The Effect of Inverter Loading Ratio on Energy Estimate Bias

Preprint

Kevin S. Anderson,[1] William B. Hobbs,[2]
William F. Holmgren,[3] Kirsten R. Perry,[1] Mark A. Mikofski,[3]
and Rounak A. Kharait[3]

1 National Renewable Energy Laboratory
2 Southern Company
3 DNV

Presented at the 49th IEEE Photovoltaic Specialists Conference (PVSC 49)
Philadelphia, Pennsylvania
June 5-10, 2022

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

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

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5K00-82812
July 2022

The Effect of Inverter Loading Ratio on Energy Estimate Bias

Preprint

Kevin S. Anderson,[1] William B. Hobbs,[2]
William F. Holmgren,[3] Kirsten R. Perry,[1] Mark A. Mikofski,[3]
and Rounak A. Kharait[3]

1 National Renewable Energy Laboratory
2 Southern Company
3 DNV

Suggested Citation
Anderson, Kevin S, William B. Hobbs, William F. Holmgren, Kirsten R. Perry, Mark A. Mikofski, and Rounak A. Kharait. 2022. *The Effect of Inverter Loading Ratio on Energy Estimate Bias: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5K00-82812. https://www.nrel.gov/docs/fy22osti/82812.pdf.

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

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

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5K00-82812
July 2022

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

NOTICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy under Solar Energy Technologies Office (SETO) Agreement Numbers 34348. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

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

The Effect of Inverter Loading Ratio on Energy Estimate Bias

Kevin S. Anderson[1], William B. Hobbs[2], William F. Holmgren[3], Kirsten R. Perry[1], Mark A. Mikofski[3], and Rounak A. Kharait[3]
[1]NREL, Golden, CO, 80401, USA
[2]Southern Company, Birmingham, AL, 35203, USA
[3]DNV, San Diego, CA, 92123, USA

***Abstract*—Subhourly effects, particularly variability in solar irradiance, can lead to underestimation of inverter clipping losses and overestimation of energy in hourly photovoltaic system performance models, particularly for systems with high inverter loading ratios. Direct simulation of this error can be complicated by factors such as the representation of spatial and temporal variability in hourly weather data and transient system conditions. In this work we take an alternative approach using real system power measurements to show that energy predictions from typical industry models suffer from a bias that increases with inverter loading ratio. We also show that this loading ratio-dependent bias is strongly correlated with an empirical subhourly inverter clipping bias derived from real power plant data. Finally, we show that this bias is not necessarily specific to any one model or weather dataset by recreating similar biases with alternatives of each.**

***Index Terms*—photovoltaic, inverter, clipping, modeling, high-frequency, subhourly, irradiance, variability**

I. Introduction

Utility-scale photovoltaic (PV) system design is increasingly trending over time to larger inverter loading ratios (ILR), also referred to as DC:AC ratios [1]. PV inverters with high loading ratios must force their arrays into reduced-efficiency operation in sunny conditions to prevent the total array power output from exceeding the inverter's maximum-rated input power. This power-limiting behavior is called clipping because it disrupts the linear relationship between irradiance and output power, resulting in curtailed performance in high irradiance conditions. An inverter might clip for several hours continuously on a clear-sky day, or only intermittently on days with highly variable irradiance when high-irradiance spikes might last for less than one minute.

The detailed system performance models used by industry developers and financiers to forecast project revenue usually include adjustments for inverter clipping and many other system performance effects. However, the conventional practice of modeling system performance at hourly scale renders these models incapable of directly simulating short-duration effects like subhourly clipping. In effect, these models assume static operating conditions over each hourly simulation interval, causing the effects of subhourly irradiance variability to be overlooked. This causes an otherwise accurate performance model to overestimate production, especially for systems with large loading ratios in climates with high irradiance variability.

