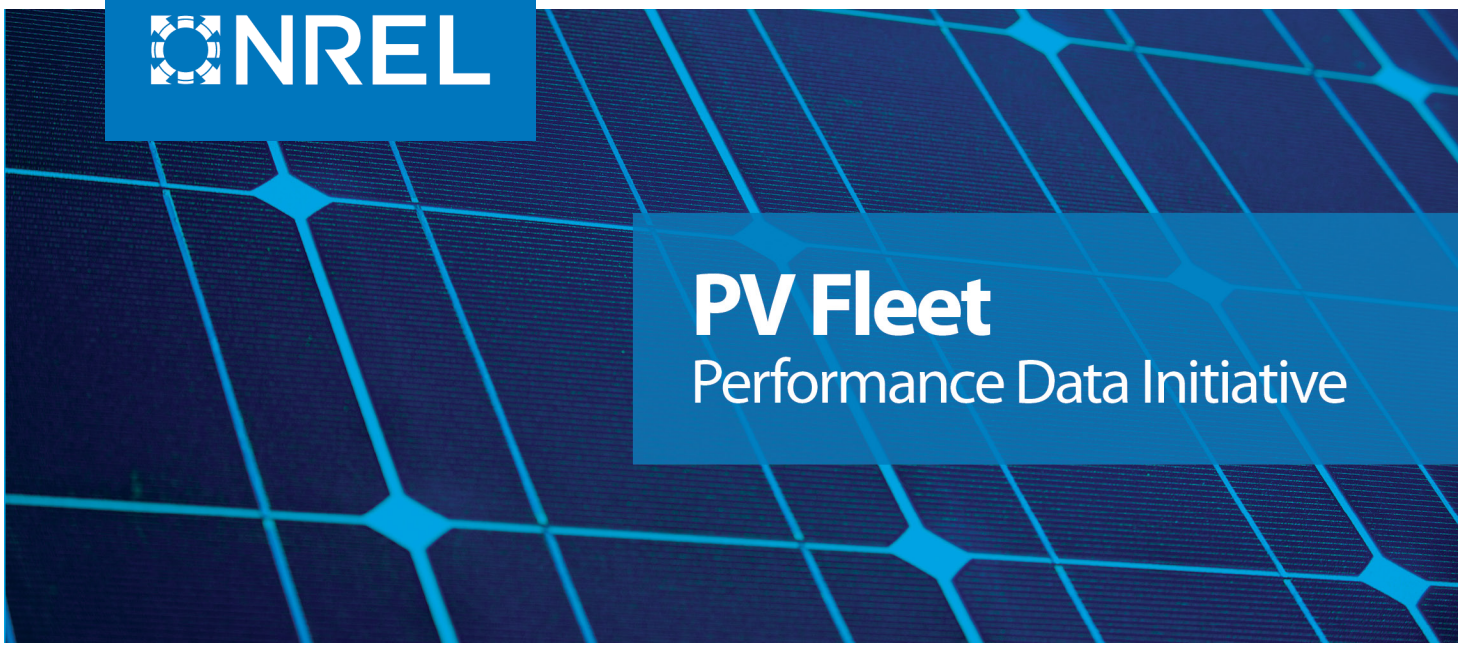

PV Fleet Performance Data Initiative: Performance Index–Based Analysis

Chris Deline, Kevin Anderson, Dirk Jordan, Andy Walker, Jal Desai, Kirsten Perry, Matt Muller, Bill Marion, 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-5K00-78720
February 2021

PV Fleet Performance Data Initiative: Performance Index–Based Analysis

Chris Deline, Kevin Anderson, Dirk Jordan, Andy Walker, Jal Desai, Kirsten Perry, Matt Muller, Bill Marion, and Robert White

National Renewable Energy Laboratory

Suggested Citation
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.

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-5K00-78720
February 2021

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. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office (SETO) under Agreement Numbers 34348 and 34172. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Acknowledgments

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

List of Acronyms and Abbreviations

AC	alternating current
DC	direct current
DHI	diffuse horizontal irradiance
DNI	direct normal irradiance
DOE	Department of Energy
E	energy
G	irradiance
GHI	global horizontal irradiance
HSAT	horizontal single-axis tracker
IEC	International Electrotechnical Commission
kW	kilowatt
L	lost energy
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
POA	plane-of-array
PR	performance ratio
PV	photovoltaic
PVLIB	PV energy modeling software
PVWatts	PV production software
STC	standard test condition
T_{cell}	cell temperature
γ	module temperature coefficient

Executive Summary

The PV Fleet Data Initiative is a DOE-sponsored program to partner with photovoltaic (PV) fleet owners to collect high-quality PV production data, which are aggregated, anonymized, and analyzed. We then report on performance trends for the PV fleet as a whole. This technical report is focused specifically on one aspect of performance—monthly alternating current (AC) energy normalized by a PVWatts expected model value, or Performance Index (PI).

In this analysis, we report on 250 PV systems throughout the United States, comprising 157 megawatts (MW) direct current of system capacity and more than 10,000 monthly PI values based on high-frequency (subhourly) energy data and satellite-based resource data. The distribution of PI values is analyzed, and multiple causes of underperformance are assessed, including first-year start-up issues, snowfall, and inverter downtime.

An initial distribution of raw monthly PI values was collected with mean measured/modeled performance of PI = 0.935. After correcting for the three identifiable loss factors mentioned above, an average monthly performance of PI = 0.994 was obtained, with a distribution closely following a Gumbel Extreme Value distribution (Figure ES-1). In particular, inverter availability was found to contribute a system energy loss of 2.3% on average, except in the first six months of operation when availability losses are closer to 8%. Other start-up issues beyond inverter downtime, such as partial string outage, contributed an additional 5% underperformance in the first year of operation across the fleet.

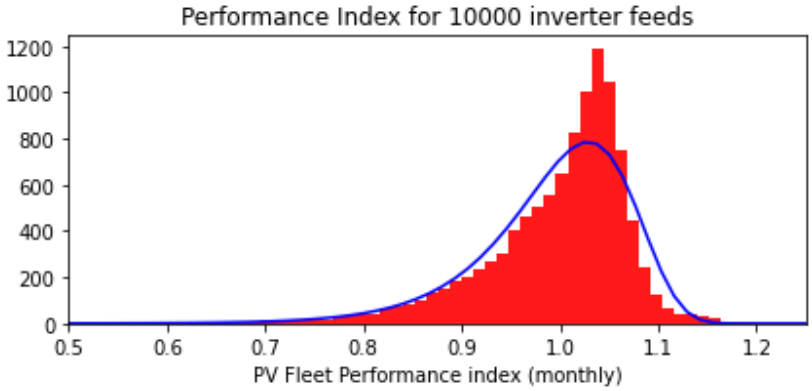

Figure ES-1. Monthly fleet-level PI distribution using high-frequency (e.g., hourly) AC data, adjusted for availability and filtered to exclude first-year start-up and snow loss. Model comparison based on PVWatts performance model and NSRDB satellite irradiance. Mean PI = 0.994 with Gumbel extreme value fit overlay in blue.

