PV Fleet Performance Data Initiative: March 2020 Methodology Report

Chris Deline, Matt Muller, Michael Deceglie, Dirk Jordan, Kevin Anderson, Lin Simpson, Kirsten Perry, and Robert White

National Renewable Energy Laboratory

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

Technical Report
NREL/TP-5K00-76687
June 2020

PV Fleet Performance Data Initiative: March 2020 Methodology Report

Chris Deline, Matt Muller, Michael Deceglie, Dirk Jordan, Kevin Anderson, Lin Simpson, Kirsten Perry, and Robert White

National Renewable Energy Laboratory

Suggested Citation
Deline, Chris, Matt Muller, Michael Deceglie, Dirk Jordan, Kevin Anderson, Lin Simpson, Kirsten Perry, and Robert White. 2020. *PV Fleet Performance Data Initiative: March 2020 Methodology Report*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5K00-76687. https://www.nrel.gov/docs/fy20osti/76687.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

Technical Report
NREL/TP-5K00-76687
June 2020

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

 Funding provided by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number 34348.

Acknowledgements

We would like to thank all the corporate partners who provided the PV system data used to perform the analyses in this document.

Executive Summary

Data for thousands of systems by multiple sources were transmitted to NREL for evaluation. Initial automated data quality assurance (QA) checks identified hundreds of systems (QA Tier 1) with high-quality meteorological and AC production data and hundreds of QA Tier 2 systems with adequate data quality. A standard RdTools analysis was conducted for these systems, evaluating performance loss on an annualized basis. To date with the systems evaluated so far, we find the median performance loss rate (aka Rd or degradation rates)is in line with historical degradation rates previously published for modules and systems (−0.5% to −0.9% / yr, Jordan et al. 2016). Updated median values and quantile statistics will be appended to this report annually as additional partners are recruited, and as data techniques are refined.

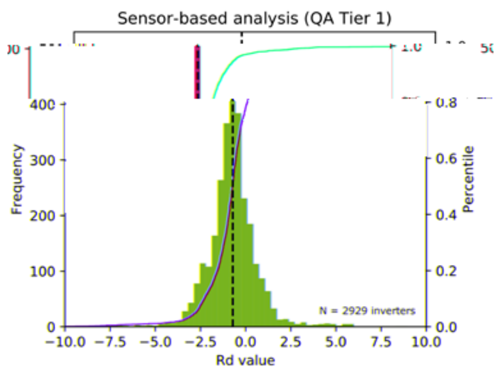


Figure ES-1. Fleet-level Rd distributions using AC inverter data. QA-compliant Tier 1 systems were analyzed using site-measured irradiance and temperature

While abnormally low or high values do appear in the above distribution, a detailed study in Section 4.3 shows that many of these entries arise from data quality issues or system configuration changes not corrected by our automatic analysis. Some negative Rd values may indeed suggest corrective action is needed on individual systems, which may be identified by individual underperforming inverters.

Table of Contents

1 PV Fleet Performance Data Initiative Introduction

The information contained in this report is part of an ongoing analysis effort by the National Renewable Energy Laboratory (NREL) and the U.S. Department of Energy (DOE). This Photovoltaics (PV) Fleet Performance Data Initiative supports the U.S. PV community by pooling and analyzing plant operation data and providing PV performance assessments of individual solar assets using standardized state-of-the-art methods. The analysis results provide plant owners and operators with a confidential detailed assessment of their fleet performance, while also providing the broader community an aggregate benchmark for the performance of the U.S. solar fleet. The outcomes will enable more efficient operation of PV installations and improve financial assessment accuracy for current and future PV power plants. All the data provided are protected by confidentiality agreements, are maintained by NREL in a secure database (i.e., the DuraMAT DataHub), and will not be shared with any person or group not authorized by the data owner.

The initiative uses power output data, mainly from medium and large (>250 kilowatts) PV installations over at least two years to provide performance assessments using RdTools. RdTools is a set of open-source Python-based PV data analysis tools used to calculate plant-level degradation rates, as well as performance impacts by soiling, inverter clipping, and plant availability. The analysis provides system-level evaluation of individual PV plants, including degradation rates and events affecting the plant performance, and compares the performance of a specific power plant to an aggregate benchmark. The analysis provided in this report is for the combined Energy fleet.

2 Data Partner Fleet Summary

2.1 Fleet Metadata Details

Performance data from the data partners consists of over a thousand systems spread across the country (Figure 1). A histogram of system sizes and types of PV indicates a variety of sizes (Figure 2, Figure 3), with median system size of 1 MW DC and 4.5 years of data on average.

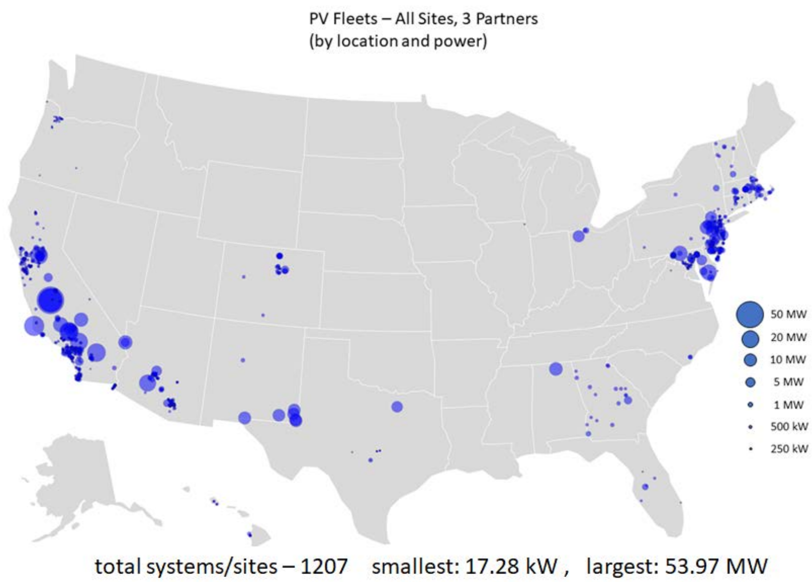


Figure 1: Spatial map of the data partner systems

As new data partners are recruited, new findings and results will be posted in semi-annual bulletin updates through the nrel.gov publications page.