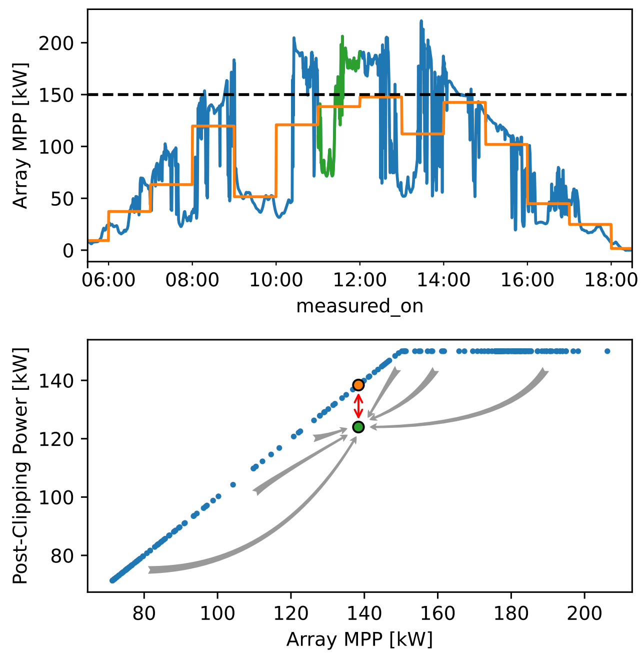

Fig. 1. Conceptual visualization of subhourly clipping. Upper subplot: high-resolution array maximum power point (MPP) data (blue line), and the corresponding average hourly values (orange line). Green line shows a particular hourly interval. Dashed black line shows a hypothetical inverter clipping point. Many MPP values lie above the clipping point at high resolution, but the hourly averages are all below the clipping point. Lower subplot: visualization of the 11:00–12:00 interval shown in green in upper subplot, with high-resolution values (blue dots), the true average of the high-resolution values (green circle), and the naive average that applies clipping at hourly scale (orange circle). Applying clipping at hourly scale overlooks the subhourly clipping loss at higher irradiance, leading to overestimated power output (red arrow).

It is worth emphasizing that this subhourly clipping bias is not a recoverable loss caused by poor system operation, but rather a failure of conventional modeling techniques to fully capture real-world PV system behavior.

A. What is subhourly clipping?

The mathematical foundation of subhourly clipping error is relatively straightforward: inverter clipping is a strongly nonlinear behavior, and the average value a nonlinear function takes across some interval is, in general, not equal to the value the function takes at the average of that interval. Thus,

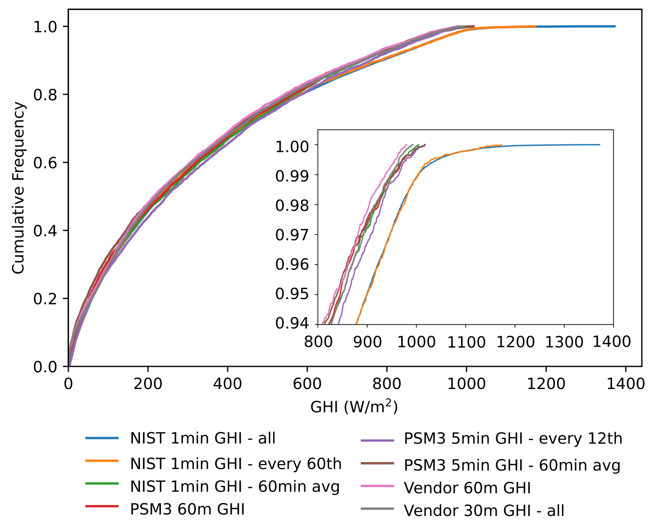

Fig. 2. Cumulative distributions of one year of Global Horizontal Irradiance (GHI) values from minute-interval ground measurements and two satellite datasets for the NIST site, along with a zoomed-in view of the upper end of the CDFs (inset). Regardless of data source (ground measurement and satellite), hourly values do not show the high-irradiance tail observed in 1-minute ground measurements.

to the extent that a particular hourly interval engages with that nonlinearity, the simulated inverter output at the average irradiance will be different from the inverter output averaged across all irradiance values in the interval. Note that similar arguments apply to the other nonlinearities in PV system response, meaning subhourly clipping is not the only contributor to hourly modeling error [2], [3]. Fig. 1 visualizes this effect, showing an example where subhourly irradiance spikes would cause intermittent clipping loss but the average irradiance is low enough for an hourly model to ignore clipping entirely.