Winter performance was also found to be 5%–10% below summer performance on average, likely due to snowfall. A simple linear relationship was found between snow loss and monthly snowfall accumulation in centimeters, indicating a 6%–40% loss in months with nonzero snowfall, depending on climate.

Additional model errors and sources of uncertainty are investigated by comparing residual influence of calculated PI on ambient weather conditions. Over the range of wind speed and ambient temperature covered by the majority of the monthly measured points, no specific trend

was apparent. Uncertainty of the method may be improved slightly by using a more detailed thermal model that accounts for specific module technology and mounting conditions; however, this was outside the scope of this initial effort.

This paper provides PV plant owners and operators with a methodology to benchmark their plant/fleet performance while also providing average performance details to guide appropriate values for use in pro-forma performance and financial models, and expectations for operation of future PV plants.

Table of Contents

1 Introduction

Accurate characterization of photovoltaic (PV) system performance is a key driver of PV system financing and deployment. Because the financial risk of developing new PV systems is tied to the uncertainty of lifetime performance projections, reducing that uncertainty carries significant financial value. Although substantial progress has been made to characterize the laboratory performance of the various components comprising a PV system, lab-based studies can only ever approximate the operating conditions seen by fielded PV systems. As PV deployment has grown over time, so has the opportunity presented by bulk analysis of the time series monitoring data recorded by the systems. By analyzing the production data and system metadata of fielded systems at fleet scale, trends of performance across system design and climate can be extracted.

In this paper we examine the performance of fielded systems collected in the PV Fleet Performance Data Initiative [Deline 2020], an ongoing analysis effort led by the National Renewable Energy Laboratory (NREL) and the U.S. Department of Energy (DOE). This 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, and will not be shared with any person or group not authorized by the data owner.

In this publication we are specifically looking at normalized monthly performance, with more than 250 systems represented, primarily comprising nameplate capacities between 100 kilowatts direct current (kWdc) and 10 megawatts direct current (MWdc). Raw 15-minute performance data are normalized by expected (modeled) production data based on satellite meteorological data and a PVWatts-based performance model. This results in a distribution of performance data over the PV fleet, aggregated monthly and corrected for climate conditions such as temperature and irradiance. Trend analysis, based on seasonality and system configuration details, provides some insight into PV system performance over time and under a variety of configurations and operating conditions.

2 Methodology

2.1 Performance Index

PV system performance is often quantified as a comparison between realized system energy production and an expected or modeled energy production. A popular metric in industry is the performance ratio (PR), which normalizes a system's energy production by system capacity and received plane-of-array (POA) insolation over some interval (here always considered on a monthly basis):

$$PR = \frac{\sum_i E_{AC,i}}{\sum_i \left[P_{STC} \left(\frac{G_{POA,i}}{G_{STC}} \right) \right]} \quad (1)$$

In addition to controlling for the effect of irradiance on production, PR may also be defined with a secondary adjustment, as in Dierauf (2013), to control for the expected effect of PV cell temperature on PV conversion efficiency:

$$PR_{corr} = \frac{\sum_i E_{AC,i}}{\sum_i \left[P_{STC} \left(\frac{G_{POA,i}}{G_{STC}} \right) \left(1 + \gamma \left(T_{cell,i} - T_{cell,ref} \right) \right) \right]} \quad (2)$$

The denominators of these equations can be viewed as simple PV performance models, making the overall metric the ratio of actual to expected production. In that vein, an improved PV performance model with reduced prediction error would lead to a more accurate performance ratio. For example, the PVWatts model (Dobos 2014) extends the denominator of Equation 2 by including the effect of module reflectance loss, inverter saturation (clipping), wiring loss, and other loss mechanisms that reduce the final power delivered at the system output.

When normalizing system production by an expected production from a more detailed model like PVWatts, the metric is called a performance index (PI) rather than a performance ratio (IEC 61724-1; Townsend 1994; Mokri 2014, among others). To the extent that the expected production model accurately captures system behavior under normal operation, a negative deviation from unity is easily interpreted as a production loss. However, the choice of inputs to the expected production model can be chosen strategically to include the effect of some losses and ignore others. For instance, a PV system operator might only be interested in the magnitude of losses that can be controlled with preventative or reactive maintenance ("controllable losses"), whereas uncontrollable loss sources like utility curtailment could be included in the expected energy calculation so that they cancel out when calculating the PI.

Aside from the specifics of the expected energy model, the details of the weather data set must be considered. High-quality irradiance and temperature data recorded from on-site sensors are preferred when available, but sensor calibration and data outages make automated analysis difficult. To get around these issues, we use satellite-based weather data from the National Solar Radiation Database (NSRDB) [Sengupta 2018]. The NSRDB data set is serially complete, is available at high (4-kilometer) spatial resolution across the United States, has a 30-minute time resolution, and avoids the complications of imperfect instrument maintenance.

The expected energy implementation used to calculate PI in this analysis is based strongly on PVWatts. Most of the modeling steps described here are implemented using the open-source PV modeling package PVLIB-python, version 0.6.3 (Holmgren 2018; Holmgren 2019). All performance indices reported here are calculated at system level from the revenue grade meter.

2.1.1 System Operating Conditions

The time series operating conditions for each system are calculated 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 the approach described in (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 true value is not provided with our system metadata.
4) POA irradiance is modeled from the NSRDB irradiance components. Diffuse horizontal irradiance (DHI) 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 (NOCT) of 45°C. This value is consistent with the assumption of a polymer-backsheet PV module on an open rack.