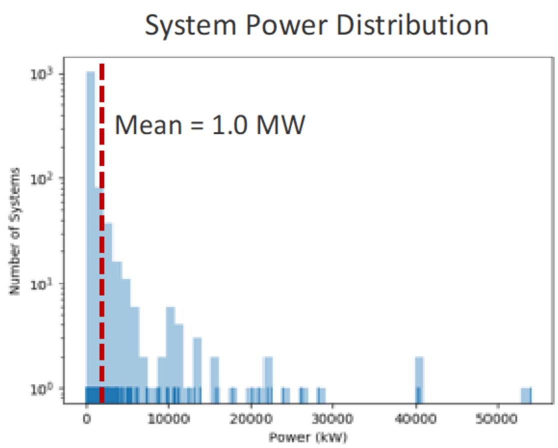


Figure 2: Distribution of power sizes fleet

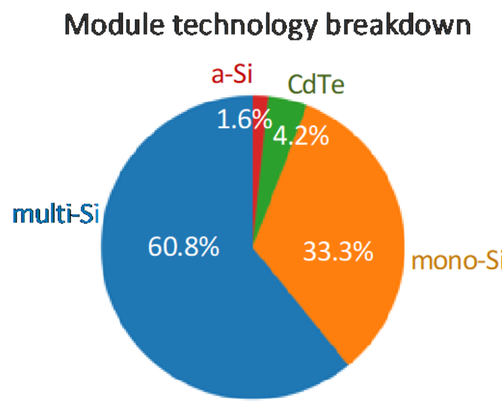


Figure 3: Module technology breakdown

Time-series data were transferred from the repository where the original system data resides. Site metadata details were also collected from there. A CSV file of site information provided by the owner/operator was used to fill gaps or missing data. This metadata included system geographic location, tilt and azimuth orientation, module, inverter components, meters, weather stations, and reference cells. We merged the available geographic location with climate data including the PV climate stressor information to evaluate possible vectors for degradation in the fleet [Karin 2019].

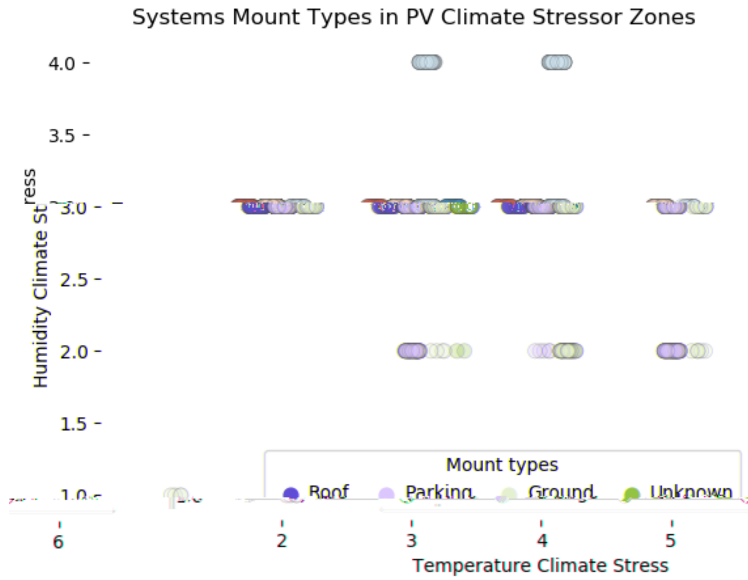


Figure 4: PV Climate stressors and PV mounting configurations. The higher climate zone numbers for temperature and humidity, the greater possible stress on the system modules. Temperature climate numbers are for ground-mount. Roof-mounted installations will tend to have even higher stress than ground due to the more restricted air flow [Karin, 2019].

For analyzing each system, we prefer to have measured data streams for AC power production data, module temperature, and plane-of-array (POA) irradiance. Other combinations of measurements could allow us to compute these values, but with perhaps increased analysis uncertainty.

2.2 Data Quality Details

EachEach system is processed through autonomous prototype quality assurance (QA) checks described in detail in Section 3.2. Initial QA results suggest that additional systems can be re-analyzed after manual corrections are applied. For example, certain systems were found to have multiple azimuth orientations while the metadata included just one orientation. In the current version of QA, all inverters or irradiance sensors that did not match the specified orientation were failed. In the future, these systems can be reclassified and contribute to the Fleet-wide statistics. In total, thousands of AC power sensors, irradiance sensors, and temperature sensors were processed through the QA software. Table 1. through Table 3 provide a breakdown of the QA reject information for each sensor type.

Table 1. AC Power - Initial Data Quality Check

AC power measurements 13% overall reject rate	
QA failure reason	**% of total sensor rejected**
Timestamp issue (timezone, daylight savings, or gross azimuth error)	7%
Smaller azimuth reporting error	2%
Poor summer or winter data fits (shading likely)	1-2%
Insufficient data for fits	1-2%
Inconclusive “tracking or fixed” check	1%
Ambiguous “tracking or fixed” due to clipping	< 1%

Table 2. Irradiance – Initial Data Quality Check

Irradiance measurements 36% overall reject rate	
QA failure reason	**% of total sensor rejected**
Number of days with unreasonably low irradiance	10%
Inconclusive “tracking or fixed” check	9%
Less than 25% valid data	8%
Timestamp issue (timezone, daylight savings, or gross azimuth error)	4%
Likely mislabeled orientation	1-2%
Poor summer or winter data fits (shading or noisy data likely)	1-2%
Insufficient data for fits	<1%

Table 3. Temperature – Initial Data Quality Check

Temperature measurements 30% overall failure rate	
QA failure reason	**% of total sensor rejected**
Less than 85% valid data	20%
Poor correlation between module temperature and valid irradiance	10%

Here data quality reject does not necessarily indicate that system data could not eventually be evaluated. Our initial focus on high-quality system data in an automated fashion identified potential issues with specific systems that should manually be evaluated.

3 Methods

3.1 Overview of PV Fleet Analysis Methods

PV Fleet performance analysis follows the workflow shown in Figure 5, starting with quality assessment of system data then proceeding to time-series degradation and loss-factor analysis. The analysis is largely automatic and unsupervised, requiring initial data screening to reject systems and channels with errors in measurement or configuration. The intent is to remove erroneous, unphysical system data prior to analysis to avoid skewing the resulting fleet performance degradation distribution. Using QA-validated system data, we then use the open-source RdTools software toolkit [RdTools 2019] to generate annualized degradation rates (annual system performance loss rates) at the meter and inverter level.

Figure 5: Analysis procedure for PV fleet data showing data QA, Rd, and Loss Factor analysis, along with partner reporting

Additional analysis, such as loss factor evaluation (soiling, data availability, and modeled vs measured Performance Index), will be included at a later date.

3.2 Initial Data Quality Checks

Each irradiance, temperature, and power measurement within a given PV system is validated by a series of data quality assurance (QA) checks. The first QA check is to determine erroneous, extreme, or stuck data points within the given time series. Stuck data is considered any four or more consecutive data points in the given time series where the values are identical. The boundaries for acceptable data are given in Table 4 per measurement type.

Table 4. Ranges of Accepted Data for Sensor QA Checks

	Minimum accepted value	Maximum accepted value
Irradiance sensor	>0	1300 W/m^2
Inverter power	>0	Mean value + 3 standard deviations
Ambient temperature	−40 °C	50 °C
Module temperature	−40 °C	85 °C

After determining erroneous or extreme data according to Table 4, the remaining data sets are applied to sequential QA checks specific to each type of sensor. If an individual sensor fails any one of the specific sequential tests, the sensor is rejected. Figure 6 shows an illustration of the sequential tests applied to irradiance sensors. A similar sequence applies to power measurements while temperature checks are significantly different.

