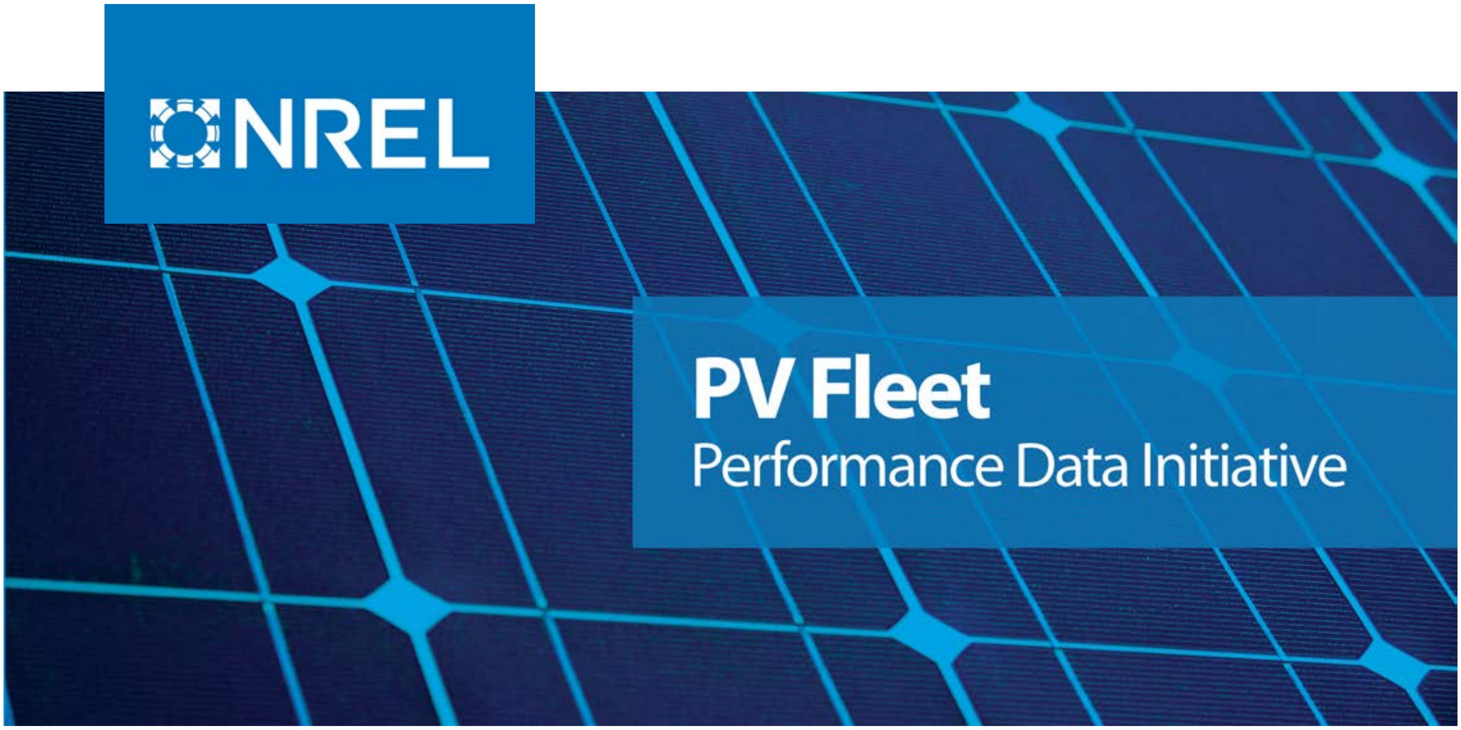


# Availability and Performance Loss Factors for U.S. PV Fleet Systems

Chris Deline, Matt Muller, Robert White, Kirsten Perry, Martin Springer, Michael Deceglie, and Dirk Jordan

*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-5K00-88769
February 2024

# Availability and Performance Loss Factors for U.S. PV Fleet Systems

Chris Deline, Matt Muller, Robert White, Kirsten Perry, Martin Springer, Michael Deceglie, and Dirk Jordan

*National Renewable Energy Laboratory*

**Suggested Citation**
Deline, Chris, Matt Muller, Robert White, Kirsten Perry, Martin Springer, Michael Deceglie and Dirk Jordan. 2024. *Availability and Performance Loss Factors for U.S. PV Fleet Systems*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5K00-88769. https://www.nrel.gov/docs/fy24osti/88769.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-5K00-88769
February 2024

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

## NOTICE

This work was authored 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. This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number 38258.

# Acknowledgments

We would like to thank all of the corporate partners who provided the PV system data used to perform the analyses in this document. This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number 38258.



Thank You to SETO and our PV Fleet Partners!

# List of Acronyms and Abbreviations

| | |
|---|---|
| AC | alternating current |
| CODS | COmbined Degradation and Soiling algorithm (implemented in RdTools) |
| DC | direct current |
| DHI | diffuse horizontal irradiance |
| DOE | Department of Energy |
| E | energy |
| G | irradiance |
| HSAT | horizontal single-axis tracker |
| IEC | International Electrotechnical Commission |
| kW | kilowatt |
| L | lost energy derate factor [fraction] |
| LID | light-induced degradation |
| MW | megawatt |
| NOCT | nominal operating cell temperature |
| NREL | National Renewable Energy Laboratory |
| NSRDB | National Solar Radiation Database |
| P | power |
| PI | Performance Index |
| PLR | Performance Loss Rate |
| POA | plane-of-array |
| PR | performance ratio |
| PV | photovoltaic |
| PVLIB | PV energy modeling software |
| PVWatts | PV production software |
| RdTools | Degradation, availability and soiling software |
| STC | standard test condition (25ºC and 1000 $Wm^{-2}$) |
| $T_{cell}$ | cell temperature |
| $\gamma$ | module temperature coefficient |

# Executive Summary

In the PV Fleet Performance Data Initiative, we partner with photovoltaic (PV) fleet owners to collect time-series PV production data and publish aggregated, anonymized results. This report is an update of our previous publications, specifically a FY 2021 performance index publication and a FY 2022 fleet degradation analysis.

In this analysis, we have increased our data participants and system totals by around 10% to 8.5 GW and 24,000 separate inverter data channels. Four major analysis topics are considered in this report—Performance Index (PI) trends, PV system availability, soiling losses, and PV system degradation.

Performance Index and inverter availability are assessed on a larger set of data from our FY 2021 report: 1,128 systems compared with 200 systems from before. The increased number of systems is due to an improved data quality methodology, as well as introducing new systems to the analysis. Overall results are similar to previously published values—overall inverter availability is low in the first six months of system performance before reaching steady-state by the end of the first year. Excluding this six-month startup period, system-level aggregated data shows a median (P50) system availability of 0.99 and a lower $10^{th}$ percentile (P90) value of 0.95 (Figure ES-1). A dependence on system size is also demonstrated, with worse inverter availability results for larger PV systems. Causes of this effect are under investigation, but may be impacted by inverter size, which also show lower availability for larger inverter sizes. This report also investigates PI, correcting for degradation, soiling, snow, and availability. Following these corrections, the median system PI over its entire lifetime is 0.95. PI values reported here are approximately 3% lower than what we presented in our previous FY 2021 report.

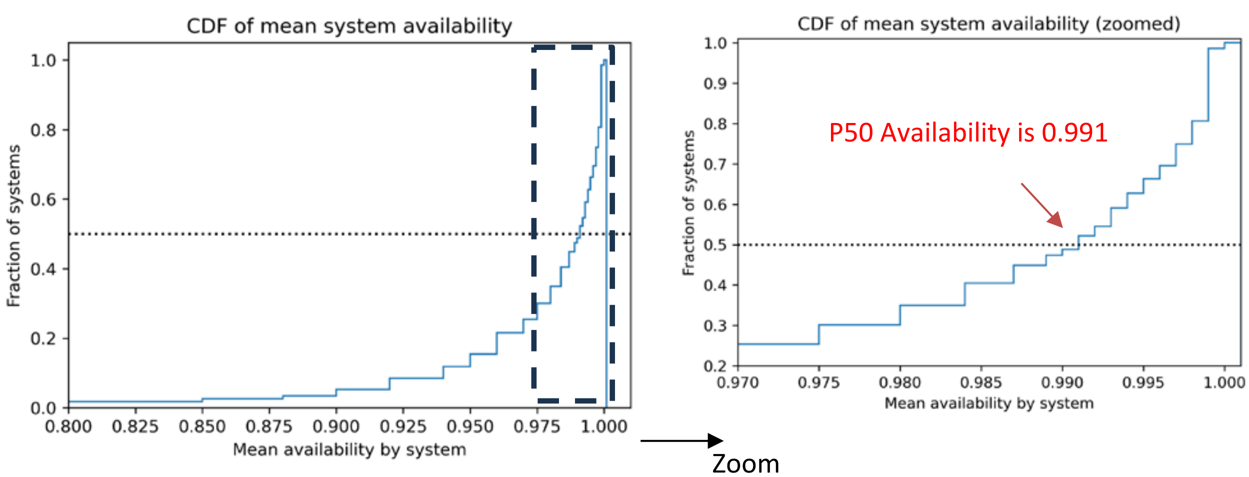


**Figure ES-1. CDF of inverter availability by system is 0.99 at P50 level and 0.95 at P90 level. Right figure shows a zoomed-in plot around the P50 value.**

Soiling loss is assessed in a comprehensive way for the first time in this report. Results are presented using the COmbined Degradation and Soiling (CODS) method, as implemented in RdTools (v3.0.0a4). Soiling values are presented for 255 systems, which indicated irradiance-weighted soiling loss greater than 1%. The values have been published in an updated NREL soiling map at nrel.gov/pv/soiling.html.