2.1.2 Expected Production

Continuing to follow the PVWatts model, the base DC array output power at timestamp *i* is calculated with:

$$P_{exp,i} = P_{STC}\left(\frac{G_{POA,i}}{G_{STC}}\right)\left[1 + \gamma\left(T_{cell,i} - T_{cell,ref}\right)\right] \quad (3)$$

Here γ is the module temperature coefficient assumed to be −0.47%/°C for silicon PV systems under study in this work. P_{STC} is the system's DC rated power as reported in metadata from the site owners. This base power is then reduced according to several system loss components. As mentioned earlier, 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%	0%
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% (0.5%/yr)
Availability	3%	0%
Total	**14.1%**	**8.23%**

Note: Values in red highlight differences in default values.

In contrast to PVWatts, which assumes no loss with system age, we assume a 0.5%/yr capacity degradation that accumulates across the system data set, and we do not include it in the static total loss value. Note that the total loss L_{total} is not calculated as the sum of constituent losses L_j but rather through the product formula:

$$L_{total} = 1 - \prod_j (1 - L_j) \tag{4}$$

By omitting the soiling and availability losses from the expected energy model, any realized production loss from those sources will show up as underperformance when measured by the PI.

After applying these static losses to the base array output power, the PVWatts inverter model is applied, using the default reference efficiency of 96.37% and a nominal efficiency of 96%. Finally, the power is curtailed to the system AC power limit to simulate the effect of inverter saturation (clipping). Because system AC power limit is not present in the available system metadata, it is calculated from the available time series data as the 99.9th percentile of the subset of system power where the power is greater than the 1st percentile of all power data.

Following collection of measured and modeled system power for each timestamp *i*, monthly PI is found by dividing monthly integrated measured power by modeled power, as per Eq. (1).

2.2 System Availability

Unlike the majority of the losses in Table 1, availability tends to be a severe but sporadic loss mechanism. Availability can be evaluated at the system level, the inverter level, or even as granular a level as individual subarrays or strings. It has been common in industry applications to define

availability in terms of equipment uptime. In this context, we define system availability as an energy-weighted metric reflecting the production loss from inverter outages:

$$Availability = 1 - \frac{E_{outages}}{E_{outages} + E_{actual}} \quad (5)$$

The challenge lies in estimating the lost production $E_{outages}$ associated with inverter outages, paying particular attention to distinguish true inverter outages from nuisance, but ultimately benign, communication interruptions. To perform this estimate, we use the method proposed 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; i.e., partial capacity reductions from thermal throttling 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. For completeness, a short overview of the approach's two branches is as follows:

2.2.1 Subsystem Outages

The first branch 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.

2.2.2 System Outages

The second branch targets "complete" outages where the entire system shows no production. In contrast with the first branch, this branch 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.

2.2.3 Example Outage Classification

Figure 1 illustrates various combinations of the two outage types described above.

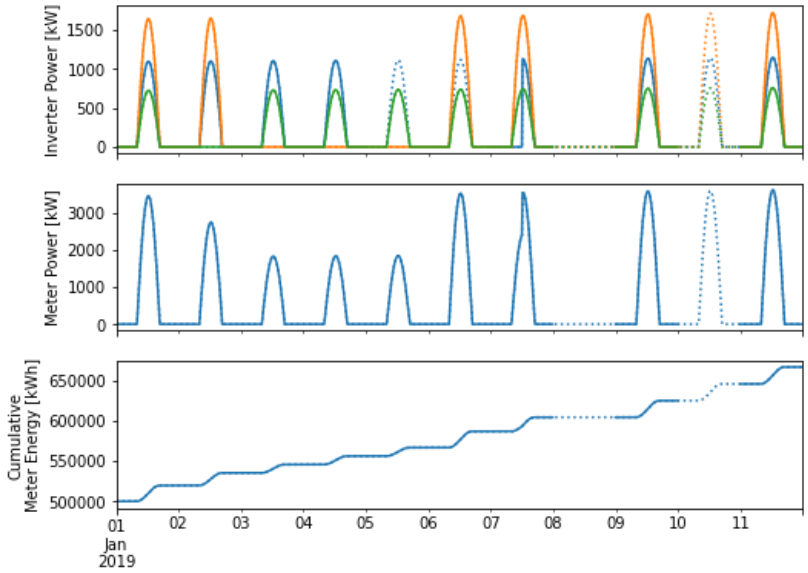

Figure 1. Example data for availability analysis. Dotted lines show "true" data that would not normally be available to the analysis routine but are shown here for illustration.

For example, day 5 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" branch because at least one peer's power data is available for comparison. In contrast, days 8 and 10 show a real systemwide outage and a full communication interruption respectively and would be handled by the "system outage" branch.

3 Data Partner Fleet Summary

3.1 Fleet Metadata Details

Performance data from the data partners are collected from of more than 1,000 systems spread across the country (Figure 2). Histograms of system sizes and types of PV indicate a variety of sizes (Figure 3, Figure 4), with median system size of 636 kW DC and 4.5 years of data on average. Following data quality checks for duration and frequency of data, gaps, and data shifts, more than 250 high-quality systems remain. Unfortunately, all nonsilicon module technologies are excluded from the analysis due to insufficient data quality.

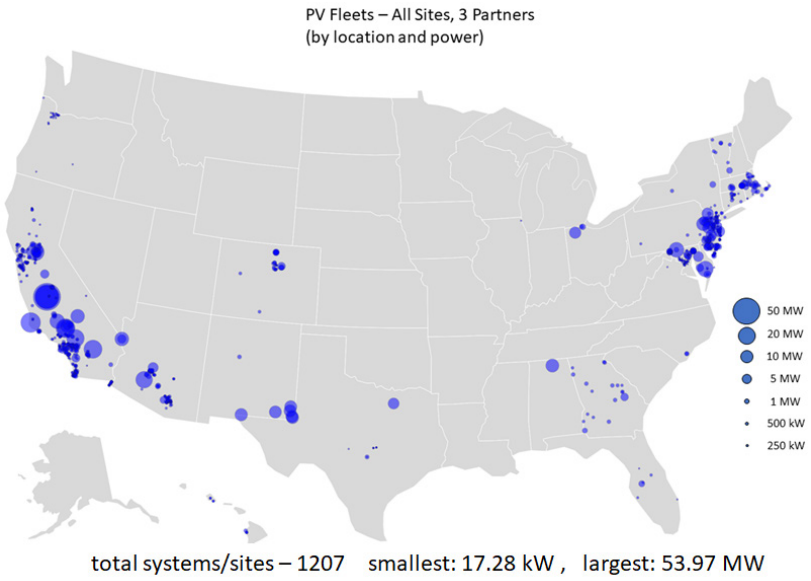

Figure 2. Spatial map of all data partner systems (pre quality assessment)