Irradiance and power QA sequence

Irradiance and power sensors are checked per the sequence shown in Figure 6:

1) Check for at least 25% remaining data after removing erroneous data (~50% remaining is typical).

2) Determine if the time series indicates a measurement on solar tracker or for a fixed orientation system. The sensor is failed if the determination of “tracking” or “fixed” does not match the specified orientation of the given sensor. The sensor is also failed if the fits to determine “tracking” or “fixed” do not pass a goodness of fit test, deemed “Inconclusive” orientation. Qualitative analysis across a large number of systems supports that an “Inconclusive” orientation result for an irradiance sensor occurs primarily due to the following reasons: sensor shading, erroneous sensor outputs that do not correlate with time-of-day changes in irradiance, significant missing data, or generally noisy data.

3) Daily clear-sky irradiance totals are generated for the location and orientation of the irradiance sensor under test. The daily clear-sky totals are then used to set an upper and lower expected boundary for the daily measured irradiance total. The total number of days in exceedance of the upper and lower boundary are determined. A sensor is rejected if 33% of the total days exceed the upper boundary or 33% of the total days exceed the lower boundary.

4) The timestamp of the sensor is then checked for discrepancies over the time period in the data set. This check involves separate steps. First, all November–February and all May–August data are fit separately to determine peak output times for the sensor under QA for the winter and summer periods. Summer and winter periods are evaluated separately to test for daylight savings shifts occurring in the timestamp as well as seasonal changes in shading. The determined summer and winter peak output times are then compared against clear-sky peak output times considering the sensor’s orientation, latitude, longitude, expected time zone, and reported daylight savings adherence. The sensor is rejected if the measured peak output time does not match the clear-sky expected value within 25 minutes, or if both winter and summer fits do not achieve suitable fit thresholds.

Qualitative analysis across a large number of systems suggests that poor summer or winter fits are primarily due to the following reasons: sensor shading in one or both seasons, erroneous sensor outputs that do not correlate with time-of-day changes in irradiance, significant missing data, or generally noisy data. If the sensor passes the above checks for timestamp discrepancies, each day in the complete time series is evaluated to determine if the day can be considered a “sunny day” (depending on daily a goodness of fit test). Peak output times are then compared for each sunny day against expected times from clear sky irradiance. The time series signal of deviation between the sensor measurement times and the clear sky times are then processed using a changepoint detection algorithm. Any periods with timestamp shifts greater than 15 minutes are considered for rejection or manual adjustment, dependent on the magnitude and length of the shift.

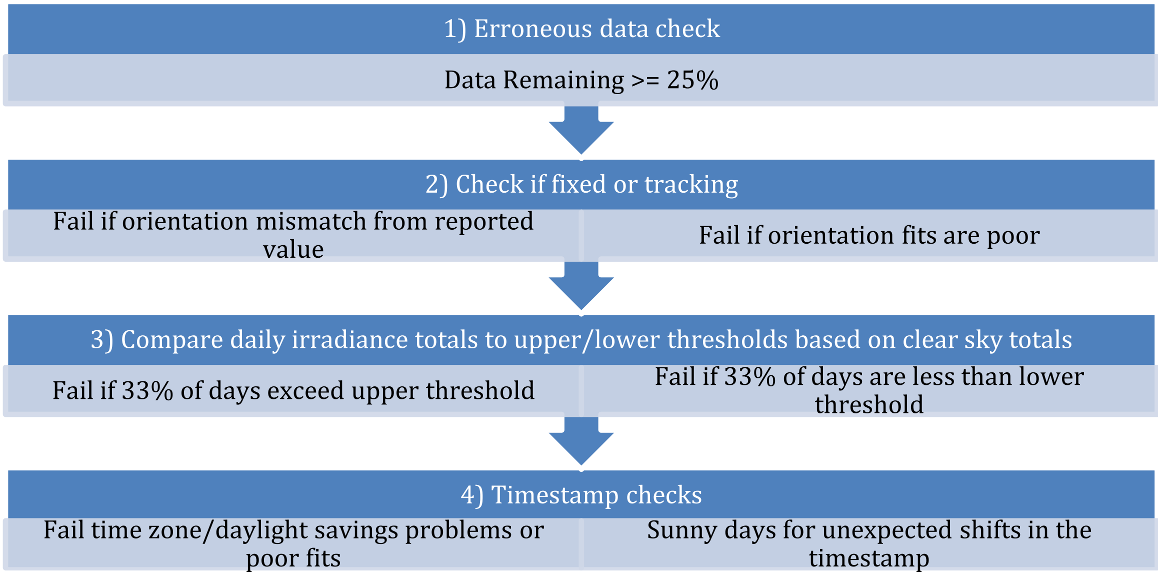


Figure 6: Flow chart of QA check sequence for irradiance sensors

The QA checks for power are slightly modified from those for irradiance. Prior to Step 2 (fixed or tracking), an algorithm is run on the power data set to determine if inverter clipping occurs. Clipping alters the power profile over the course of a day and therefore this information is input into the "tracking or fixed" logic in the step 2 QA. Step 3 is applied to daily power totals as compared to irradiance totals. Thresholds for the upper and lower bounds on the daily power totals are based on first applying a sinusoidal fit to the measured daily power totals rather than using clear sky data. Step 4 is identical for both power and irradiance data.

Temperature measurement QA

Erroneous data is first identified for module and ambient temperature per Table 4. Erroneous data comprising greater than 15% of the entire time series results in sensor rejection. A module temperature sensor fails if its mean value is outside the range of 5°C – 40°C, and ambient temperature sensor fails if its mean is outside the range of −5°C – 30°C. No additional checks are performed on ambient temperature, but a linear regression is performed between module temperature time series and a valid irradiance time series (meaning an irradiance sensor for the same PV system that has passed QA checks). If the correlation coefficient between the module temperature and a valid irradiance times series is less than 0.7, the module temperature sensor is rejected. If no valid irradiance sensor exists for the given PV system, the module temperature passes with no further checks.

System QA

While each individual sensor is subjected to the QA process described, a PV system must have at least one valid irradiance sensor, AC power measurement, and temperature measurement in order to be subjected to the RdTools ground sensor-based degradation analysis. To run the clear-sky-based degradation analysis, each system must have a valid AC power measurement and an available irradiance measurement.

Appendix B provides flow diagrams for selecting the irradiance, power, and temperature sensors with the highest data quality in each system. Sensors with the highest quality for each system are further subjected to the RdTools analysis. Each sensor must pass basic requirements for use in the analysis, including:

- 2+ years of data availability
- Pass initial QA checks outlined in Section 3.2

See Appendix B for further logic on selecting the highest quality measurement per a system.

After applying the logic described in Appendix B to each system, each system is further categorized into data quality tiers, which are used to determine which RdTools analysis to run:

1. Tier 1 System: System has at least one valid temperature, AC power, and irradiance measurement. The highest data quality measurements in each category are used to run a RdTools ground sensor-based analysis and clear-sky analysis.
2. Tier 2 System: System does not meet Tier 1 requirements but has at least one valid AC power measurement. Additionally, the system has an irradiance measurement lasting more than two years (can pass or fail QA checks). An RdTools clear-sky analysis is run on Tier 2- classified data.