Finally, we investigated system degradation using three different data analysis techniques: conventional RdTools (year-on-year (YOY)), CODS, and Performance Loss Rate (PLR) analysis. Overall degradation results are consistent with our previous publications. Rerunning conventional RdTools on our updated fleet shows that some data partners have systematically fallen below the median system degradation rate (change over time) of -0.75 %/year. A comparison with PLR analysis, which looks at change in annual PI over time, shows that median system degradation is consistent with -0.5% to -0.75% per year change. However, at the P90 value, system degradation is substantially faster. These two results are consistent and indicate that resulting degradation statistics depend to a great degree on the population of PV systems making up the analysis cohort and whether soiling impacts the systems. The use of CODS for degradation analysis provides a different method for degradation assessment, which explicitly excludes the impact of recoverable soiling on degradation analysis. Excluding soiling effects yields an annual system degradation around -0.5% per year on average. This indicates that a portion of system performance loss may be attributed to periodic soiling that is not fully recovered.

This report provides PV system owners/operators with background and methods to analyze PV system performance, give guidance for expected cohort performance, and performance loss values for use in pro-forma financial models, which guide new-build system design and bankability reports.

# Table of Contents

# 1 Introduction

Rapid deployment of solar PV systems to supplant carbon-producing electricity alternatives is imperative to meet carbon reduction targets [Department of Energy, 2021]. PV is also the most cost-effective source of new electricity generation in many cases [Lazard, 2023]. A major assumption of future-year projections of solar PV project financials is long-lived performance with project lifetimes of 30+ years and only modest annual declines in performance [Jordan, 2012]. In particular, with the anticipated scale-up of solar PV manufacturing to a Terawatt scale [Haegel, 2023], communicating the regional or technology-based differences in PV performance will allow for tailored solutions as systems are widely deployed. In the PV Fleet Performance Data Initiative [Deline, 2020], a U.S. Department of Energy (DOE)-sponsored program, we use 3rd-party data supplied under NDA and anonymized and aggregated to develop public reports on overall U.S. fleet performance to ensure that systemic risks in the U.S. PV fleet do not go undetected.

PV system performance can be reported in several ways. We focus on Performance Index (PI), which is a monthly sum of system AC energy divided by a weather-corrected estimate. This approach was taken in an earlier report to look at monthly cumulative system performance relative to modeled expectation [Deline 2021]. This PI analysis identified trends in performance and inverter availability, including that median system performance was at 0.994 of expected, and 90% of systems performed within 10% of expected. Additional trends of inverter availability were also reported, finding that availability in the first six months of system operation was depressed but tended to reach steady state within the first year of operation. Also, empirical snow loss factors were identified based on the cumulative monthly snowfall received.

System performance loss over time is also a crucial parameter to understand because PV system financial returns are dependent on stable long-term performance. Typical system losses include recoverable losses, such as snow cover, downtime for maintenance, curtailment, among other factors, and non-recoverable losses like module material degradation. Soiling loss can be recoverable or non-recoverable depending on the soiling mechanism and maintenance particulars [Bessa, 2024]. Performance Loss Rate (PLR) is defined to be a system-level combination of all recoverable and non-recoverable factors, determined from annual change in the PI [Deceglie 2023, Lindig 2022, French 2021]. This high-level metric is useful in that it is closely related to the financial performance of an asset, but it is not useful for diagnostic or predictive purposes, in contrast to other techniques to isolate individual degradation factors from system performance. However, PLR represents an upper bound to maximum degradation, which is instructive, so we will take it up in Section 6.

In prior publications we have summarized the annual degradation of PV systems using the open-source software RdTools [Jordan 2018], which attempts to isolate nonrecoverable loss factors by excluding some seasonal and recoverable effects. For instance, the method tends to be insensitive to the effects of inverter clipping, outages, seasonal soiling, and other temporary losses which are recovered in later years. Applying this method to our PV Fleet systems, we found a median system degradation rate (Rd) of -0.75%/yr [Jordan 2022]. This previous report also identified climate-dependent differences in system degradation, ranging from -0.5%/yr in temperate zones to -0.88%/yr in hotter areas. Other technology-based differences were also investigated, such as CdTe vs Si module types, or fixed-tilt vs tracked. Within the uncertainty of the data, no differences were identified for these various system configurations.

The current report can be considered an update to these previous publications. We have conducted additional analysis on an expanded PV fleet following introduction of 500 new systems into our database, bringing the total to 8.5 GW and 24,000 separate inverter data channels. Four major analysis topics are considered here—**Performance Index (PI)** trends; **System availability** focused on both fleet and system-level roll-ups; **Soiling loss** using two different data analysis techniques (stochastic rate and recovery (SRR) and CODS); and **System Degradation analysis** considering three techniques: PLR, RdTools and CODS.

## 1.1 PV Fleet Composition

Performance data from our PV Fleet data partners are collected from more than 2,200 systems spread across the country (Figure 1). Here, a system is defined as one or more PV inverters combined through a single metered point of connection. Therefore, some large PV projects deployed in multiple stages could show up as multiple split PV systems in our analysis. Histograms of system sizes and types of PV indicate a variety of sizes, with median system size of 3.8MW DC and 5.2 years of data on average. Compared with our previous publication in [Jordan, 2022], this represents roughly a 10% increase in the number of sites and inverter channels considered. However, the system size, age, and technology distributions are roughly unchanged. The geographic extent of systems is also largely unchanged.

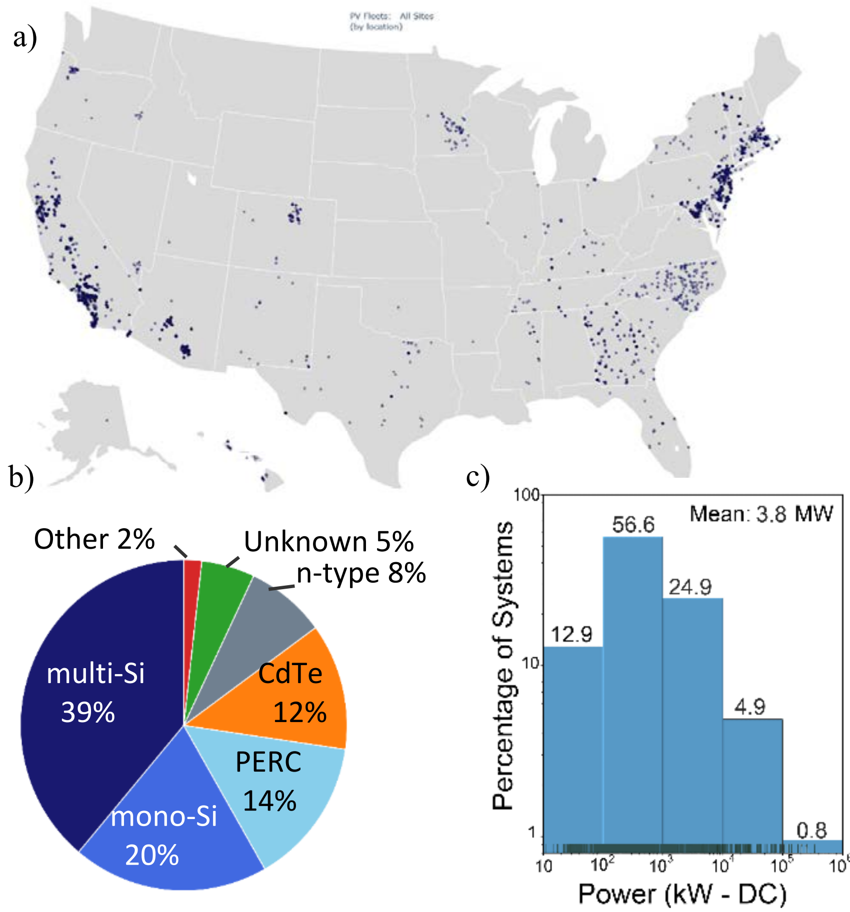


**Figure 1: Geographical distribution of the PV systems considered here (blue dots) (a). Technology pie chart by number of inverters (b) and system size distribution (c). Multi- and mono-Si indicate older Al-BSF module types. Details here are prior to data quality filtering.**

Time-series data are collected for both inverter-level and revenue-grade meter AC data, typically on a 15-minute time interval. Depending on the type of analysis, results are presented at the inverter-level (RdTools, soiling) or the system-level (availability, PI, PLR). Site metadata information, including module and inverter type, tilt/orientation, and location are used to generate corresponding expected energy values based on a PVLib-Python PVWatts energy model and NSRDB satellite resource data [Sengupta 2018].

Additional information about our methodology—including data quality checks, data warehouse information, and other processes—are documented in Deline (2021) and Jordan (2022).

## 1.2 Quality assessment and filtering

The aphorism "garbage-in, garbage-out" applies to large-scale automated analysis of fleet performance data. To attempt to exclude or correct common input data issues, we leverage the PVAnalytics data package [Perry 2022a, PVAnalytics 2023]. Leveraging the PVAnalytics package, each data stream was screened for outliers and data outages, and the longest data shift-free period for the time series was taken

[Perry 2022b]. Each data stream was then screened for time shift/time zone issues, and time issues were corrected if detected [Perry 2023a].

Additionally, each data stream time series was used to estimate azimuth and tilt [Perry 2023b] and mounting configuration (fixed tilt or tracking). After initial screening, the remaining 1,700 systems and 28,700 inverter and energy meter channels (81%) had sufficient data quality to be analyzed.