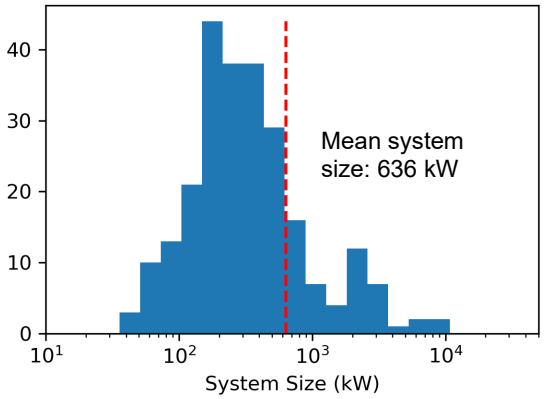

Figure 3. Distribution of system DC nameplate rating (post-data QA)

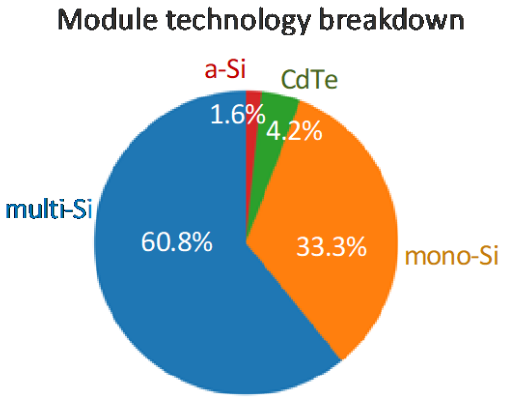

Figure 4. Module technology breakdown (pre-data QA)

Time-series data are collected for both inverter-level and revenue-grade meter AC data, typically on a 15-minute time interval. All performance indices are reported at the system level; the inverter data are used for availability analysis as described in Section 2.2. 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.

For each unique location, a scalar PV climate stressor value is generated, based on the method of Karin (2019). This value combines the annual temperature and humidity of the zone to generate a PV-specific score reflecting the relative stress of field deployment based on known weathering and degradation phenomena (Figure 5). Additional information about our methodology including data quality checks, data warehouse information, and other processes are documented in Deline (2020).

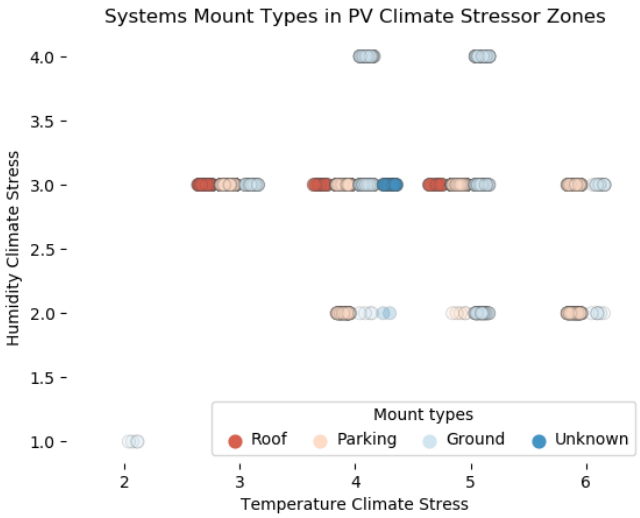

Figure 5. PV Climate stressors and PV mounting configurations for each system. The higher the climate zone numbers for temperature and humidity, the greater the possible stress on the system modules. Temperature climate numbers are for ground-mounted systems. Roof-mounted installations will tend to have even higher stress than ground-mounted systems due to the more restricted air flow (Karin, 2019).

4 Performance Index Results and Analysis

More than 13,000 separate monthly performance index values are collected for the 250 high-data-quality PV systems. The histogram of PI values is shown in Figure 6, with a skewed distribution centered near 1, but with a long negative tail extending down to zero. Several monthly instances greater than 1 are also shown, but these are of decreasing frequency for values greater than 1.1. Standard statistic values show a mean of the distribution at 0.935, and a 1σ standard deviation of 0.18.

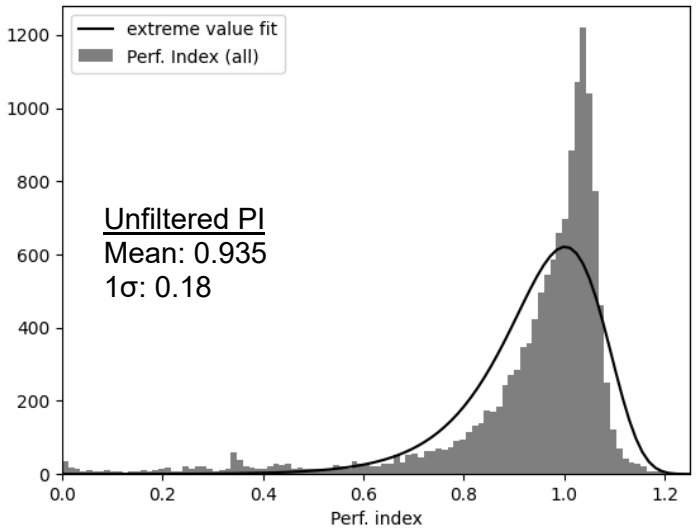

Figure 6. Unfiltered distribution of system-level monthly PI for 250 systems showing a mean value of 0.935 and a negative skewed distribution, approximated by an extreme value distribution fit.

The histogram distribution is clearly not normally distributed; nor does it follow a log-normal distribution. A better fit is achieved with a Gumbel extreme-value distribution, which is often used to

assess tail-risk scenarios (Jordan 2012). It is part of a class of distribution colloquially known as "weakest link" distributions, where multiple modes of failure or underperformance can lead to a tail of low-performing values:

$$P(x) = \frac{1}{\beta}\exp\left[\frac{x-\mu}{\beta} - \exp\left(\frac{x-\mu}{\beta}\right)\right] \quad (6)$$

where best-fit parameters for the raw distribution are μ= 1.0 and β =0.096. The best-fit line to the unfiltered monthly data is also shown in Figure 6. The fit is not perfect, and many values are shown underperforming at low PI value. In the next sections we investigate some of the causes of monthly underperformance in this cross-section of 250 systems and attempt to isolate them from a more fundamental underlying distribution, which will have a better fit. There is also a large grouping of system-months very near PI = 1.0 which may represent well-tended systems that perform better than the extreme value distribution would predict.