3.3 Degradation Rate Assessment

The performance loss rate for each system in the fleet is estimated using the year-on-year approach as implemented in RdTools (version 2.0.0-alpha.0). The year-on-year approach compares the daily yield values normalized to a model (termed normalized daily yield), separated by exactly one year. The median of all these slopes is taken as an estimate of the underlying performance loss rate, and the uncertainty in the median is reported as a confidence interval. The reported confidence intervals do not account for sensor or measurement uncertainty. Performance loss rate is analyzed based on individual AC inverter power and energy meter data streams. We also perform the analysis on the aggregated power from each site's inverters and meters.

Two different analyses are considered for each power time series (inverter, meter, and the sums). The first is a sensor-based approach in which measurements made on site are used to model expected PV performance. The second is a clear-sky-based approach in which POA and module temperature are modeled based on expected weather assuming clear sky conditions [Jordan, 2017]. The sensor-based approach generally yields tighter confidence intervals in the analysis, but these confidence intervals do not capture possible bias from drifting sensors. The clear sky approach is more robust against drifting irradiance and temperature sensors, but it is susceptible to bias due to year-to-year atmospheric condition variations, especially on shorter data sets.

The systems are classified into two tiers of systems, as described in Section 3.2, for which we perform degradation analysis. For Tier 1 systems, both the sensor-based and clear-sky-based analyses are performed. For Tier 2 systems, only the clear-sky analysis is performed.

The RdTools analysis requires time series of plane-of-array (POA) irradiance, module temperature, and power. Time series data is first regularized by interpolating onto a regular time index with a period determined by the median period of the data set. If no QA-passing

POA irradiance measurements are available for a given system, POA is modeled based on measured global horizontal irradiance (GHI) according to the Erbs diffuse radiation model [Erbs 1982] and the isotropic sky model [Hottel 1942]. If no QA-passing module temperature measurement is available, it is modeled from ambient temperature measurement according to the thermal model of the Sandia Array Performance Model [King 2004]. However, due to low data availability and the relative insignificance of wind speed to thermal modeling, wind speed was set to zero in the thermal model for all systems. After the temperature and irradiance input types are determined based on this logic, the sensors with the highest QA score for each type are used as input into the Tier 1 RdTools analysis. The median measured or modeled (from GHI) value of POA at each time step is used to detect clear-sky conditions for the tier 2 analysis.

The first step in the RdTools workflow is to normalize the energy associated with each time step in the data set to modeled energy based on a simple temperature and irradiance model for power:

$$P_{model} = \frac{G_{POA}}{1000\ Wm^{-2}} P_0(1 + \gamma(T_{mod} - 25°\text{C}))$$

Where P_{model} is the modeled power, G_{POA} is POA irradiance, P_0 is the system's DC nameplate capacity, γ is the temperature coefficient of power and T_{mod} is the module temperature. The temperature coefficient γ was determined by examining a system's module metadata. If the module metadata indicated thin-film modules, the value was set to the First Solar Series 3 value of −0.25%/C, and otherwise set to −0.47%/C to match the "standard" module value in NREL's PVWatts model [Dobos 2014]. If multiple temperature coefficients are inferred from a system's module metadata, the median of the values is used. The next step is to filter the high-time resolution normalized data using built-in RdTools filtering. Data are filtered to remove:

- Normalized energy less than or equal to 5%
- POA outside the range of 200 to 1,200 W/m^2
- Module temperature outside the range of −50 to 110°C
- Points affected by clipping and power values below 0.01 W.
- *For the clear sky analysis only*: points that occur in non-clear-sky conditions

The third step is to use an irradiance-weighted mean to aggregate the remaining normalized energy values to daily frequency. Finally, the year-on-year analysis is performed to estimate the energy yield degradation rate and the associated confidence interval.

4 Results

4.1 Statement of Uncertainty

The calculation of Rd in an automated fashion makes sense only when data shifts and system outages are corrected. While we have conducted a limited data quality screening, the fleet distributions described below are affected by outliers at both ends of the distribution. This will influence distribution statistics, particularly P90 values, which are not reported here. Systems displaying large magnitude Rd values may be due to temporary changes in system configuration or inverter outage rather than unrecoverable module degradation.

Other sources of uncertainty for individual system analysis includes lack of a detailed model for system performance. Our PVWatts performance model is necessarily limited due to the nature of the production data and metadata available to us. This may influence the overall accuracy of Rd calculation for a single system, although this effect is mostly covered by the reported confidence interval (CI) in the detailed system-level Appendix A.

4.2 Partner Fleet Level Results

Fleet-level Rd results are collected for the combined fleet. We have analyzed hundreds of QA Tier 1 inverter-level subsystems, which met all data quality checks for irradiance, temperature, and AC inverter power, and provided their degradation distributions in Figure 7. This sensor-based analysis uses site-measured temperature and irradiance data, as described in Section 3.3. The median degradation rate for this group of systems is preliminary and subject to change with updated methods and fleet composition. We have also analyzed thousands of QA Tiers 1 and 2 inverter-level data streams using our clear-sky degradation methodology based on modeled (not site-measured) temperature and irradiance. Agreement with the sensor-based analysis is good, providing an important cross-validation of the degradation methodology.

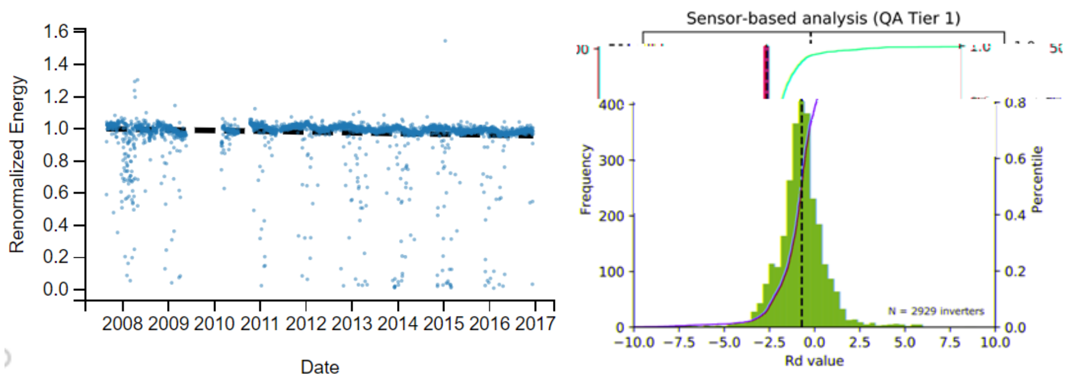


Figure 7: Example of renormalized energy plot from RdTools (left) and fleet-level Rd distributions for all systems using AC inverter data (right); QA-compliant Tier 1 systems analyzed with site-measured irradiance and temperature

In general, there appears to be good agreement between the clear-sky analysis distribution and sensor-based analysis. The small difference gives good confidence in the adequate calibration and cleaning status of pyranometers and reference cells for these systems. Parenthetically, for those systems where high quality pyranometers were deployed, we found better agreement with clear-sky values, although this may just be a factor of O&M budget at larger sites.

