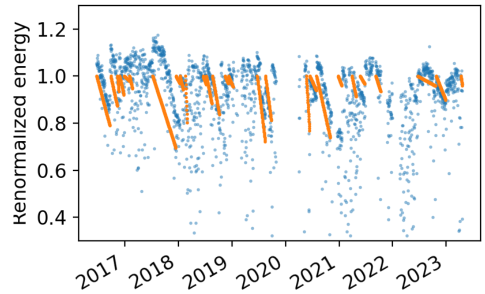

Fig. 9. Inverter-level renormalized energy time series for a system with heavy soiling.

presence of soiling. Total system module soiling defect counts for no/minimal soiling systems were compared to defect counts for systems with soiling present. A box plot showing these distributions is provided in Figure 10. Additionally, soiling/non-soiling distributions of percentage soiling defects is displayed in Figure 11. It is important to include this distribution, as systems can vary significantly in size.

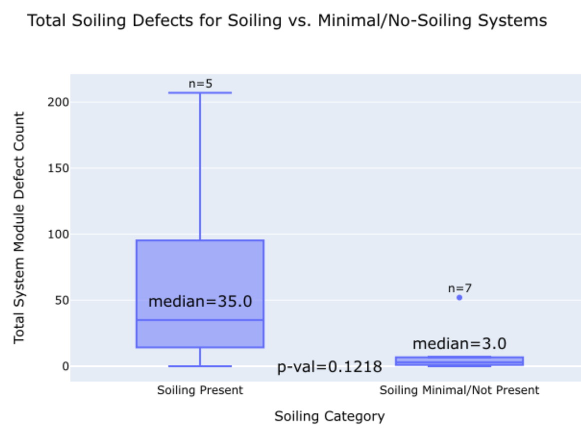

Fig. 10. Number of detected soiling defects for systems that display soiling time series behaviors vs. systems that do not.

When juxtaposing Figures 10 and 11, aIRT analysis correctly identifies systems with soiling issues, but it doesn't do so proportionally when looking at percentage of modules affected. For the systems that displayed soiling behaviors via time series analysis, soiling signals were prevalent across all inverter AC power time series data analyzed. However, only a few modules were flagged for soiling defects via the aIRT analysis.

C. Short-Term Performance

1) Time series: Time series data from a one-week period before and after each site scan was captured, with the intent of analyzing short-term site performance around the aIRT scan date. In particular, we examined cases with string outages

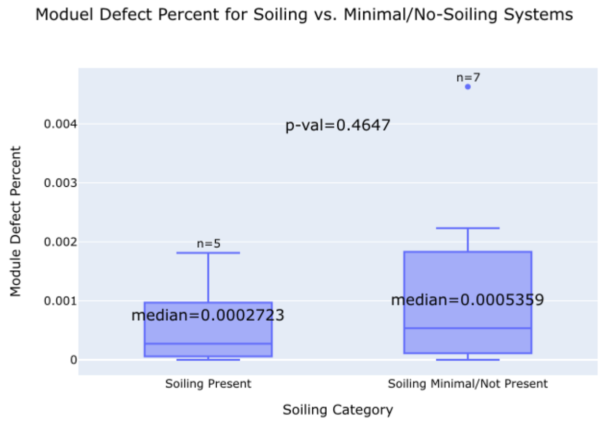

Fig. 11. Module percentage of soiling defects for systems that display soiling time series behaviors vs. systems that do not.

and misaligned modules, which were the two more frequently occurring defects in the entire data set. We isolated a subset of inverter block cases with over 500 module defects for the associated defect category, and looked at the associated AC power time series output. In total, we examined the AC power time series data associated with 35 unique inverter blocks (1 block had both misaligned modules and offline strings issues). Each block was classified into one of three categories based on its time series behaviors:

- Producing (AC power data signal looks normal)
- Curtailment or other unknown issues (flatline in the middle of the day indicating curtailment behavior or other strange behavior)
- Offline (AC power data is flatlined at 0)

Figure 12 shows an example normalized AC power output for a system block flagged for 640 offline modules. The block appears to have an outage until the scan date, and then is heavily curtailed. The inverter set-points were checked with the system operator to determine system curtailment.