4.1 Inverter Availability

The first loss factor we investigate is the downtime of systems. System availability is discussed in Section 2.2, and we find it has a strong influence on monthly PI. It is evident that systems with an availability of only 75% in a given month should therefore have PI that is only 75% of the expected value. A plot of monthly PI versus monthly availability is shown in Figure 7, and it indicates that for a number of systems, there is a 1:1 correlation between PI and availability. We can therefore apply a correction factor to expected monthly energy and multiply by measured monthly availability, which should remove some of the impact of availability loss to the calculated PI.

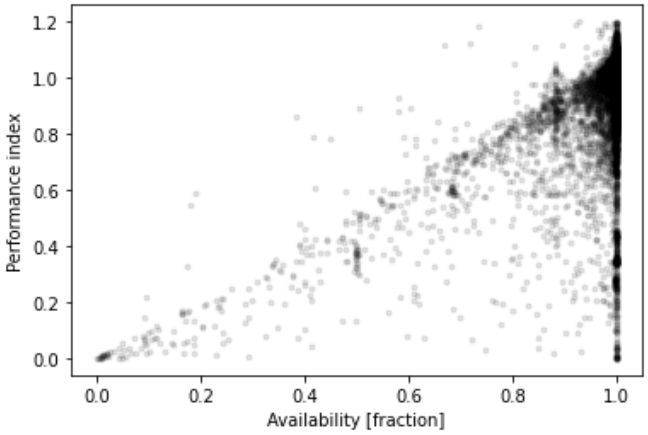

Figure 7. Performance Index vs. availability. Some system-months with low performance index also show low availability, indicating that their low PI is at least partially due to availability. Further PI values are corrected by availability % to account for this effect where noted.

An investigation into the timing of system availability losses indicates most inverter outages occur early in the system's life, possibly because of initial start-up issues or early equipment failures (Figure 8).

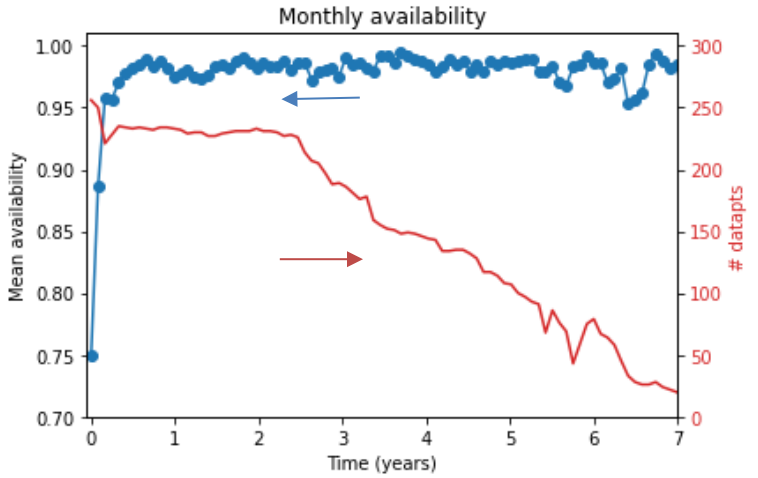

Figure 8. Monthly Inverter availability vs. time. Early-life failures or start-up issues occur most frequently in the first year of system lifetime.

The overall availability in this data set is 97.7%, excluding the first year. As shown in Figure 8, the first 6 months have a lower availability at 92% on average due to start-up issues. By comparison, the standard PVWatts default value assumes a static 97% availability over the system lifetime. Other financial modeling cases use different values—for instance, the default value used in the independent engineering (IE) community is typically 99% availability, which may be aggressive based on these results.

4.2 Start-Up Issues

As shown in Figure 8, early-life outages can suppress monthly performance. Even after correcting for availability as described in Section 4.1, we still find low overall performance in the early months of system life. In Figure 9, we show the PI as a function of year of operation, which indicates steady performance in years 1–6, but low performance in the initial year. (Years 7 and 8 are also low but only have a small number of data points and may not be representative). Start-up issues in Year 0 can have several different causes, even after excluding inverter availability. In general, these can reflect the difficulty of bringing large PV systems online, including "partial outages" where DC strings or modules are offline, but inverters are generating partial power. Some new components may not work when initially installed ("infant mortality"), and it may take months to source replacements and identify and correct all installation errors. This is often called a "start-up, "break-in," or "shake-down" period, and data indicates it is on the order of 4–12 months duration following commencement of operations.

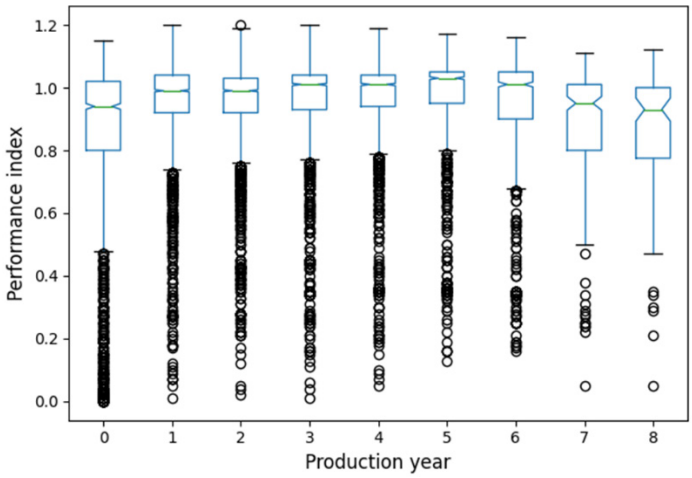

Figure 9. Box plots of monthly PI for sequential production year. The median for each year is indicated by a horizontal line. Year 1 performance is generally lower for a percentage of systems. Performance drops again in years 7 and 8, although the sample size is very small.

Because of the low year-1 performance for a percentage of systems, we exclude initial performance from subsequent PI distributions, assumed to be the first four months of system operation.

4.3 Snow and Winter Underperformance

Winter system underperformance is well-documented, particularly in northern climates that experience snowfall. Other causes of underperformance can include increased shading due to low sun angles. In Figure 10 we show monthly performance index after correcting for measured system availability and excluding start-up data. It is clear that winter months exhibit underperformance of 5%–10%, averaged across the entire fleet.