Additional filtering is required for each analysis method because each method depends on a different level of data quality or available instrumentation. For the availability analysis, inverter- and revenue-grade cumulative energy data are required for each system. This results in only 1,128 passing systems for availability and PI analysis. For soiling analysis, data quality is more rigorous, requiring manual verification of soiling interval data. For this run, only 255 systems contribute to soiling analysis. Finally, RdTools system degradation analysis requires two years of performance data, which resulted in 23,464 inverter and energy meter channels from 1,575 passing systems.

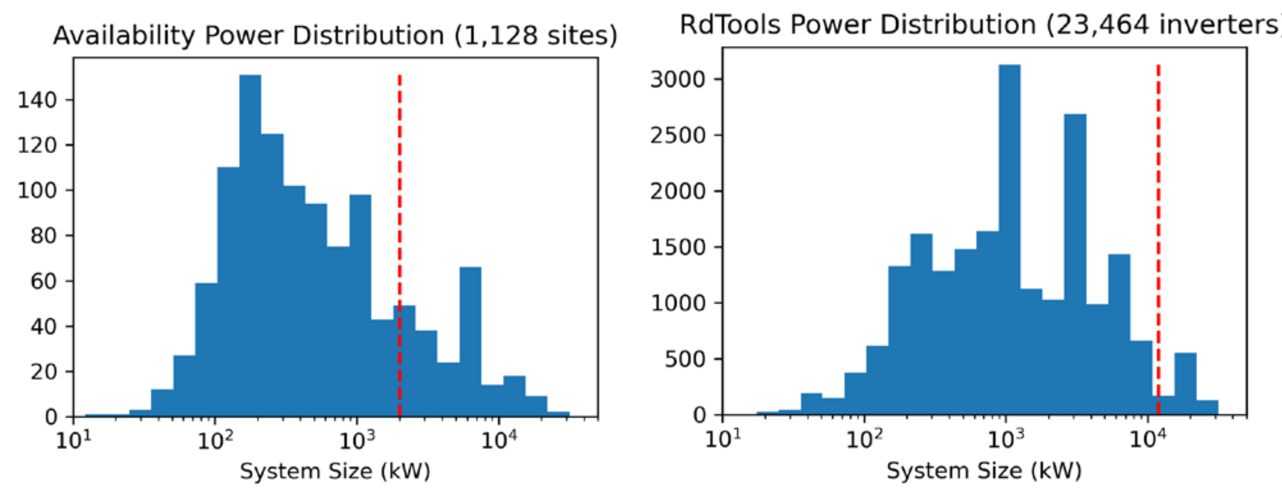


**Figure 2: Post-filtering system DC capacity distributions for availability analysis done at the system level (a) and RdTools (b) analysis at the inverter level. Red dashed line indicates mean system power for the distribution.**

# 2 Analysis Methodology

## 2.1 System Availability Methods

PV system availability is defined as the up-time of equipment available for production. We consider AC production of the system at the inverter level to be the level of focus. Therefore, subsystem outages of other equipment—such as trackers or DC strings, which do not result in the complete loss of an inverter's production—are not captured by this methodology.

For this analysis, we define inverter availability as an energy-weighted metric reflecting the production loss from inverter outages:

$$Availability = 1 - \frac{E_{outages}}{E_{outages} + E_{actual}} \tag{1}$$

To perform this estimate, we use the method published in Anderson (2020) and implemented in RdTools (2020), an open-source PV data analysis package developed by NREL. Note that this method only estimates loss when an inverter is completely offline; partial capacity reductions from clipping, thermal derating or partial curtailment are not classified as an inverter outage and do not count against availability. The same applies to partial DC-side outages; the inverter output must go to zero to count against availability. One aspect of this analysis method is that it is robust to communication outages—if an inverter data channel reads zero production but the cumulative energy meter shows full production, this period is logged merely as a data gap, not as an actual availability loss. The method includes two analysis routines, one for conditions when communication is available, and another where there is a communications outage.

The first approach targets partial outages where some, but not all, of the inverters in a system are not reporting data. This method operates time stamp by time stamp to detect inverter outages and estimate the lost production if applicable. The detection is performed by comparing the aggregated reported inverter power with the system meter power. In cases where an inverter is online but not reporting production because of a communication fault, the meter power will read correspondingly higher than the aggregated inverter power. In cases where the nonreporting inverter is truly offline, the meter power will align with aggregated inverter power. By examining the difference between meter power and aggregated inverter power, it is possible to estimate the offline fraction of system capacity and the associated power loss.

The second analysis routine targets complete outages where the entire system shows no production. In contrast with the first approach, this one cannot perform a peer-to-peer comparison and must rely on data outside the outage and external data. Any system production during the outage duration can be calculated by taking the difference in cumulative production data reported by the system meter from just before and just after the outage. This actual production is then compared against an expected production from an NSRDB-based expected energy model. If the actual production is sufficiently low (i.e., too low to be explained by the uncertainty in the energy model), the difference is assumed to be production loss associated with the outage and is distributed across the outage duration in proportion to expected production.

Figure 3 illustrates various combinations of the two outage types described above.

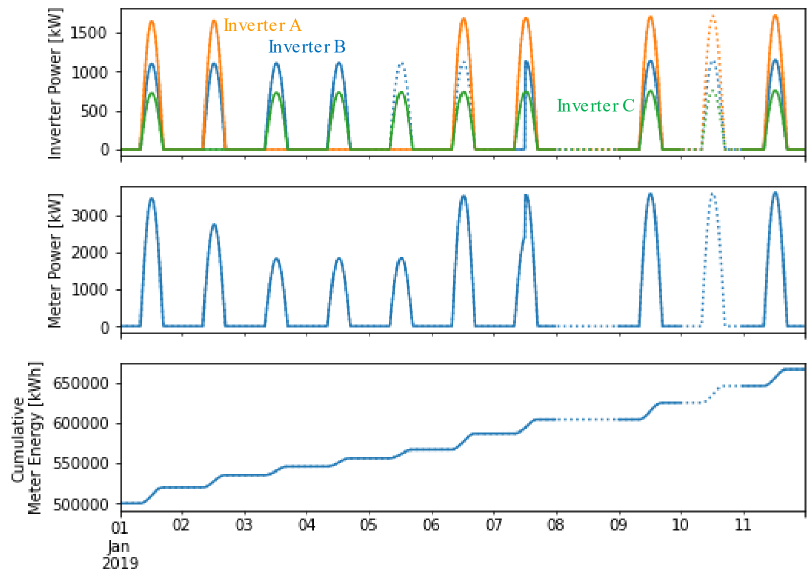


**Figure 3. Example data for availability analysis. Dotted lines show "true" data that would not normally be available to the analysis routine but are shown here for illustration.**

For example, day five has one inverter operating normally (green), one inverter tripped offline (orange), and one inverter producing but experiencing a communication outage (blue). This day would be handled by the subsystem outage routine because at least one peer's power data is available for comparison. In contrast, days eight and 10 show a real systemwide outage and a full communication interruption respectively and would be handled by the system outage approach.

## 2.2 Performance Index Methods

PV system PI is a comparison between actual system energy production and expected (modeled) energy production, calculated on a monthly basis (Eq 2).

$$PI_{month} = \frac{\sum^{i} E_{meas,i}}{\sum^{i} E_{exp,i}} \tag{2}$$

Measured production in the numerator is simply based on system-level revenue-grade meter Energy in kWh, summed over monthly periods.

Modeled results in the denominator are based on a PVWatts v5 production model using as-measured weather data for each time period $i$. While on-site weather data can have higher accuracy, the frequent cleaning, calibration issues and data outages of on-site solar resource measurement (on-site pyranometer) make automated analysis difficult. Instead, we use satellite-based weather data from the National Solar Radiation Database (NSRDB) [Sengupta 2018]. The NSRDB data set is available at various temporal and spatial resolutions across the United States. In this work we use the 4 km, 30-minute resolution. The use of NSRDB data avoids the complications of imperfect instrument maintenance.

The remaining modeling steps are implemented using the open-source PV modeling package, PVLIB-python, version 0.9.0 (Holmgren 2018; Holmgren 2019). The specific PVWatts algorithm is somewhat complicated, and accounts for a number of effects including temperature dependence, which is user-specified. Where known, we use the manufacturer's actual specified temperature coefficient. When unknown, we use the PVWatts default value of −0.47%/°C for standard silicon modules. The nameplate $P_{STC}$ power for the simulation is the system's DC rated power as reported in metadata from the site owners.

A complete accounting of the modeling steps are as follows:

1) 30-minute irradiance, ambient temperature, and wind speed data are retrieved from the NSRDB for the system location.
2) The solar position time series for the location is calculated using (Reda 2004).
3) For systems with horizontal 1-axis tracking arrays (HSAT), the time-dependent module orientation (surface tilt and azimuth angles) is calculated using the approach described in Lorenzo (2011) and Anderson (2020). All systems are modeled with backtracking activated using a ground coverage ratio of 0.4 if a known value is not provided with our system metadata.
4) Plane-of-arrary (POA) irradiance is modeled from the NSRDB irradiance components. Diffuse horizontal irradiance is transposed to POA with the Perez 1990 model (Perez 1990), and ground-reflected irradiance is calculated assuming a ground albedo of 0.25.
5) The beam component of POA irradiance is reduced to account for reflection off the module surface using Snell's and Fresnel's laws, assuming a module glass index of refraction of 1.526.
6) Cell temperature is modeled from POA irradiance, ambient temperature, and wind speed, using the transient thermal model described in Fuentes (1987), assuming a nominal operating cell temperature of 45°C. This value is consistent with the assumption of a polymer-backsheet PV module on an open rack.
7) Base PVWatts power is reduced according to system loss factors as described below in 2.2.1.
8) The PVWatts inverter model is applied, using the default reference efficiency of 96.37% and a nominal efficiency of 96%.
9) Power is curtailed to the system AC power limit. Because system AC power limit is not present in the available system metadata, it is calculated from the available time series data as the 99.9$^{th}$ percentile of system power
10) Monthly PI is calculated by dividing monthly integrated measured power by modeled power, per Eq. (2).