In addition, the median degradation value for the fleet is in-line with previously published median system degradation rates of –0.6% to –0.9% /yr (Jordan, 2016). While some systems appear in the tails of the above distribution, detailed analysis in Section 4.4 will show that many of these entries arise from data quality issues or system configuration changes not corrected by our automatic data assessment routines. Some large negative Rd values may indeed be cause for action on individual systems; this would be highlighted by detailed analysis of associated inverters and meters.

4.3 Detailed Fleet Distribution Statistics, Based on Inverter Data

System Capacity

Figure 8 depicts a scatterplot of system degradation (%) versus overall system capacity, given on a logarithmic scale. The data is color-coded based on analysis type with three categories: Tier 1 sensor analysis, Tier 1 clear-sky analysis, and Tier 2 clear-sky analysis. Overall system capacity for all analysis types is heavily concentrated between 100 kW and 1 MW. Degradation Rd percentage is heavily concentrated between 0% and −2.5% across all analysis types. Specific trends of degradation versus system size are currently under investigation.

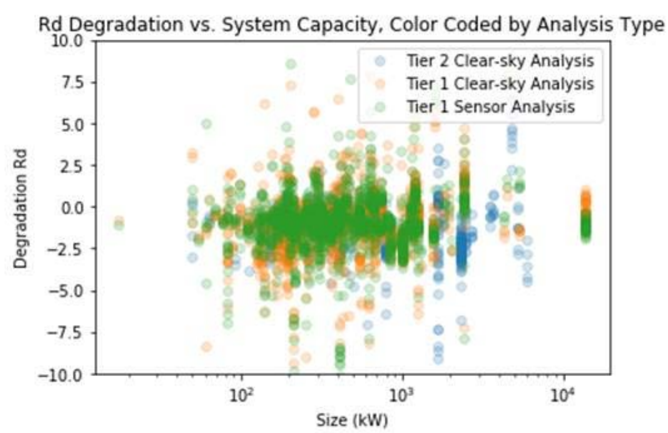


Figure 8: Scatterplot of Rd degradation vs. system capacity, where data points are color-coded on analysis type (Tier 1 sensor analysis, Tier 1 clear-sky analysis, and Tier 2 clear-sky analysis). Outliers are omitted.

System Size and Age

Figure 9 below depicts a series of histograms displaying system degradation by system size and age, with three size categories: >1 MW, 0.3-1 MW, and <0.3 MW; and three age categories: 2 – 3 years, 3 – 5 years, and > 5 years. Median degradation for all age categories is between −0.6 to −0.7% / year for sensor-based analysis. Median degradation for all size categories is between −0.5 to −0.8% / year for sensor-based analysis.

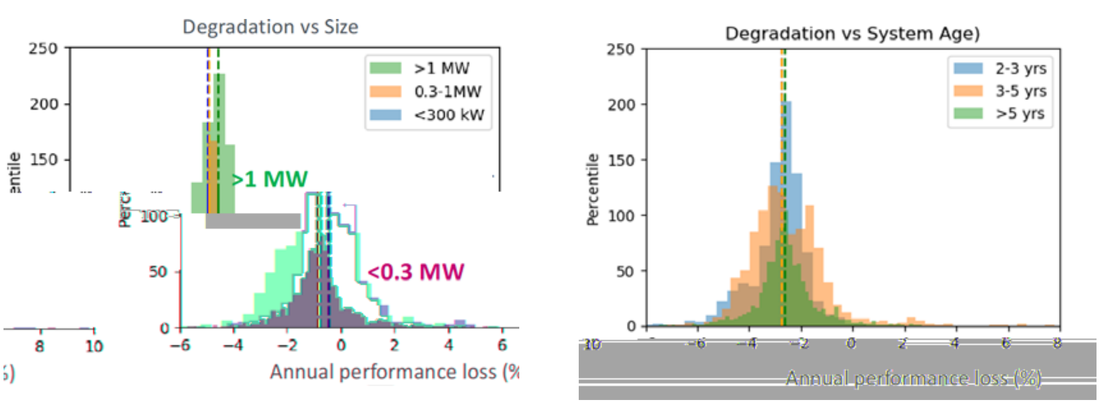


Figure 9: Rd histograms for sensor-based analysis, with outliers omitted. Data is binned in three system size categories (left) and three age categories (right). **PRELIMINARY RESULTS SUBJECT TO CHANGE**

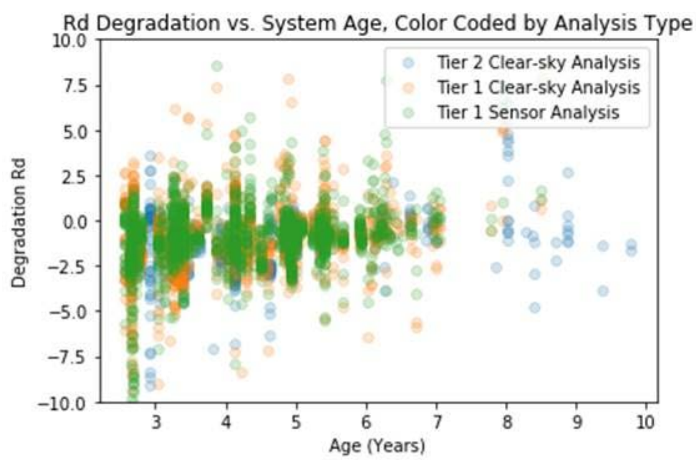


Figure 10: Scatterplot of Rd degradation vs. system age, where data points are color-coded by analysis type (Tier 1 sensor analysis, Tier 1 clear-sky analysis, and Tier 2 clear-sky analysis). Outliers omitted.

A scatterplot of system degradation rate vs. system age is depicted in Figure 10. The data is binned by analysis type, with three categories: Tier 1 sensor analysis, Tier 1 clear-sky analysis, and Tier 2 clear-sky analysis. System age is heavily concentrated between 2.5 and 7 years; degradation rate is heavily concentrated between 0 and -2.5% for all three analysis categories.

A slight trend is apparent, suggesting older systems (> 5 years) are correlated with a more modest degradation rate. There may be several explanations for this, including primarily that our RdTools confidence intervals are better when systems have more data. Other possible contributors may also include initial system start-up issues or nonlinear degradation (e.g. LID, LeTID), although the results are not conclusive at this point.

Meter-based Rd Analysis

Rd analysis based on system revenue-grade meters was also conducted and compared with the inverter-based analysis above. The meter-based analysis shows more modest degradation values compared to what we found in the inverter analysis above. The causes of this discrepancy may be related to inverter outages, which are properly detected and removed in the inverter-level analysis, but may erroneously influence the meter-level analysis. This is because inverter outages occur more frequently early in system life, therefore artificially depressing initial performance and resulting in positive degradation rate slopes (Figure 11).