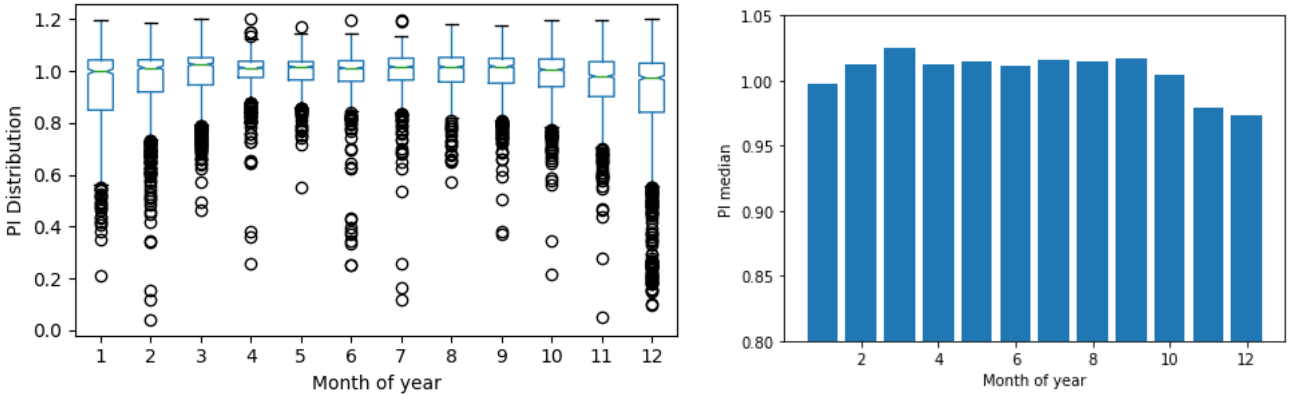

Figure 10. Distributions of monthly PI (left) and median values (right), corrected for system availability and excluding the first 4 months of system operation. Winter months show underperformance compared to summer months, likely due to snow.

In Figure 11 we show the impact of measured snowfall on mean PI, corrected as above. Snowfall values are monthly observed snow data obtained from The Weather Company (2018). For this figure, monthly data are sorted by snowfall amount, with an PI values averaged at each 1cm bin, reducing our 12,000 data points to 90 discrete points. A measurable decrease in PI is seen for monthly snowfall of 1 cm or more, with greater loss in PI with more measured snowfall. Even 1 cm of monthly snowfall was enough

to lower monthly PI from 0.99 to 0.92 on average. For months with measurable snow, the data reveal a simple linear dependence on cumulative monthly snowfall: PI = 0.89 – 0.003*snowfall[cm].

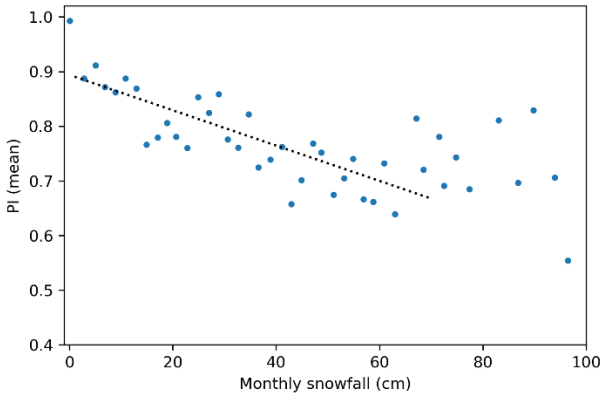

Figure 11. Average PI vs. monthly snowfall in centimeters. Linear regression shows monthly PI loss increases with cumulative monthly snowfall.

After excluding months with >1 cm of monthly snowfall recorded, winter underperformance is improved considerably compared with what is shown in Figure 10. Post-filtering, the lowest-performing months of November and December exhibit a median of PI = 0.985, which brings them within 2.5% of the average, a major improvement from the previous case. The small remaining offset is possibly due to shading or other effects, which have not been isolated from the data set.

4.4 Filtered Performance Index Distribution

After correcting for inverter availability, removing the first 4 months of performance data to eliminate start-up issues, and eliminating points with over 1cm of monthly snowfall, we have a remaining 10,565 monthly data points. These show a much narrower distribution, with a mean at PI = 0.994 and median value of 1.02. The P90 value (90% probability of exceedance) is improved from 0.81 to 0.90. The results also follow more closely to a Gumbel extreme-value distribution, from Eq. (6), where best-fit parameters for our filtered distribution are μ= 1.03 and β =0.06. RMS fit error is reduced by half compared with the original unfiltered distribution.

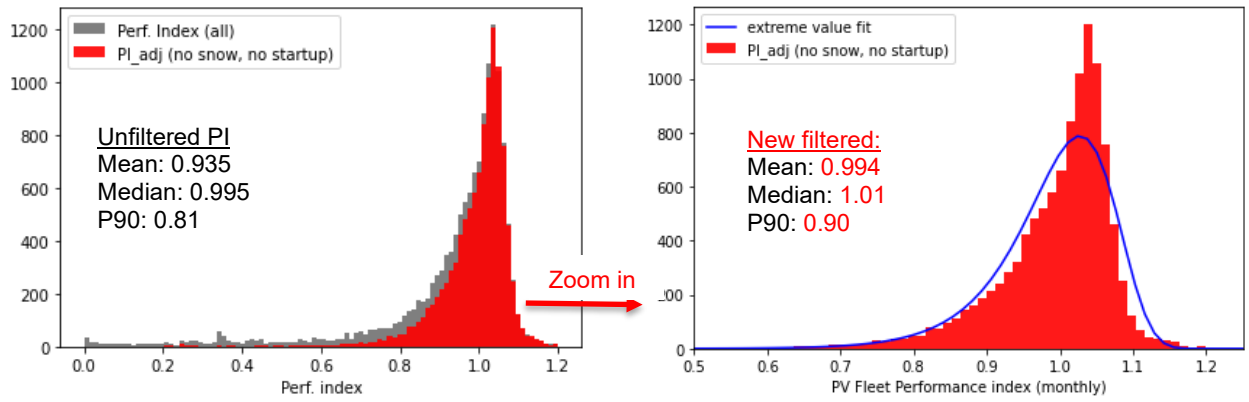

Figure 12. Postfiltering PI distribution (red), correcting for availability loss and filtering out initial 4 months of data and monthly snowfall >1cm. An improved fit to the Gumbel distribution and tighter P90 value are shown.

4.5 Residual Model Effects

Residual model effects are evaluated by comparing calculated PI (availability-corrected and filtering out start-up and snow effects) to other model parameters. The comparisons in Figure 13 show an overall regression of monthly PI versus ambient wind and temperature conditions, averaged monthly and weighted for irradiance. The irradiance weighting is to remove the influence of nighttime conditions, which are irrelevant for energy model purposes. If our performance model is doing a reasonably good job of correcting for wind speed, temperature, and irradiance, there should be little to no trend relative to these parameters.