### 2.2.1 PVWatts loss factors

The loss assumptions in an expected energy model can be adjusted to customize the PI for a particular application. For this effort, we choose to include losses for uncontrollable effects like wiring loss and shading while omitting controllable losses like array soiling and inverter downtime from the expected energy. Table 1 shows a complete listing of the loss assumptions, with the default PVWatts recommendations for comparison.

Table 1. Expected Energy Loss Assumptions

| Energy Loss Term | PVWatts Default | PV Fleets Loss |
|---|---|---|
| Soiling | 2% | **variable** |
| Shading | 3% | 3% |
| Snow | 0% | **0%** |
| Mismatch | 2% | 2% |
| Wiring | 2% | 2% |
| Connections | 0.5% | 0.5% |
| LID | 1.5% | 0% |
| Nameplate | 1% | 1% |
| Age | 0% | **0.5%/yr** |
| Availability | 3% | **variable** |
| **Total** | **14.1%** | **8.23%** |

**Note: Loss factors in red are investigated separately.**

The total linear loss factor $L_{total}$ is calculated through a product of individual loss factors $L_j$:

$$L_{total} = 1 - \prod_j (1 - L_j) \tag{3}$$

In our modeled power, we separately handle some loss factors because we are able to isolate and apply site-specific values. This is the case for the Table 1 factors in red: Age, Snow, Availability and Soiling. Starting with the first factor, we assume an age-based loss factor of 0.5%/year that accumulates across the system data set, and we do not include it in the static total loss value.

Snow loss is covered in more detail in Deline, 2021, and is handled here by excluding months in which recorded snowfall is greater than 1cm. Therefore, the default PVWatts snow loss value of 0% is suitable for the remaining snow-free months. System Availability is calculated per Section 2.1 and applied to the monthly expected power. Soiling loss is assumed to be 0% unless the system has been identified to have measurable recoverable soiling loss, as described below. In this case, the monthly insolation weighted soiling ratio (IWSR) loss is applied to the monthly expected power.

## 2.3 Soiling Loss Methods

Previous versions of PV Fleets soiling loss analysis have been conducted on a subset of the fleet using the stochastic rate and recovery algorithm (SRR) algorithm separate from the RdTools workflow [Deceglie 2018]. In the current analysis the RdTools workflow was updated to include both the SRR and the combined degradation and soiling (CODS) models [Skomedal 2020]. CODS provides an iterative methodology which extracts signals for degradation, soiling, seasonality, and noise from within the PV power time series data; alternatively SRR only provides a soiling signal. Validation and comparison of SRR and CODS has been historically challenging due to the computational intensity of the CODS analysis, as well as limited truth data sets available for soiling. The first issue was solved by deploying our analysis on NREL's Peregrine High Performance Computing (HPC) cluster. The second issue was solved recently by NREL's development of a comprehensive synthetic PV data set designed for testing algorithms like SRR and CODS [Muller 2023]. SRR and CODS were compared to predict the soiling losses and soiling rates embedded in synthetic data. It was found that neither method performed particularly well when there was high residual seasonality in the daily performance index. For low (<15%) seasonality cases, there was a slight difference in method performance: CODS better predicted soiling **losses** while SRR better predicted soiling **rates**. As neither model was clearly dominant, it was decided to proceed with CODS with a full PV Fleets analysis, both because HPC incorporation offers this as a new opportunity and because decoupling soiling from degradation enables an investigation of how high soiling losses may impact system degradation rate calculations.

As noted, investigations with synthetic data demonstrated that CODS soiling loss results are highly inaccurate when high residual seasonality remains in the daily PI for a given inverter. For this reason all CODS soiling results were examined against plots of the daily PI. All plots with annual seasonality exceeding 15% were excluded from soiling analysis. Additionally, it was found that data outages or data shifts could cause unintended results from the CODS algorithm and therefore any inverter with such problems were excluded.

## 2.4 Degradation Rate and PLR Methods

Degradation Rate and Performance Loss Rate (PLR) are similar-sounding concepts which have a distinction in our treatment here. In (Deceglie, 2023), PLR is defined as the annual change in system PI, inclusive of recoverable and non-recoverable performance losses such as soiling, availability and curtailment. This is in contrast with the system degradation value reported by RdTools, which by design only focuses on certain non-recovered loss factors. It includes both DC degradation of the modules (degradation rate) as well as some system-level performance loss factors such as unrecovered soiling, string outages and tracker pointing failures, to the extent these factors are progressive over time (Jordan, 2022).

In Section 6, system degradation and PLR are reported in several ways. A simple assessment of PLR is calculated by taking the system-level performance index (PI) of each system on a monthly basis, and looking for the linear trend of the annual PI average over time. A single value per system is calculated using an approach similar to the YOY method of RdTools where annual performance changes from year to year are each calculated and the median value for the system is reported as the overall PLR.

A more typical approach of ours is to leverage the RdTools software (version 3.0.0a4 for this paper) to report on inverter-level degradation rate. Specific details are given in (Jordan, 2018) and (Jordan, 2022). Following filtering of high-frequency system data, time-series data is corrected by temperature and incident irradiance, then rolled up into daily normalized energy production for each inverter passing data QA. Daily production in subsequent years is compared for each calendar day to collect YOY slopes of performance change. The median of the distribution of these slope values is taken as the overall linear degradation rate for that inverter channel. Some specific details on the filtering approach taken in this analysis are as follows—the RdTools default values are used, which filter for data outages, high DC-AC clipping [Perry 2021], and low irradiance. Also, NSRDB satellite-based irradiance is used for all analysis rather than site-reported irradiance. This is to avoid calibration or bias errors for incorrect resource data.

An additional outlier filter is applied to the daily aggregated data. This two-way filter requires each datapoint to be within 3% of both a forward-looking and backward-looking seven-day moving window.

A third degradation approach considered here is the linear degradation trend extracted from the CODS soiling analysis. In addition to soiling loss estimation, CODS also provides for overall linear annual degradation on a system-level basis once seasonality and soiling trends are removed. This can be an interesting analysis to compare against because the effects of soiling are not explicitly removed from either the RdTools or PLR methods, and in many cases recoverable soiling will influence our analysis by showing up as faster-than-expected degradation.

# 3 System Availability

Availability of each PV system was calculated by looking for individual inverter outages, as described in Section 2.1. As we found before, there is a strong dependence of availability versus time mostly in the first six months of system performance. Once this initial start-up period concludes, average availability stays relatively constant. Figure 4 shows overall availability of all systems, binned by months since the beginning of the dataset. Note that this is not necessarily the commercial operation date, just the first datapoints received for this system by our partners. Results in Figure 4 show both the original 2020 results (right) and the updated 2023 results (left). Despite having almost 4x the monthly availability datapoints, the overall trends of the two analyses are very similar. After excluding the initial 6-month startup period, overall average availability remains around 0.98.

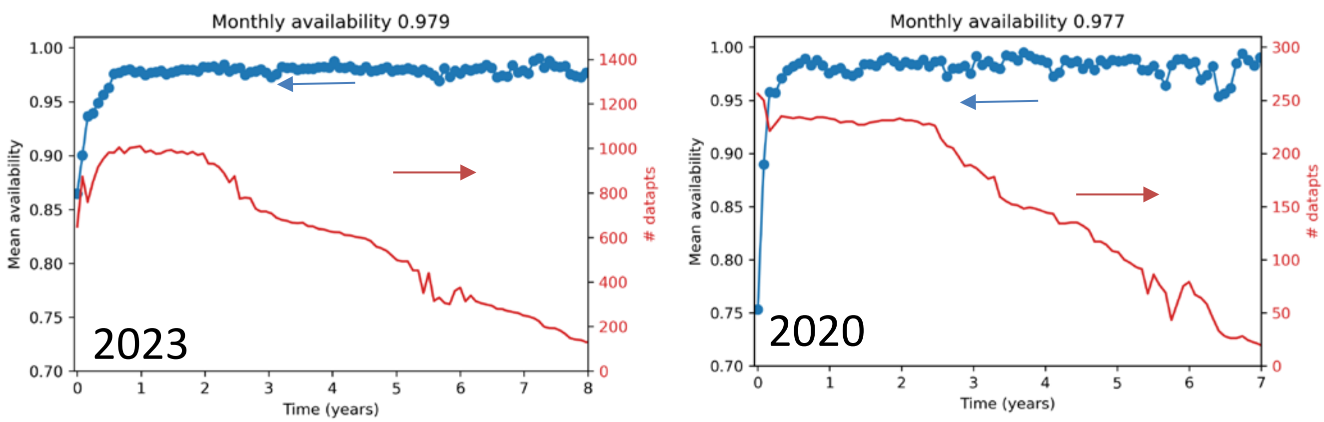


**Figure 4. Monthly mean availability of all systems grouped by time since start of dataset. Not grouped by system. Left plot shows present updated analysis. Right plot shows prior 2020 results.**

A slightly different view of the availability data gives more actionable results for performance expectations. Averaging over all datapoints like in the above plot gives outsized weight to those systems that may be experiencing particularly poor availability. A more useful way to present the results is to group availability by each system, and present average monthly availability for each system. In this case, we focus on steady-state availability excluding the initial six-month startup period. The histogram of mean monthly availability for each system is shown in Figure 5. Here the median (P50) value of 0.991 is shown by a dashed line at the 50th percentile. Other statistical values are 20th percentile (P80): 0.968 and 10th percentile (P90): 0.947.