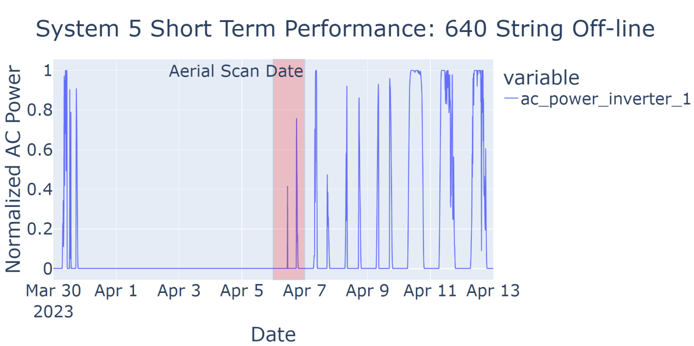

Fig. 12. Normalized AC power production for a system block flagged for string offline issues, with the aIRT scan date highlighted in red.

The percentage of modules affected on an inverter block level was calculated and compared across classifications and defect types, as shown in Figures 13 and 14. These boxplots show good alignment between the time series and aIRT defect data. When over 80% of the modules are affected by offline string defects, the entire inverter block is generally offline. When only a small fraction of modules are flagged with offline string defects (<10%), the AC power data associated with the inverter block is still producing as expected. This also applies to curtailment; the inverter block is still producing, but production is throttled.

These trends also hold for misaligned module/stuck tracker defects, as shown in Figure 14. These results indicate that inverter blocks were completely offline when over 80% of their trackers were misaligned. These results were confirmed directly with Zeitview; in their process, modules were originally flagged as offline in the aIRT analysis, and their designation was set as "misaligned module" if the associated tracker was misaligned. Consequently, these flagged modules are both offline and associated with stuck trackers.

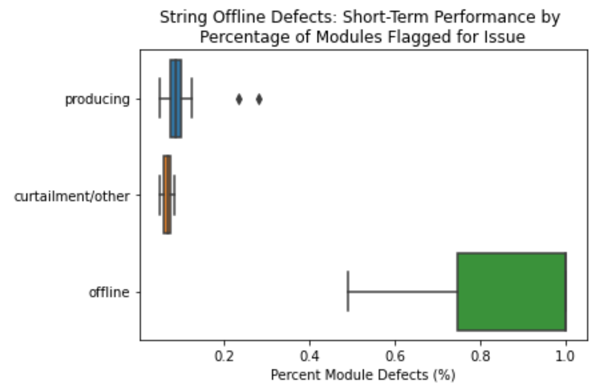

Fig. 13. Box plot of percentage module defects by inverter block by outage category, for inverter blocks flagged for offline strings.

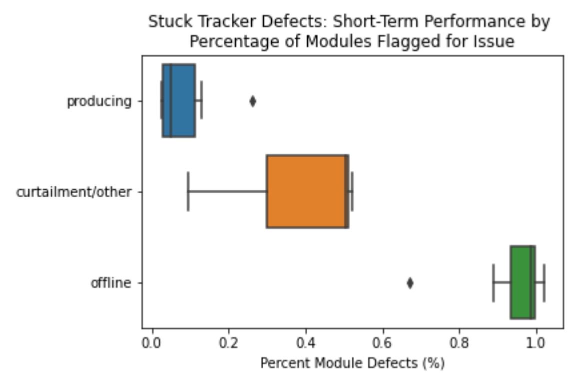

Fig. 14. Box plot of percentage module defects by inverter block by outage category, for inverter blocks flagged for misaligned modules/stuck trackers.

IV. Conclusions & Next Steps

In this research, a process for automating joining aIRT defect data and solar time series data is introduced. This includes both the automation of specific site diagram formats into GeoJSON format, and the fusing of data set GeoJSON's via Python. Following this process, defect data is compared to time series data from 12 systems in the United States. Most defects are infrequently occurring, with the exception of offline strings and misaligned modules/stuck trackers.

Due to the infrequency of long-term and BoS defects, systems did not experience faster degradation rates when presence of these defects was detected. However, one particular system, system 4, had frequently occurring multiple hotspot defects. When comparing the inverter block with these defects to all other blocks, a near statistically significant relationship indicating faster degradation in the presence of hotspots was determined. This indicates a possible relationship between hotspots and degradation in field installations.

The relationship between short-term recoverable defects, including offline strings and stuck trackers/misaligned modules, and AC power time series data was investigated. Percentage module defects aligned closely with time series performance. In general, when over 80% of modules in an inverter block are labeled with one of these defects, the associated time series data show that the block is completely offline.