The left image shows PI binned by average monthly wind speed in 0.1 m/s increments. This displays fairly constant performance in the range of 2 m/s–4 m/s, where most of our collected data lie. Expanding to a larger range of wind speed values suggests an apparent negative trend in PI versus wind speed. It is not entirely clear whether this is a real trend or an aberration based on other correlated factors when there are only a few data points.

Ambient temperature (Figure 13, right) shows a similar trend when PI is binned by temperature in 0.3 ºC increments. In the ambient temperature range of 15°–35°C, where most of our monthly data lie, there is a relatively constant PI value. However, expanding the range to the full 0°C–45°C range suggests a consistent positive trend, with greater calculated PI at higher ambient temperature. This may indicate temperature is being overcorrected, such as if our assumed temperature coefficient of –0.47 %/°C is too large in magnitude. It may also just be an effect of correlated parameters occurring where data availability is low.

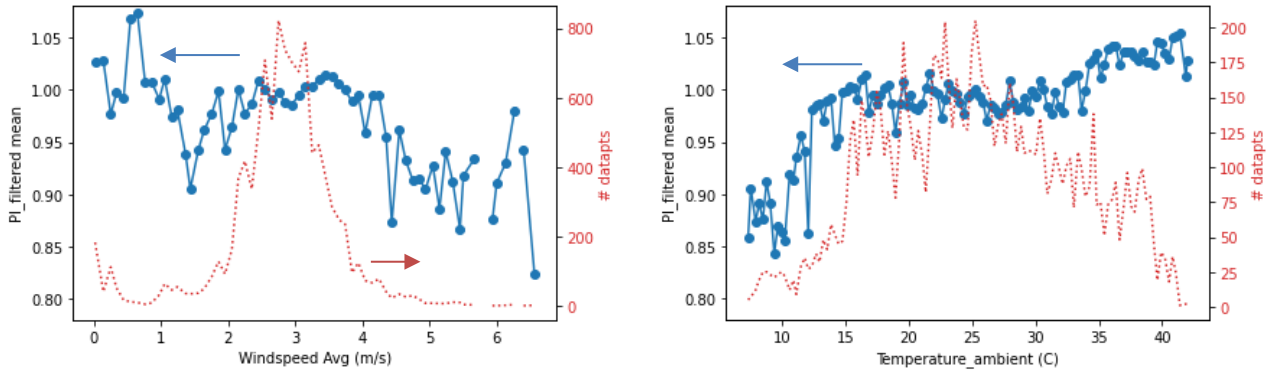

Figure 13. Monthly PI (corrected for availability and excluding snow and start-up) averaged for each wind speed value (left) showing no consistent trend. A consistent positive correlation is seen between PI and ambient temperature (right).

An additional trend is revealed in Figure 14, where the PI is grouped by the PV mounting configuration type. Commercial roof-mounted systems exhibit a ca. 4% lower PI than ground-mounted systems. Although this difference is statistically significant, it is possible that other correlated effects, such as module technology and installation quality, may still be present and are difficult to separate. However, a

similar impact of mounting configuration was noted in a previous publication (Jordan 2020). This study is primarily concerned with larger commercial roof-mount systems, not residential-scale systems.

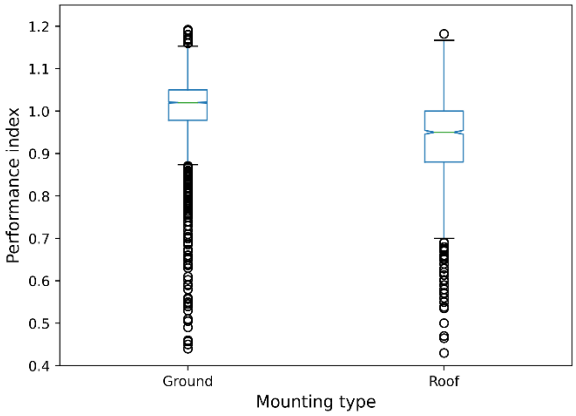

Figure 14. PI (corrected for availability and excluding snow and start-up) grouped by the mounting configuration. The size of the notch indicates the 95% confidence interval. Roof-mounted systems have a statistically lower performance than ground-mounted systems.

4.6 Statement of Uncertainty

Expected energy from Eq. 3 forms the denominator of the PI calculation and therefore directly affects the accuracy of the monthly PI distribution. The accuracy of the NSRDB satellite irradiance used in Eq. 3 is therefore of utmost importance in defining the uncertainty of our approach.

Here we estimate modeled monthly uncertainty of POA irradiance based on NSRDB resource data to be +/- 10% at the 95% confidence interval. This is based on two components of uncertainty: one being uncertainty of the global horizontal irradiance (GHI) reported from the NSRDB, and the other being errors from transposition of direct normal irradiance (DNI) and DHI components to the tilted POA. The first uncertainty component is addressed in Figure 2 of Habte and Sengupta (2017), showing uncertainty of the GHI at a monthly timescale ranges from +/-6% to +/-12% for nine validation stations. We take a root-sum-square of each of these individual station uncertainties to generate an uncertainty for the stations as a group of +/- 8.3% for GHI. This can be confirmed by our own comparison with ground-measured SURFRAD sites in Figure 15 below, showing a similar 6%–12% accuracy in satellite-measured GHI. Note that accuracy is better in high-DNI, cloud-free climates.