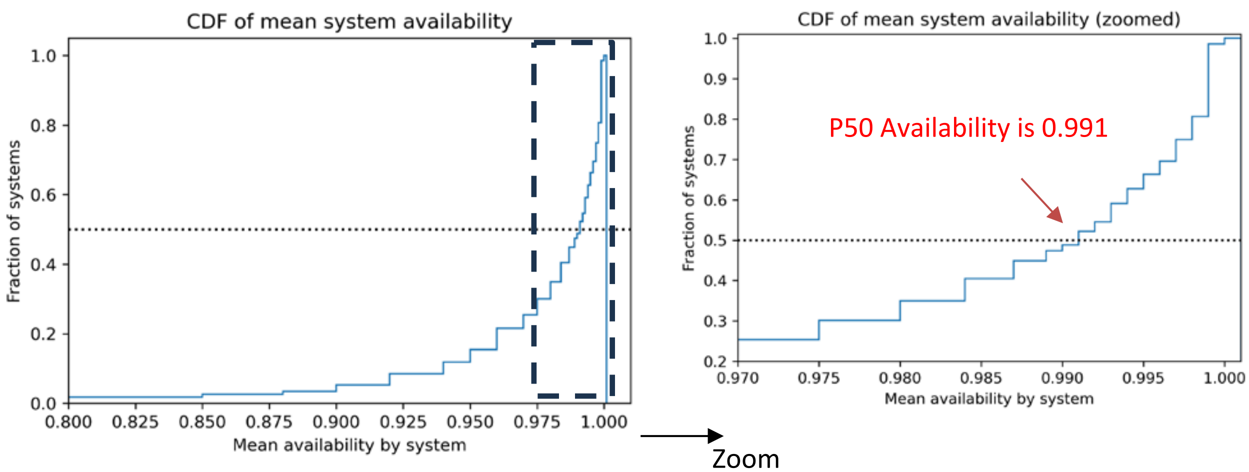


**Figure 5. CDF of availability by system is 0.99 at P50 level and 0.95 at P90 level. Right figure shows a zoomed-in plot around the P50 value.**

This figure illustrates that the median 50th percentile system demonstrates a system availability of 0.99. That is an improvement over the overall average availability of 0.979 shown in Figure 4. To further dig into what may influence system availability, we look at the above system availability data and further break it down by system size. We use 20 variable-width system power bins, each containing the same number of systems (N=56). The plot of Figure 6 shows the median (P50) and P90 system availability for each system power bin from 63kW to 13MW. As is clear from the figure, there is a negative correlation of availability vs system power, at both the P50 and P90 level.

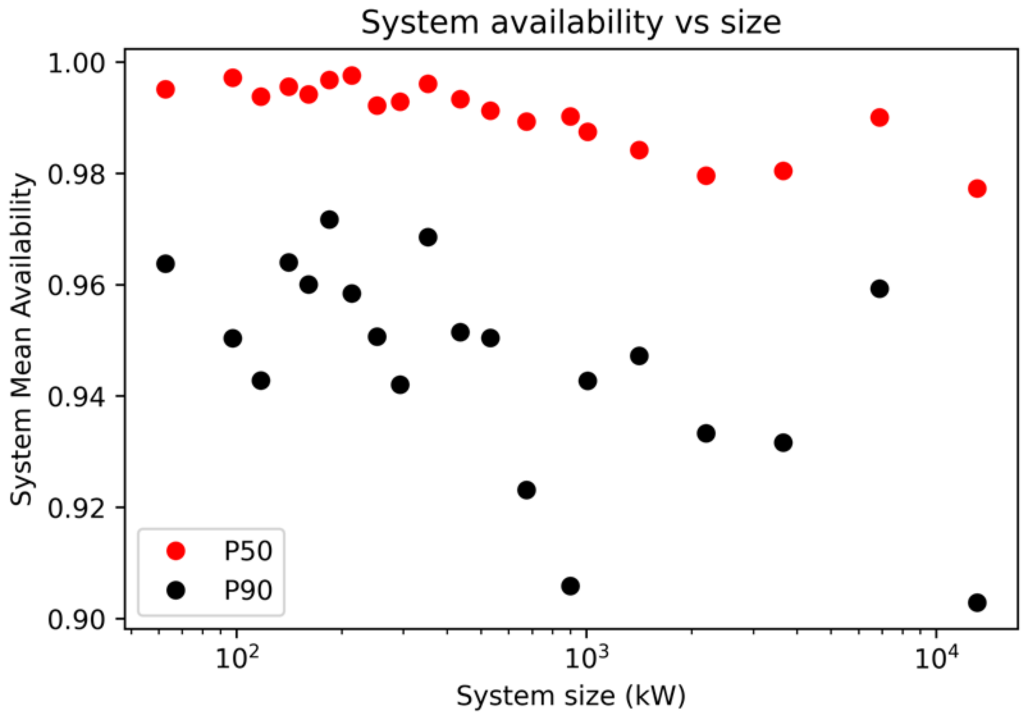


**Figure 6. System availability versus system size shows negative trend. P50 (red) and P90 (black) quantile values shown.**

Overall, by averaging systems above or below 1MW, we can say that the median system availability for small systems 0.02 MW- 1 MW is 0.994. For larger systems between 1MW and 30MW, median system availability is 0.984.

To dig into possible causes for the 1% difference in availability at larger system size, we investigate the impact of the PV inverter type on availability. While we can differentiate by individual inverter manufacturers, which will result in groupings of more or less reliable systems, it is hard to draw specific conclusions from that. What was helpful was to take the median and P90 system availability based on 20 equal-sized (N=56) bins of inverter size.[1] Individual inverter sizes range from 6kW to 4MW based on this estimate. A comparison of system availability vs inverter size is shown in Figure 7 at the P50 and P90 level. Focusing first on median (P50) values in red, there appears to be a grouping of higher availability systems between 0.99–1.0 for smaller inverters <250kW. Median availability for larger inverters 250kW–4MW is somewhat lower, from 0.97–0.99.

At the P90 level, the same trend does not hold as clearly. System availability at the 10th percentile is spread all over, from 0.9 to 0.98 for small inverters <250kW, while larger inverters range from 0.91–0.96. From this plot, we can draw some possible conclusions, one being that smaller string inverters can achieve higher system availability, possibly due to removal of string combiner boxes in systems. However, the

[1] Inverter AC capacities were estimated based on system DC rating, number of inverters per system and an assumed 1.2 DC/AC ratio.

large P90 spread indicates that there can still be manufacturer differences that result in variability of system availability. So, performance is not guaranteed.

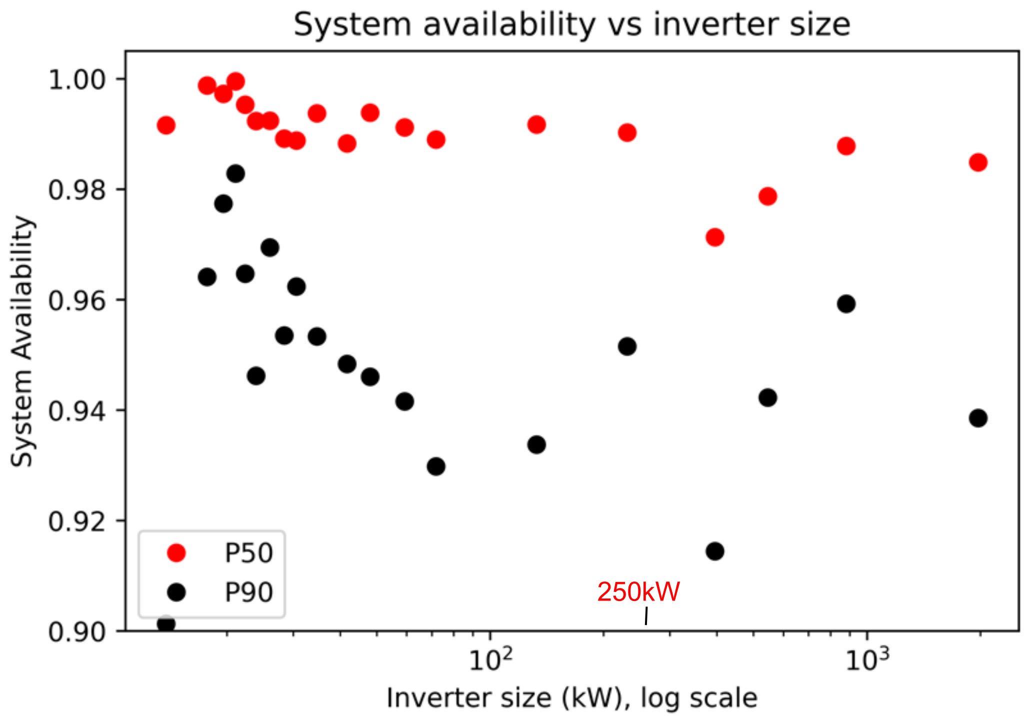


**Figure 7. System availability vs inverter size shows a negative correlation vs inverter size. Note that microinverters are not deployed on any of the systems in this study.**

In summary, inverter-level availability values that we have found in this study are higher than the PVWatts default value of 0.97 and closer to other published system availability values of 0.99. The values are also higher than our previously published average values of 0.977 due to a change in methodology—namely looking at the availability on a per-system basis rather than overall fleet availability for all inverters over their entire operating period. In all cases, these statistics exclude a low-availability six-month startup period.