Identification of inverter outages for the meter-based assessment may improve the agreement between these two analysis approaches.

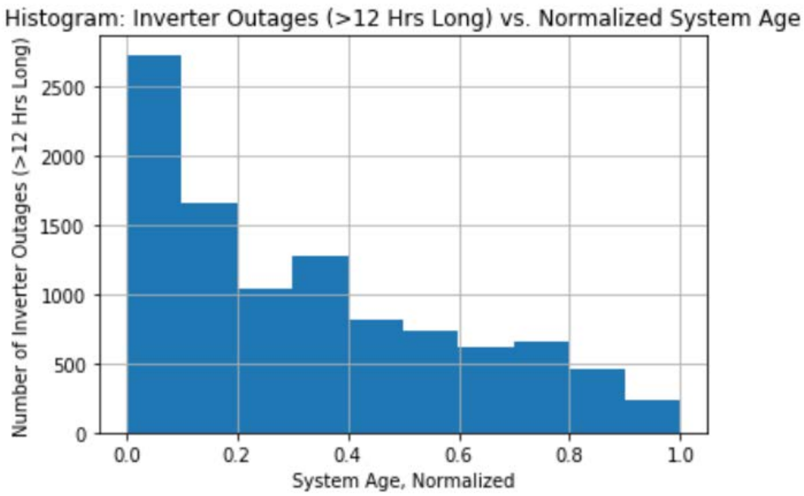


Figure 11: Inverter outage detected by periods of zero production vs normalized system age. Artificially depressed initial production leads to erroneous degradation rate if not properly detected.

4.4 Individual Case Study Results

Here we provide details on specific individual systems for illustrative purposes. The first case, shown in Figure 12, is a well-behaved example from a 400kW inverter at a California site. This inverter shows no evidence of downtime or data issues and gives results in the middle of the overall distribution with a median performance loss rate of −0.81%/yr.

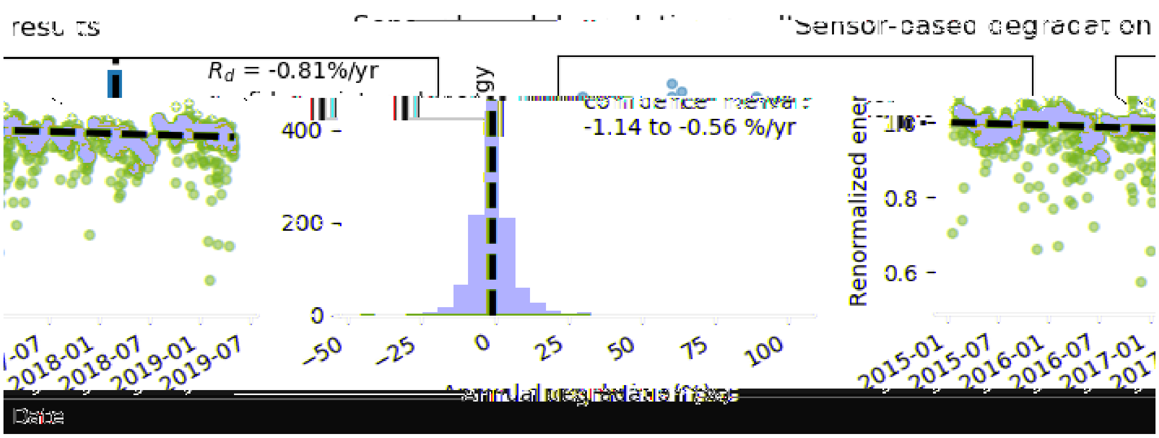


Figure 12: RdTools degradation summary plots. The left plot shows inverter production normalized against expected production over time, with no major deviations from the trend from downtime, soiling, or other sources of production loss. The right plot shows the distribution of year-over-year degradation rates extracted from the normalized production data. The black dashed lines reflect the median degradation rate.

The second example (shown in Figure 13) is of a system with results biased from soiling accumulation on the panels. These are the results for a 200kW inverter in the Los Angeles area. The severe loss feature (likely soiling) in 2018 caused the analysis to calculate rapid year-over-year degradation from 2017 to 2018. However, the analysis data set extended only partially through 2019, excluding the presumed performance recovery after removal of soiling from the performance loss rate distribution. This significantly biased the calculation for a performance loss rate estimate of −10.3%/yr.

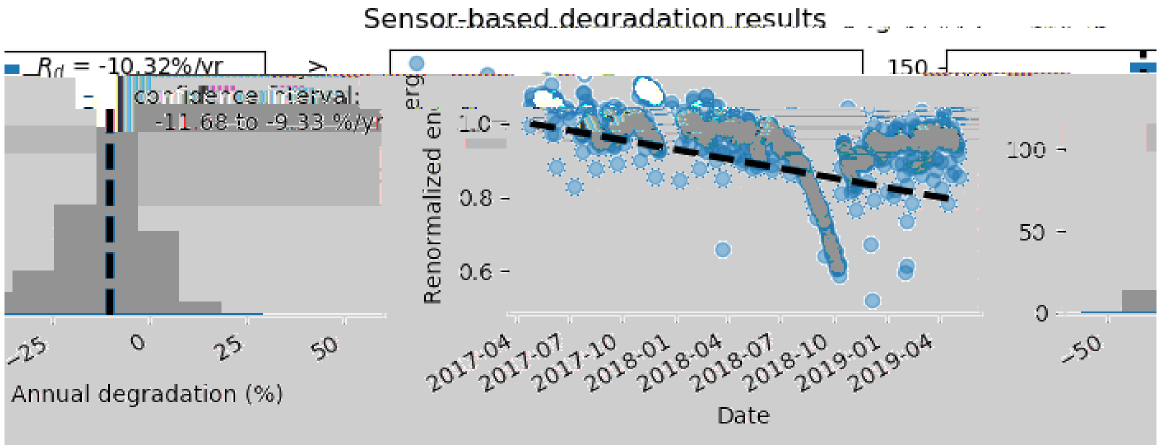


Figure 13: An example of a system with significant soiling. The normalized production signal drops significantly as soiling accumulates on the array, skewing the distribution of degradation rates.

The third example (shown in Figure 14) is from a nearby LA system. The array feeding this inverter had reduced performance in the first year of its data set, potentially due to downed strings. The increase in production capacity between years 1 and 2 causes an apparent increase in performance, resulting in a calculated positive performance loss rate of 28.8%/yr.