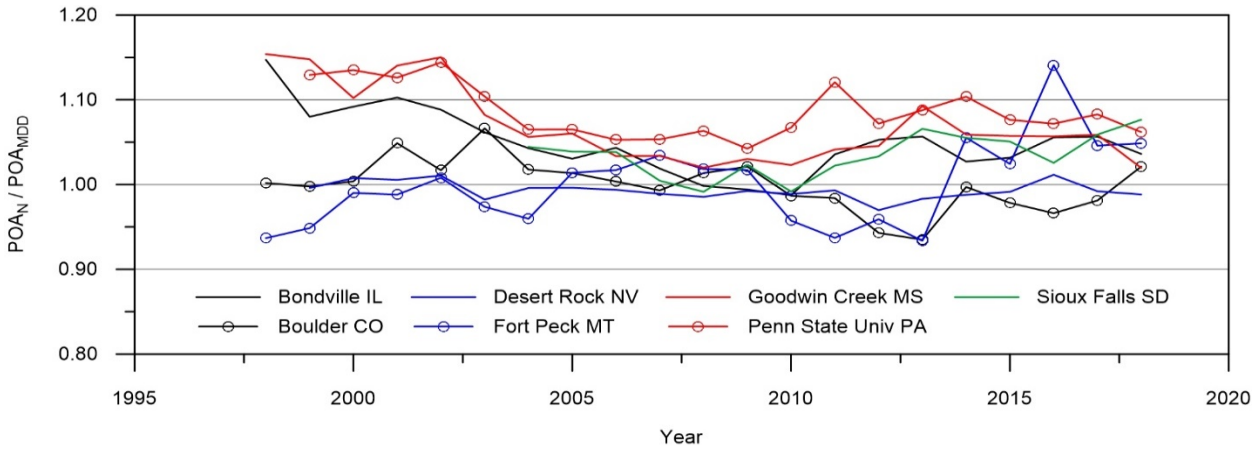

Figure 14. Ratio of annual NSRDB all-sky irradiance to field-measured irradiance for seven SURFRAD ground station sensor sites. NSRDB irradiance tends to be overestimated for cloudy climates, which can result in low calculated PI at these sites (Marion, 2020).

The second uncertainty component for horizontal-to-tilted transposition error comes from Hay and McKay (1988), where the uncertainty of the model transposition to POA was found to be +/- 5%. A root sum squared combination of the GHI and model transposition uncertainty results in a POA uncertainty of +/- 9.7%, rounding to +/- 10%.

Other sources of uncertainty in the method include revenue-grade meter accuracy of 0.1%, which is neglected. Other sources of variability in the PI distribution can include differences in inverter efficiency, assumed to be at 96.4% for all inverters, and assumptions about system DC nameplate rating. Uncertainties in these values could result in 1%–2% over- or underestimates of modeled energy. As described in Section 4.5, other errors in our thermal model due to wind-speed resource data, module temperature coefficient and normal operating cell temperature (NOCT) can also contribute several percent to model uncertainty. We also consider larger systems with multiple stated tilt and azimuth values as a single combined system with uniform orientation for the purposes of this analysis, which will reduce the accuracy of the modeled performance for these systems.

References

Anderson K. and M. Mikofski. “Slope-Aware Backtracking for Single-Axis Trackers”, NREL/TP-5K00-76626, National Renewable Energy Laboratory, 2020. doi: 10.2172/1660126

Anderson, K. and Ryan Blumenthal. "Overcoming Communications Outages in Inverter Downtime Analysis." Presented at the 47th IEEE Photovoltaic Specialists Conference (PVSC 47). 2020.

Deline C., Muller M., Deceglie M., Jordan D., Anderson K., Simpson L., Perry K.,White R.“ PV Fleet Performance Data Initiative: March 2020 Methodology Report”, NREL Technical report NREL/TP-5K00-76687 https://www.nrel.gov/docs/fy20osti/76687.pdf. 2020

Dierauf, T., Growitz, A., Kurtz, S., Cruz, J. L. B., Riley, E., and Hansen, C. Weather-Corrected Performance Ratio. United States: N. p., 2013. Web. doi:10.2172/1078057.

Dobos, A. P. PVWatts Version 5 Manual. United States: N. p., 2014. Web. doi:10.2172/1158421.

Fuentes, M. K., 1987, “A Simplifed Thermal Model for Flat-Plate Photovoltaic Arrays”, SAND85-0330, Sandia National Laboratories, Albuquerque NM. http://prod.sandia.gov/techlib/access-control.cgi/1985/850330.pdf

Habte, A., Sengupta, M., 2017. Best Practices of Uncertainty Estimation for the National Solar Radiation Database (NSRDB 1998-2015), in 33rd EUPVSEC.

Hay ,J., McKay, D., 1988. Final Report IEA Task IX – Calculation of Solar Irradiances for Inclined Surfaces: Verification of Models which use Hourly and Daily Data. International Energy Agency Solar Heating and Cooling Programme.

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

Holmgren W., Calama-Consulting, Tony Lorenzo, Uwe Krien, bmu, Mark Mikofski, … Alaina Kafkes. (2019, May 15). pvlib/pvlib-python: v0.6.3 (Version v0.6.3). Zenodo. http://doi.org/10.5281/zenodo.2850192

IEC 61724-1:2017

Jordan D.C, Wohlgemuth J.H, Kurtz S.R., 2012, Technology and climate trends in PV module degradation. in 27th European Photovoltaic Solar Energy Conference. Frankfurt: Germany, 2012. 3118–3124.

Jordan D.C, Marion B., Deline C. Barnes T., Bolinger M., PV field reliability status—Analysis of 100,000 solar systems, Progress in PV 28 (8), 2020, 739 – 754.

Karin T, Jones C.B, Jain, A "Photovoltaic Degradation Climate Zones," 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), Chicago, IL, USA, 2019. doi: 10.1109/PVSC40753.2019.8980831

Lorenzo, E et al., 2011, “Tracking and back-tracking”, Prog. in Photovoltaics: Research and Applications, v. 19, pp. 747-753.

Marion B, “Measured and satellite-derived albedo data for estimating bifacial photovoltaic system performance”, Solar Energy 2020, submitted

Mokri, J. and Cunningham, J. PV System Performance Assessment. 2014.

Perez, R., Ineichen, P., Seals, R., Michalsky, J., Stewart, R., 1990. Modeling daylight availability and irradiance components from direct and global irradiance. Solar Energy 44 (5), 271-289.

RdTools, version 2.0.0, https://github.com/NREL/rdtools, DOI:10.5281/zenodo.4300113

Reda I. and A. Andreas, Solar position algorithm for solar radiation applications. Solar Energy, vol. 76, no. 5, pp. 577-589, 2004.

Sengupta, M. et al. The National Solar Radiation Data Base (NSRDB). Renewable and Sustainable Energy Reviews, 2018. doi: 10.1016/j.rser.2018.03.003

Townsend T., C. Whitaker, B. Farmer and H. Wenger, “A new performance index for PV system analysis,” Proceedings of 1994 IEEE 1st World Conference on Photovoltaic Energy Conversion - WCPEC, Waikoloa, HI, 1994, pp. 1036-1039 vol.1, doi: 10.1109/WCPEC.1994.520138.

The Weather Company, IBM “Cleaned Observation API Documentation”, April 2018, http://cleanedobservations.wsi.com/documents/WSI%20Cleaned%20Observations%20API%20Documentation.pdf