# 4 Performance Index Results and Analysis

In this updated analysis of our 2020 work, we report on actual monthly system performance relative to our predicted model expectation based on measured NSRDB weather data. As before, raw PI was filtered to remove effects of snow, availability, and initial startup losses. These are removed because these effects can be accounted for in a proper back-casting where weather, availability, and soiling loss information is known. What remains is a tighter distribution of system performance, potentially including partial shading, non-recoverable soiling, degradation at faster than a -0.5%/year rate, and partial outages that are not caught by our availability algorithm.

In Figure 8 the post-filtered PI distribution is shown for all system-months in our fleet. In our previous 2020 analysis, 10,565 system-months made it into the distribution, versus 43,936 for our current analysis. This increase is due to additional systems being added to the PV Fleet, as well as improved data cleaning and metadata correction which excluded fewer systems. Looking at the statistical values of this distribution, the median performance index is 0.986 and the mean value is 0.96, reflecting the long negative tail. This is a decrease of around 4% from the previous 2020 analysis. Likewise, the P90 value came out lower at 0.85.

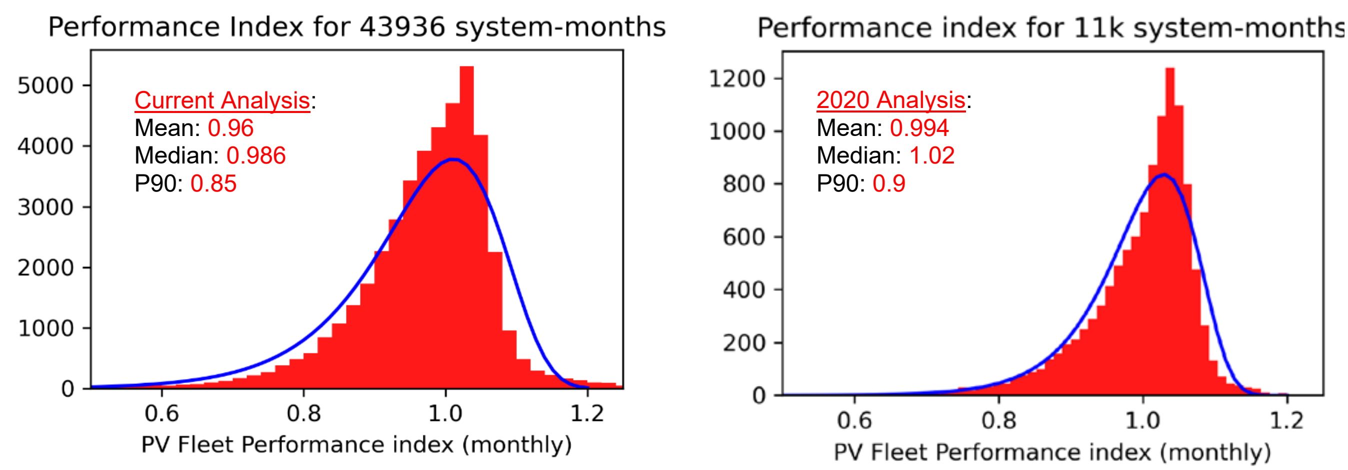


**Figure 8. Performance Index of the current fleet analysis (left) compared with the earlier 2020 report (right). PI accounts for availability loss and filters out initial four months of data and monthly snowfall >1cm. Gumbel extreme value fit shown in blue.**

As before, we can look at per-system statistics to gain information on the median (50th percentile) system and P90 (10th percentile) system. To accomplish this, we group the above overall histogram by system, and plot each system's total PI value in Figure 9. Total PI is the cumulative sum of actual monthly production over the sum of expected monthly production. This method weights monthly values by their relative importance on overall performance. The statistical outcome of this on a system basis: P50 system performance index is 0.95 and P90 system performance index is 0.83

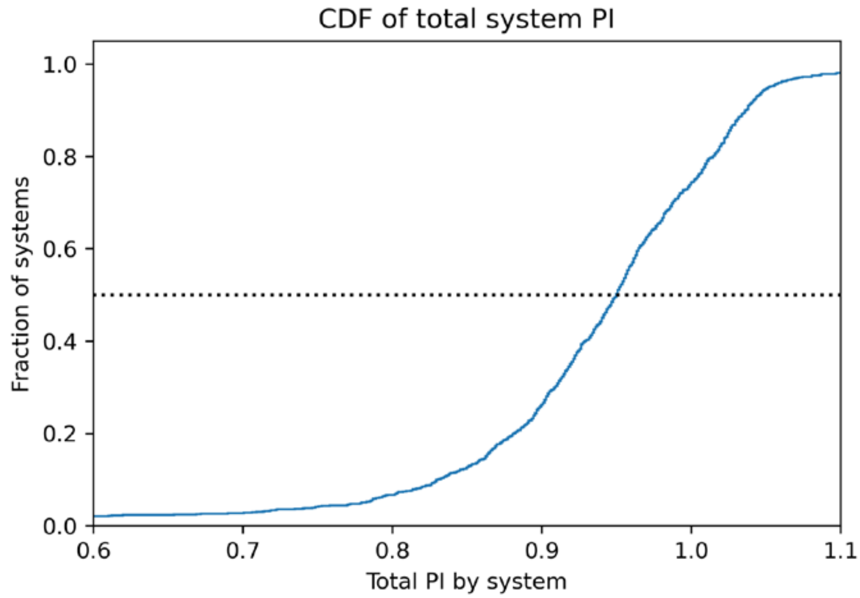


**Figure 9. Filtered Performance Index, summed by system. P50 median: 0.95. P75 value: 0.9. P90 value: 0.83.**

In this case, the median system underperforms expectation by around 5%, and 75% of systems perform within 10% of expectation. This is a somewhat worse outcome than we presented in 2020, which had a brighter result of 90% of systems performing within 10% of expected. Causes of this change in analysis outcome are being investigated.

# 5 Soiling Results and Analysis

As described in Section 2.3, soiling results and analysis are based on application of the CODS methodology included in RdTools. In this method, CODS uses an iterative approach to determine fits for soiling, degradation, seasonality, and residual noise. Following analysis, all results were manually reviewed to exclude inverter time-series data that had residual seasonality greater than 15% or visible problems with data outages or data shifts.

Using the described methods resulted in an update to the NREL soiling map with data for a total of 255 locations. Figure 10 provides the color-coded insolation weighted soiling ratio (IWSR) for the 255 sites. Sites with an IWSR<0.99 provide additional data on the annual IWSR for all years in the data set as well as monthly soiling loss statistics.

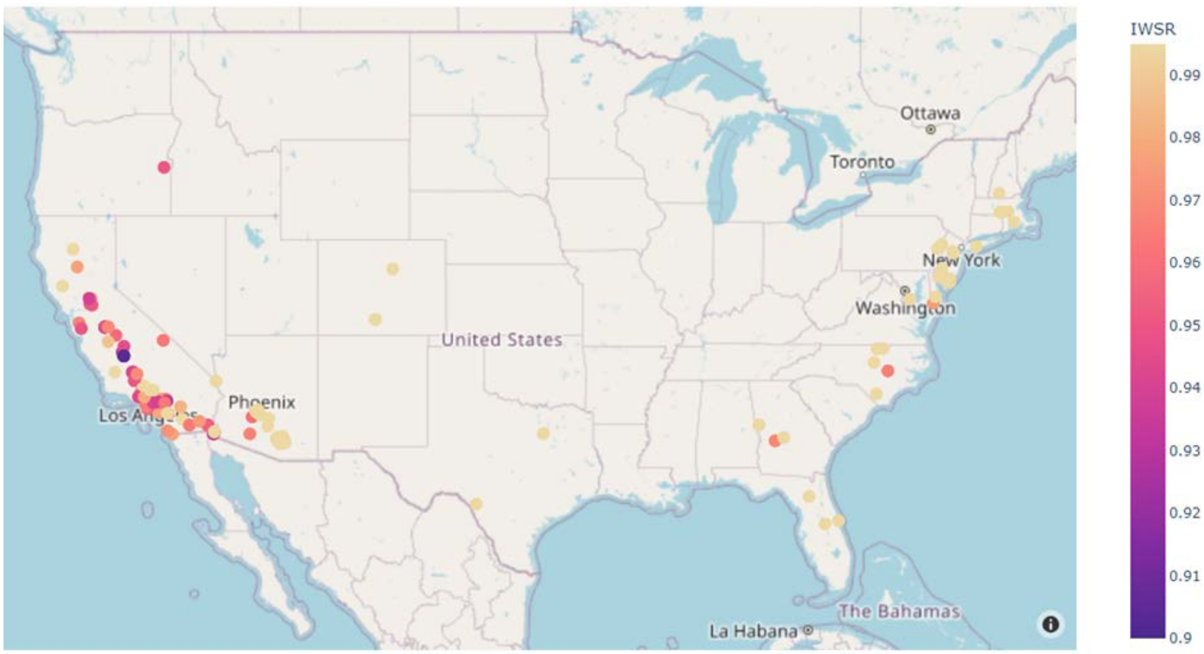


**Figure 10. Color-coded IWSR for the 255 sites on the updated NREL soiling map.**

Figure 11 provides a histogram of all the annual IWSR values for all the locations on the map that have an overall IWSR<0.99. Note that IWSR values greater than 0.99 are present in the histogram as some individual years have soiling losses less than 1% (IWSR>0.99). The histogram shows that annualized soiling losses typical range from 0–15% with the median annualized losses being in the range of 2–3%.