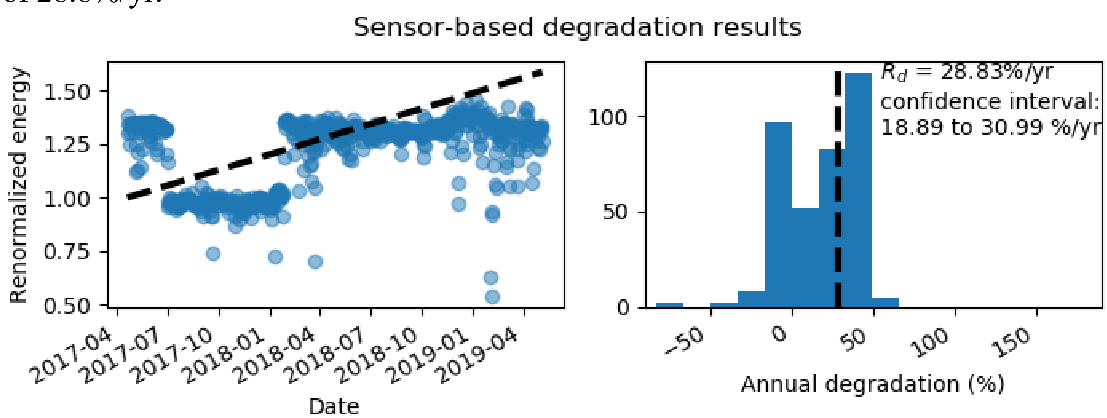


Figure 14: An example of a system with reduced array performance. The reduced performance causes apparent regeneration between years 1 and 2, resulting in a bimodal degradation rate distribution and an apparent positive degradation rate.

Capacity change is a source of many of the outlier performance loss rates. Because capacity reduction manifests at the inverter level, it can also often lead to spread in performance loss rate estimates for a single system. Table 6 shows a selection of the outlier results in the Fleet data set and associated root causes:

5 Loss Factor Analysis (Under Development)

In addition to degradation rate over time, different temporary or seasonal loss factors can be investigated that tend to be excluded from the RdTools analysis. These include inverter outage, clipping, soiling, and temperature-corrected Performance Ratio (PR). Best practices for these analysis methods are currently in development.

5.1 Inverter Availability

To detect specific inverter outages, we incorporate a comparison with system meter data to distinguish true inverter outages from communications outages. This method is the subject of an upcoming PVSC publication. In addition to identifying periods of unavailability, the method can estimate the energy loss associated with the downtime.

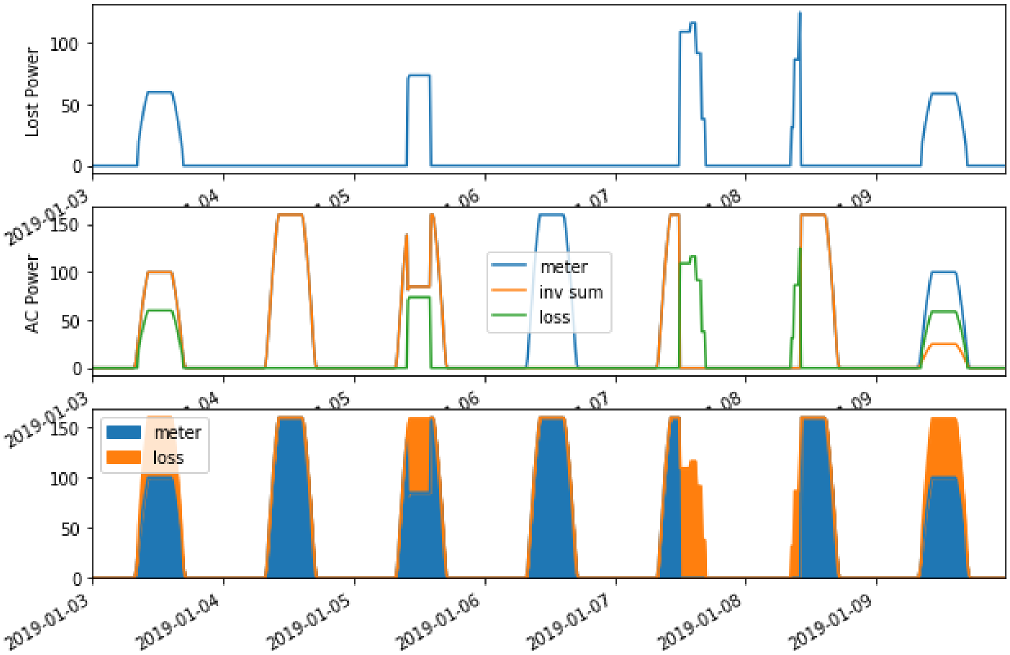


Figure 15: An example inverter availability analysis showing robust loss estimates despite inverter-level communications outages

5.2 Soiling

RdTools includes a soiling loss factor module, which identifies periodic loss and recovery trends which may be attributable to seasonal soiling. The method is still in development and has requirements for high quality irradiance and power sensor data. The method is described in Deceglie et al. "Quantifying soiling loss directly from PV yield" JPV, 2018 [Deceglie 2018].

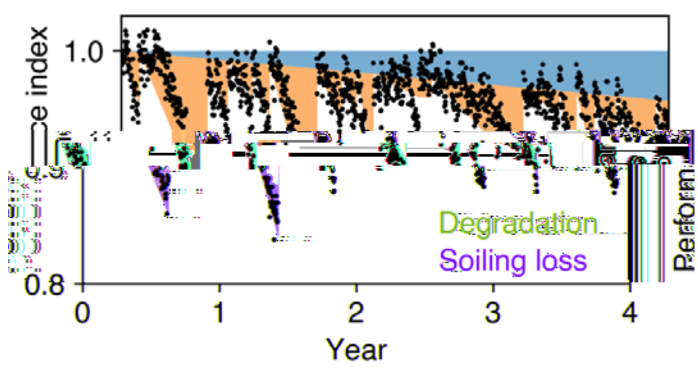
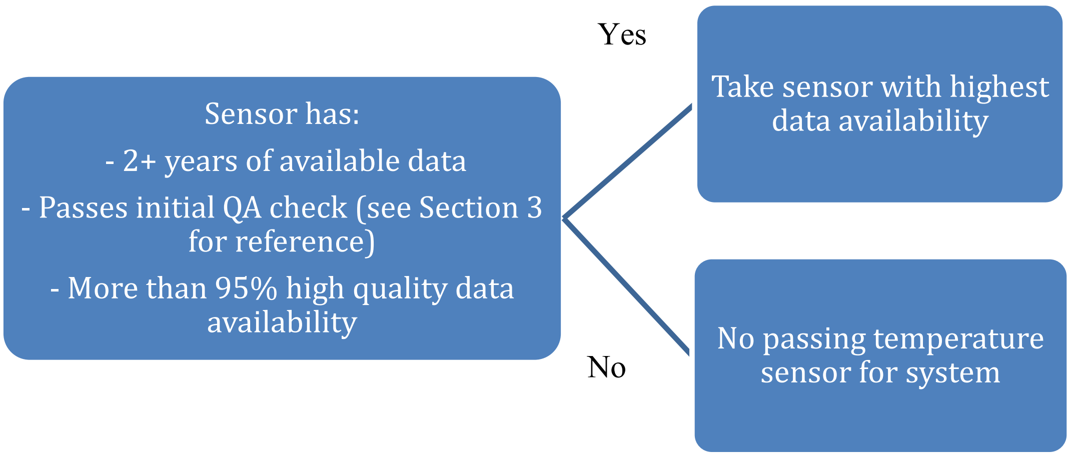