Because this is a largely unexplored topic in the literature, there is still much work to be done in this space. In particular, the authors of this work plan to expand the analysis to more systems in the NREL PV Fleets Initiative. Additionally, we are working with Zeitview to obtain aIRT scans for sites over time, allowing for time-dependent defect analysis. By adding this time component, we can examine how defects change over time for a particular system and further relate this to time series data.

The standardization and automation of site diagrams into GeoJSON format would significantly aid in performing analyses such as this one. We plan to further investigate this field of research, and develop a Python toolkit to aid in automating this process. Although this is a problem that is not generalizable, identifying common formats and automating the GeoJSON conversion could save solar site operators considerable time when fusing aIRT and PV site performance data.

Acknowledgment

This work was authored by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by Zeitview under agreement CRD-22-22941, and by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number 38258.

References

[1] Solar industry research data. https://www.seia.org/solar-industry-research-data. Accessed: 2024-01-17.

[2] John A. Tsanakas, Long D. Ha, and F. Al Shakarchi. Advanced inspection of photovoltaic installations by aerial triangulation and terrestrial georeferencing of thermal/visual imagery. *Renewable Energy*, 2017.

[3] Aline Kirsten Vidal de Oliveira, Mohammadreza Aghaei, and Ricardo Rüther. Aerial infrared thermography for low-cost and fast fault detection in utility-scale pv power plants. *Solar Energy*, 2020.

[4] A.W. Kandeal, M.R. Elkadeem, Amrit Kumar Thakur, Gamal B. Abdelaziz, Ravishankar Sathyamurthy, A.E. Kabeel, Nuo Yang, and Swellam W. Sharshir. Infrared thermography-based condition monitoring of solar photovoltaic systems: A mini review of recent advances. *Solar Energy*, 2021.

[5] Minhhuy Le, Van Su Luong, Dang Khoa Nguyen, Van-Duong Dao, Ngoc Hung Vu, and Hong Ha Thi Vu. Remote anomaly detection and classification of solar photovoltaic modules based on deep neural network. *Sustainable Energy Technologies and Assessments*, 2021.

[6] Roberto Pierdicca, Marina Paolanti, Andrea Felicetti, Fabio Piccinini, and Primo Zingaretti. Automatic faults detection of photovoltaic farms: solair, a deep learning-based system for thermal images. *Energies*, 2020.

[7] Michael Simon and Edson L. Meyer. Detection and analysis of hot-spot formation in solar cells. *Solar Energy Materials and Solar Cells*, 94(2):106–113, 2010. URL: https://www.sciencedirect.com/science/article/pii/S0927024809003420, doi:10.1016/j.solmat.2009.09.016.

[8] Mahmoud Dhimish, Violeta Holmes, Peter Mather, and Martin Sibley. Novel hot spot mitigation technique to enhance photovoltaic solar panels output power performance. *Solar Energy Materials and Solar Cells*, 179:72–79, 2018. URL: https://www.sciencedirect.com/science/article/pii/S0927024818300709, doi:10.1016/j.solmat.2018.02.019.

[9] Dirk C. Jordan, Kevin Anderson, Kirsten Perry, Matthew Muller, Michael Deceglie, Robert White, and Chris Deline. Photovoltaic fleet degradation insights. *Progress in Photovoltaics: Research and Applications*, 2022.

[10] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[11] Easyocr. https://github.com/JaidedAI/EasyOCR.

[12] Khoj Badami. How to: Extract table from image in python (opencv ocr). https://livefiredev.com/how-to-extract-table-from-image-in-python-opencv-ocr/, 2023.

[13] Pyproj. https://github.com/pyproj4/pyproj.

[14] Alphashape. https://github.com/bellockk/alphashape.

[15] Shapely. https://github.com/shapely/shapely.

[16] Kirsten Perry, William Vining, Kevin Anderson, Matthew Muller, and Cliff Hansen. Pvanalytics: A python package for automated processing of solar time series data. 2022.

[17] Kirsten Perry and Quyen Nguyen. Aerial ir-pv time series fusion, 6 2024. URL: https://github.com/kperrynrel/aerial_IR_PV_time_series_analysis.

[18] python visualization. Folium. URL: https://python-visualization.github.io/folium/.

[19] Plotly Technologies Inc. Collaborative data science, 2015. URL: https://plot.ly.

[20] Michael G. Deceglie, Leonardo Micheli, and Matthew Muller. Quantifying soiling loss directly from pv yield. *IEEE Journal of Photovoltaics*, 8(2):547–551, 2018. doi:10.1109/JPHOTOV.2017.2784682.