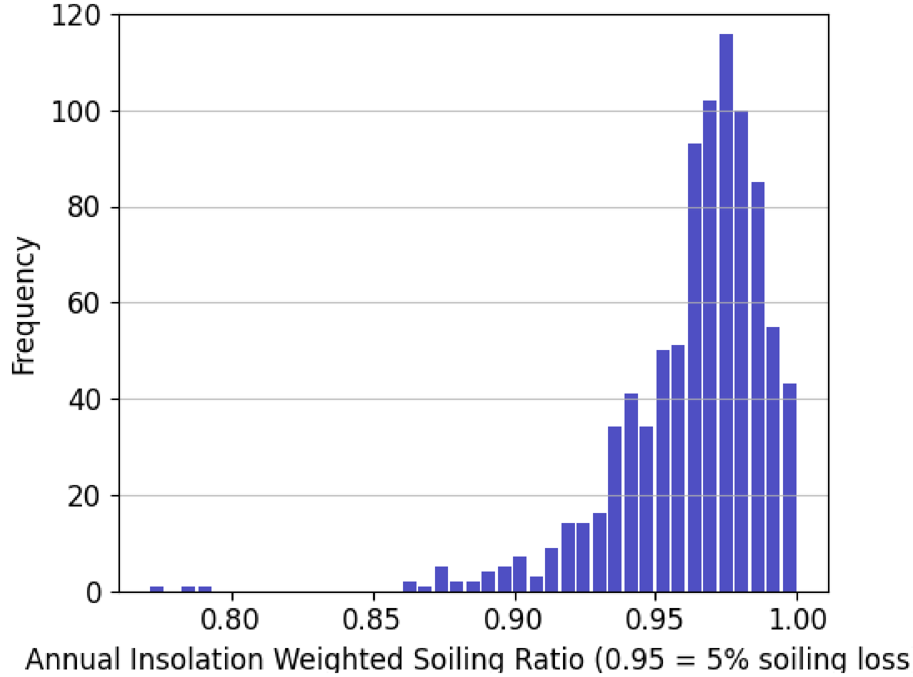


**Figure 11. Histogram of all the annual IWSRs from the sites on the soiling map that have an overall IWSR that is less than 0.99.**

# 6 System Degradation Results and Analysis

In 2022, an RdTools degradation analysis of the entire PV Fleet resulted in median system degradation rate of -0.75 %/year (Jordan 2022). An updated analysis was conducted here based on the ~10% expanded PV fleet. Overall trends were similar, but the addition of new data partners enabled a new way of visualizing the results. In Fig. 12, overall system degradation values are reported by individual data partner number using the regular YOY approach and the combined soiling and degradation (CODS) analysis. The reference loss rate of -0.75 %/year from the 2022 study is indicated by a horizontal dashed line. Most of the data partners show a median loss that is consistent with the previous finding; however, a minority of sub-fleets show substantially more rapid decline. For the data partners that exhibit rapid decline by the YOY method, the CODS analysis shows much less rapid decline, indicating that soling has a substantial impact on the performance of these partners. Other recoverable losses may still be present such as tracker outages, partial outages etc. even after removing soiling losses. It is also possible that for certain partners, the population used in this dataset is biased in favor of underperforming systems. Further analysis will be required to determine the specific reasons for each category, but certain populations are uniformly underperforming their peers.

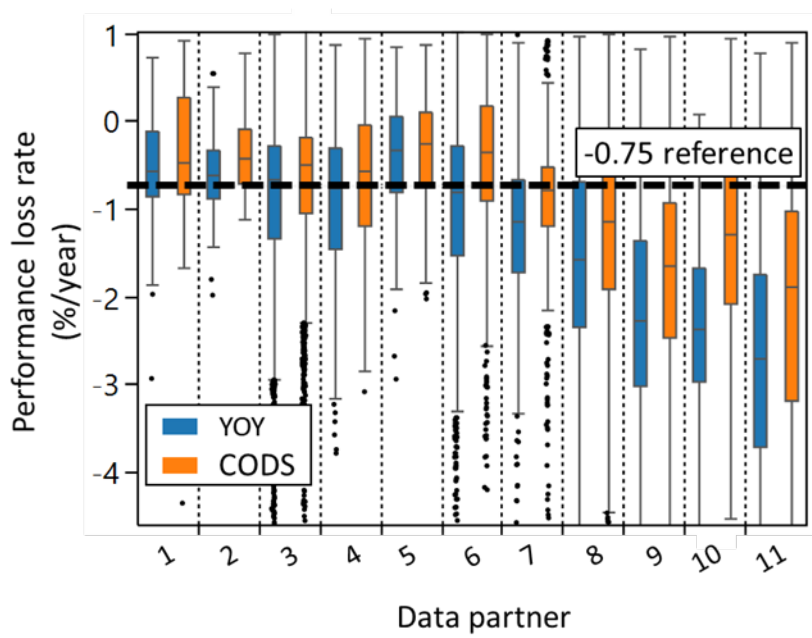


**Figure 12. RdTools degradation rate by data partner. The reference line of -0.75 %/year is the median degradation rate found in the previous analysis.**

In Section 2.4, we mentioned that multiple degradation measurement methods would be deployed on the same dataset. In this case, we compared the reported degradation rate from the standard RdTools analysis vs that extracted by CODS for each inverter in our dataset. To review, in addition to estimating soiling losses, CODS enables a soiling-aware estimate of system degradation rate. We have shown previously, with synthetic test data, that CODS provides a more accurate estimate of degradation rate in the presence of strong soiling trends [Skomedal 2020]. In Figure 13, we present a scatterplot of the RdTools YOY degradation versus CODS degradation for each inverter channel. For most inverters investigated, CODS results in a degradation rate that is less severe than the conventional RdTools analysis, a difference of approximately 0.25%/yr on average. Relative to the previous -0.75%/year degradation rate published, this represents a median system degradation value of -0.50%/year for the CODS results. Note that a direct comparison of overall system degradation rate results is not possible because the systems included in soiling analysis was just a subset of the total. But it does tend to suggest that soiling losses may contribute to apparent system degradation rates, and if not accounted for, the resulting YOY degradation values may be biased towards more rapid loss rates.

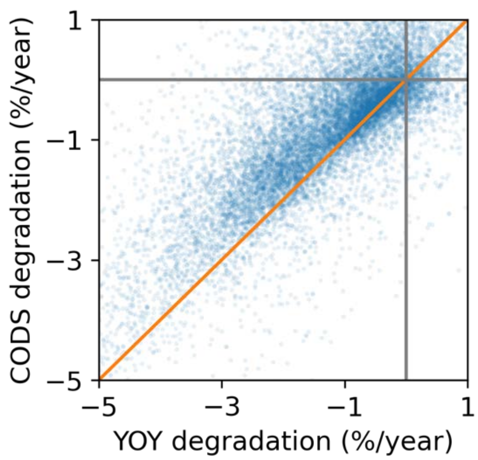


**Figure 13. Comparison between the degradation rate extracted by CODS and the RdTools YOY degradation rate. For many inverters, CODS estimates a slower degradation than the YOY suggesting that soiling trends may be biasing the YOY analysis.**

Our third way to assess system degradation is by examining the change in distribution of annual performance index over time. This conforms most closely to the definition of PLR and is shown for a subset of systems in Figure 13. The systems selected were those with at least six years of valid PI results between 0.5 and 1.5. This down-selection limits the analysis to a consistent set of systems with reasonable results over the whole time period. The PI values in Figure 14 are normalized to their value in the 2$^{nd}$ year. Normalization helps to highlight temporal trends that would otherwise be lost in the model offset noise [Deceglie 2023]. Choice of year two instead of year one reduces scatter in the normalization step introduced from year one underperformance associated with variable first year availability as discussed in Section 3. The figure also shows P10 and P90 decile values instead of the typical quartile values.

The effects of low year one availability are reflected in the low year-one median shown in Figure 14. The figure also shows that over time, while the median value remains relatively constant, the lower decile (P90) value continues to decrease after the second year. This reflects a higher risk of low PI in some poorly performing systems as time progresses. This decline is in addition to the 0.5%/year loss that is baked into the expected energy in this analysis. The takeaway is that the median system has modest annual degradation, but a subset degrades more quickly.

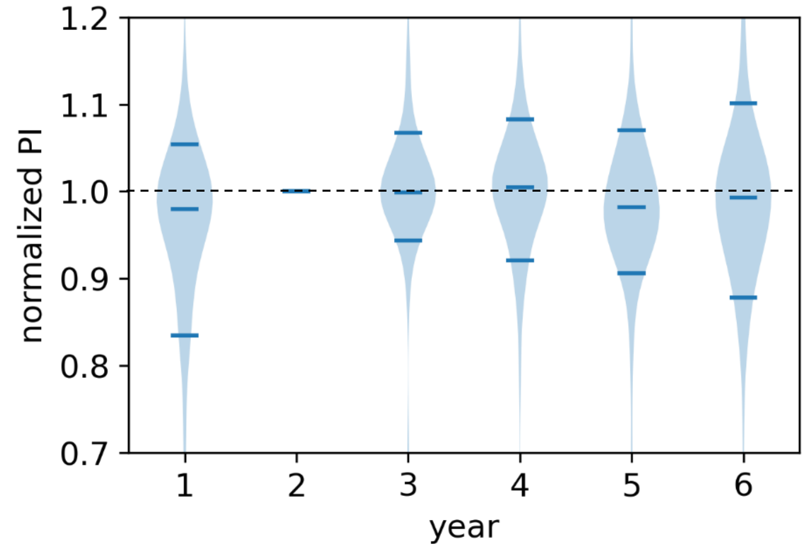


**Figure 14. Violin plot showing the evolution of PI distribution over time. PI values are normalized to year two to minimize the impacts of model offset and low year one availability. The horizonal lines indicate the median, P10 and P90 values of the distribution for each year. Black dashed line shows PI=1. Note: -0.5%/year annual degradation is factored into the expected energy model.**

# 7 Conclusions

In conclusion, our updated degradation analysis shows that multiple analysis methods provide relatively consistent results. Re-running RdTools on our updated fleet shows that overall median degradation remains approximately the same as from 2022, but some of the system data provided by data partners fall below the average value of -0.75%/year. Potential factors leading to this systemic under-performance are being investigated. A second degradation analysis method (CODS) was also compared with conventional RdTools analysis. Median system degradation using CODS was less severe by approximately 0.25%/year, primarily due to exclusion of soiling from the rolled-up system degradation analysis. This brings our degradation median value closer to the -0.5%/year value previously published for module DC degradation in [Jordan, 2012]. Our third degradation approach was to plot the change in annual PI over time (relative to Year-2 performance). This plot presupposes a -0.5%/year degradation rate in our calculation of expected energy. The median value shows flat PI over the first six years, but the 10$^{th}$ percentile value (P90) shows substantially faster degradation.