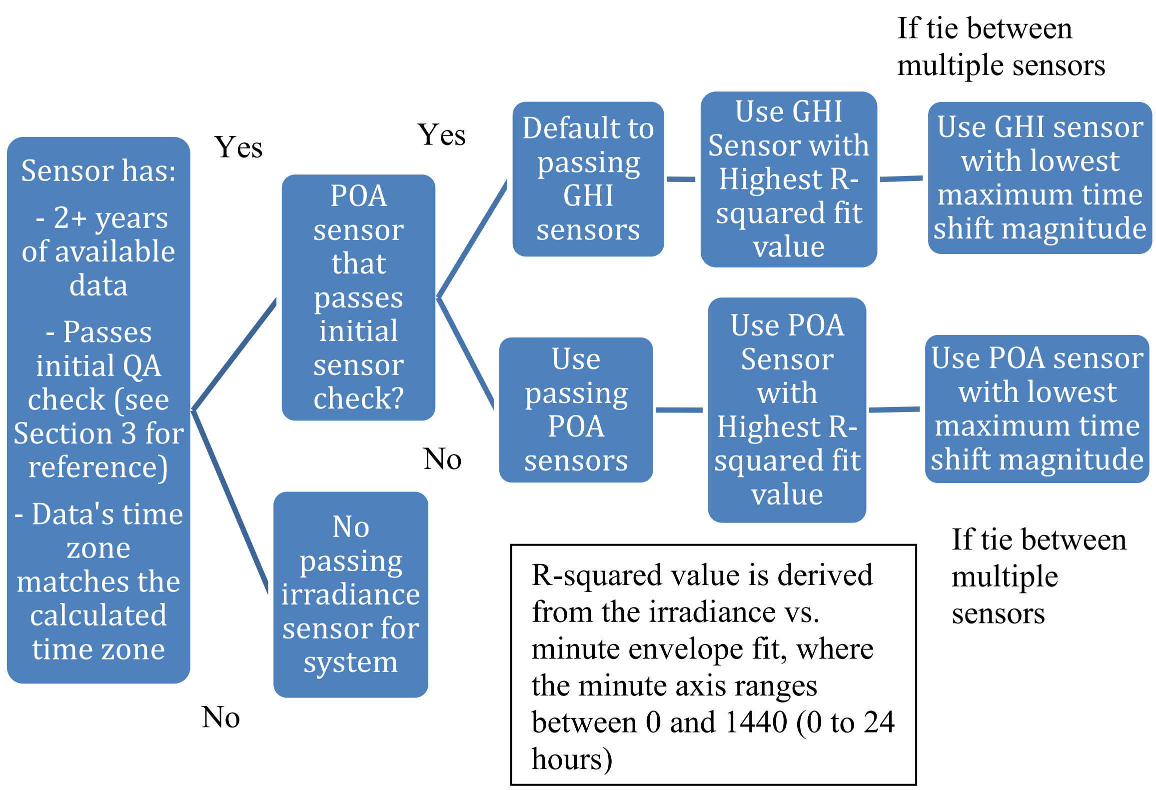
Figure 16: Example hybrid soiling and degradation analysis showing how underperformance can be partitioned between soiling and degradation by the algorithm. These results are for simulated PV data.

Appendix A. Logic for Selecting the Best Sensors for Each System

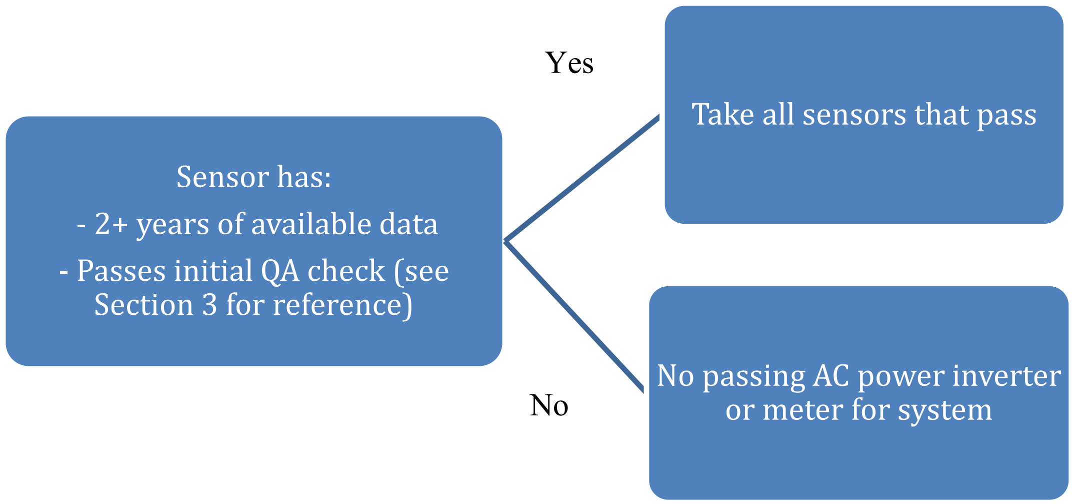
Logic for Acquiring Highest Data Quality Temperature Sensor in a System



Logic for Acquiring Highest Data Quality Irradiance Sensor in a System:



Logic for Acquiring Highest Data Quality AC Power Sensor(s) in a System



References

D.C. Jordan, S.R. Kurtz, K. VanSant, and J. Newmiller, “Compendium of photovoltaic degradation rates,” *Progress in Photovoltaics*, 24 pp. 978-989, 2016

D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, "Robust PV Degradation Methodology and Application*," IEEE Journal of Photovoltaics*, 8:2, pp. 525-531, 2018.

D. G. Erbs, S. A. Klein, and J. A. Duffie, “Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation,” *Solar Energy* 28(4), pp 293-302, 1982. Eq. 1

H.C Hottel and B.B Woertz, “Evaluation of flat-plate solar heat collector,” *Trans. ASME* 64, 1942.

D.L. King, W.E. Boyson, and J.A. Kratochvill, “Sandia Photovoltaic Array Performance Model”, *SAND Report 3535*, Sandia National Laboratories, Albuquerque, NM. 2004.

T. Karin, C.B. Jones, and A. Jain, “Photovoltaic Climate Zones”. *IEEE Photovoltaic Specialist Conference 46*, Chicago, IL., June 2019 – See also https://pvtools.lbl.gov/pv-climate-stressors

RdTools, version 2.0.0-alpha.0, https://github.com/NREL/rdtools, DOI:10.5281/zenodo.1210316

M. G. Deceglie, L. Micheli and M. Muller, “Quantifying Soiling Loss Directly from PV Yield,” in IEEE Journal of Photovoltaics, 8(2), pp. 547-551, 2018