# References

Anderson, Kevin, and Mark Mikofski M. 2020. *Slope-Aware Backtracking for Single-Axis Trackers*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5K00-76626. https://www.nrel.gov/docs/fy20osti/76626.pdf

Anderson, Kevin, and Ryan Blumenthal. 2020. *Overcoming Communications Outages in Inverter Downtime Analysis*. Presented at the 47th IEEE Photovoltaic Specialists Conference (PVSC 47). https://doi.org/10.1109/PVSC45281.2020.9300635.

Bessa João Gabriel , Michael Valerino, Matthew Muller, Mike Bergin, Leonardo Micheli, Florencia Almonacid, and Eduardo Fernandez. 2024. "An Investigation on the Pollen-Induced Soiling Losses in Utility-Scale PV Plants." *IEEE Journal of Photovoltaics* 14 (1): 178–184. https://doi.org/10.1109/JPHOTOV.2023.3326560.

Deceglie, Michael G., Leonardo Micheli, and Matthew Muller . 2018. "Quantifying Soiling Loss Directly from PV Yield," *IEEE Journal of Photovoltaics* 8 (2): 547–551. https://doi.org/10.1109/JPHOTOV.2017.2784682.

Deceglie, Michael G., Kevin Anderson, Daniel Fregosi, William B. Hobbs, Mark Mikofski, Marios Theristis, and Bennet Meyers. 2023. "Perspective: Performance Loss Rate in Photovoltaic Systems" *Solar Rapid Research Letters*. https://doi.org/10.1002/solr.202300196.

Deline, Chris, Matt Muller, Michae 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.

Deline, Chris, Kevin Anderson, Dirk Jordan, Andy Walker, Jal Desai, Kirsten Perry, Matt Muller, Bill Marion, and Robert White. 2021. "PV Fleet Performance Data Initiative: Performance Index-Based Analysis." Golden, CO: National Renewable Energy Laboratory. NREL/TP-5K00-78720. https://www.nrel.gov/docs/fy21osti/78720.pdf.

French Roger H., Laura S. Bruckman, David Moser, Sascha Lindig, Mike van Iseghem, Björn Müller, Joshua Stein, Mauricio Richter, Magnus Herz, Wilfried van Sark, and Franz Baumgartner. "Assessment of Performance Loss Rate of PV Power Systems." Report IEA-PVPS T13-22:2021. https://iea-pvps.org/wp-content/uploads/2021/04/IEA-PVPS-T13-22_2021-Assessment-of-Performance-Loss-Rate-of-PV-Power-Systems-report.pdf/.

Fuentes, MK. 1987. "A Simplified Thermal Model for Flat-Plate Photovoltaic Arrays." Albuquerque, NM: Sandia National Laboratories,. SAND-85-0330. http://prod.sandia.gov/techlib/access-control.cgi/1985/850330.pdf

Haegel, Nancy M. et al. 2023. "Photovoltaics at multi-terawatt scale: Waiting is not an option" *Science*. 380 (6640): 39–42. https://doi.org/10.1126/science.adf6957.

Holmgren, William F., Clifford W. Hansen, and Mark A. Mikofski. 2018. "pvlib python: a python package for modeling solar energy systems." *Journal of Open Source Software*, 3 (29), 884. https://doi.org/10.21105/joss.00884.

Holmgren William F., et al. 2021. "pvlib/pvlib-python: v0.9.0 (Version v0.9.0)." Zenodo. http://doi.org/10.5281/zenodo.5366883.

Jordan Dirk C., John J. Wohlgemuth, and Sarah R. Kurtz. 2012. "Technology and climate trends in PV module degradation." Presented at the 27th European Photovoltaic Solar Energy Conference; Frankfurt, Germany. Golden, CO: National Renwable Energy Laboratory. https://www.nrel.gov/docs/fy13osti/56485.pdf.

Jordan Dirk C., Chris Deline, Sarah R. Kurtz, Gregory M. Kimball, Mike Anderson. 2017. "Robust PV Degradation methodology and application." *IEEE Journal of Photovoltaics*, 8 (2): 525–531. https://doi.org/10.1109/JPHOTOV.2017.2779779.

Jordan, Dirk C., Kevin Anderson, Kirsten Perry, Matthew Muller, Michael Deceglie, Robert White, and Chris Deline. 2022. "Photovoltaic fleet degradation insights." *Progress in Photovoltaics*, 30 (10), 1166–1175. https://doi.org/10.1002/pip.3566.

Lazard Ltd. 2023. "Lazard's Levelized Cost of Energy Analysis (LCOE 16.0)." https://www.lazard.com/media/2ozoovyg/lazards-lcoeplus-april-2023.pdf.

Lindig, Sascha, Marios Theristis, David Moser. 2022. "Best practices for photovoltaic performance loss rate calculations." Progress in Energy, 4 (2): 022003. https://doi.org/ 10.1088/2516-1083/ac655f.

Lorenzo, E., L. Navarte, and J. Muñoz . 2011. "Tracking and back-tracking." *Progress in Photovoltaics*: Research and Applications, 19, 747–753. https://doi.org/10.1002/pip.1085.

U.S. Department of Energy. 2021. "The Solar Futures Study". Golden, CO: National Renewable Energy Laboratory. https://www.osti.gov/servlets/purl/1820105.

Muller, Matthew, Kevin Anderson, and Michael Deceglie. 2023. "Generating Synthetic Time Series Photovoltaic Data with Real-World Physical Challenges and Noise for Use in Algorithm Test and Validation." Golden, CO: National Renewable Energy Laboratory. NREL/TP-5K00-86459. https://www.nrel.gov/docs/fy23osti/86459.pdf.

Perez, Richard,, Pierre Ineichen, Robert Seals, Joseph Michalsky, and Ronald Stewart. 1990. "Modeling daylight availability and irradiance components from direct and global irradiance." *Solar Energy*, 44 (5), 271–289. https://doi.org/10.1016/0038-092X(90)90055-H.

Perry, Kirsten, Matthew Muller, and Kevin Anderson. 2021. "Performance Comparison of Clipping Detection Techniques in AC Power Time Series." Presented at the IEEE Photovoltaic Specialists Conference. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10360063.

Perry, Kirsten, William Vining, Kevin Anderson, Matthew Muller, and Clifford Hansen. 2022. "PVAnalytics: A Python Package for Automated Processing of Solar Time Series Data." 2022 PV

Performance Modeling and Monitoring Workshop. Presented at the PV Performance Modeling and Monitoring Workshop, Salt Lake City, Utah. nrel.gov/docs/fy22osti/83824.pdf.

Perry Kirsten, and Matthew Muller. 2022. "Automated Shift Detection in Sensor-Based PV Power and Irradiance Time Series." Presented at the IEEE Photovoltaic Specialists Conference. https://doi.org/10.1109/PVSC48317.2022.9938675.

Perry, Kirsten, Bennet Meyers, and Matthew Muller. 2023. "Survey of Time Shift Detection Algorithms for Measured PV Data." Presented at the 2023 Photovoltaic Reliabilty Workshop. National Renewable Energy Laboratory, Golden, CO, NREL/PO-5K00-85699. https://www.nrel.gov/docs/fy23osti/85699.pdf.

Perry, Kirsten, Bennet Meyers, Kevin Anderson, and Matthew Muller. 2023. "A Reproducible Validation of Algorithms for Estimating Array Tilt and Azimuth from Photovoltaic Power Time Series". Presented at the IEEE Photovoltaic Specialists Conference. National Renewable Energy Laboratory, Golden, CO, NREL/PR-5K00-86513. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10360063.

PVAnalytics, version 0.1.3, https://github.com/pvlib/pvanalytics.

RdTools, version 3.0.0a4, https://github.com/NREL/rdtools, DOI:10.5281/zenodo.4300113.

Reda, Ibrahim, and Afshin Andreas. 2008. "Solar position algorithm for solar radiation applications." *Solar Energy*, 76(5), 577–589. https://doi.org/10.1016/j.solener.2003.12.003.

Sengupta, Manajit, Yu Xie, Anthony Lopez, Aron Habte, Galen Maclaurin, and James Shelby. 2018. The National Solar Radiation Data Base (NSRDB). *Renewable and Sustainable Energy Reviews*, 89, 51–60. https://doi.org/10.1016/j.rser.2018.03.003.

Skomedal, Åsmund, and Michael Deceglie. 2020. "Combined Estimation of Degradation and Soiling Losses in Photovoltaic Systems," *IEEE Journal of Photovoltaics*, 10(6), 1788–1796. https://doi.org/10.1109/JPHOTOV.2020.3018219.