B. *What are challenges with using satellite data?*

Two limitations of satellite-based irradiance datasets are worth mentioning here. The first is that, because these datasets are derived in part from geostationary satellite imagery, their spatial resolution is limited by the the satellite's imaging optics and sensors. In particular, today's datasets are limited to kilometer-scale resolution, which is large compared with individual PV arrays. The second is the limited temporal frequency, which again is constrained by the corresponding imaging limits of the satellite. Together, these two limitations mean that our current satellite-based irradiance datasets are unable to recreate irradiance at the scale of individual PV arrays, at least without a statistical rescaling or similar downscaling step. Some combination of spatial resolution limitations and steps taken to convert imagery to irradiance results in a distribution of irradiance values that do not completely recreate the distribution of ground measurements, even when the satellite data are not time averages.

This is shown in Fig. 2, where maximum Global Horizontal Irradiance (GHI) values are clearly higher in both full and sub-sampled ground measurements than for averaged and "instantaneous" satellite-based data from both PSM3 and a commercial vendor. This means that, despite previous work suggesting that using "instantaneous" satellite-based irradiance data can partially avoid subhourly clipping bias [4], these satellite-based datasets are still missing a key characteristic needed to fully model the effect of inverter clipping, regardless of temporal resolution.

C. *This work*

In the absence of an industry-wide shift to higher resolution models and weather data, previous efforts to correct this bias have focused on estimating post-hoc adjustments to forecasts from conventional models. Several approaches for generating these correction factors have been explored, including temperature-corrected insolation ratios [5], machine learning [6]–[8], and direct simulation using ground-measured 1 minute irradiance [9] or NSRDB PSM3 5 minute irradiance [4], [10], [11]. However, it is not obvious how to rigorously validate these correction methods: direct comparison of modeled and measured clipping loss is usually only possible in small research installations with different performance characteristics from utility-scale systems, while comparisons of overall power output are confounded by a myriad of unrelated modeling errors and real-world plant performance issues.

Furthermore, it may be premature to even attempt to correct a subhourly clipping bias, as there remains doubt in the community that real-world revenue projections are meaningfully affected by this error [11]. In this work, we seek to bring clarity to the issue by recreating typical industry modeling procedures, comparing the model output with measured data from real systems to investigate how the overall model bias varies with ILR, and showing that the model bias variation with ILR is consistent with the bias expected from subhourly clipping. This is made possible by two key insights: first, we can generate hourly production for a population of systems spanning a range of ILRs but with otherwise identical performance characteristics by artificially applying inverter clipping to high-resolution power data from a low-ILR system; and second, we can use the same high-resolution data to calculate an empirical ILR-dependent subhourly clipping bias to compare with the overall ILR-dependent model bias. Together, these insights allow us to avoid the confounding effects inherent to a more conventional multi-system analysis.

II. Methods

A. *Weather Datasets*

Several satellite-based weather datasets are commonly used for PV performance modeling. Although publicly available satellite-based irradiance datasets exist (for example, the Physical Solar Model v3 (PSM3), part of the National Solar Radiation Data Base (NSRDB) [12]), commercial systems are typically financed using models based on proprietary commercial irradiance data. Here we use hourly interval data from a popular commercial data vendor. We choose single-year weather datasets so that we can directly compare modeled system output with measured system output for the same time periods. This is a divergence from the typical modeling procedure which uses a hypothetical "typical meteorological year" (TMY) dataset. However, because TMY datasets are simply combinations of subsets of single-year datasets, we expect single-year and TMY datasets to share whatever characteristics are relevant to subhourly modeling and therefore do not expect this difference to materially affect our conclusions.

B. PV System Datasets

We examine datasets from six PV systems in the United States representing a range of climates and configurations, summarized in Table I. In all cases, the high-resolution power data are 60 second averages of higher frequency data sampled at 1–15 s intervals depending on the system. Note that, because we perform this analysis at the inverter level (instead of the more typical plant level), the system metadata in Table I reflects the configuration of the chosen inverter rather than the plant as a whole. To add relevant geographical and climate details for the sites, particularly for the anonymous commercial plants, we also included Solar Forecast Arbiter Climate Zone (SFACZ) [13] and Solar Variability Zone (SVZ) [14] for each site.

In each case we analyze data from a single calendar year to facilitate associating the ground-measured data with the corresponding satellite-based weather dataset. The specific calendar year chosen for each systems varies in order to minimize the effect of data gaps and substantial performance issues. However, some data were still excluded because of abnormal operation (three days for Commercial Plant 2 for tracking issues; 41 days for NREL for snow coverage on the array). Additionally, any hourly intervals in the measured power containing zero, negative, or null values were dropped from both the measured and modeled datasets.

Although most of these datasets are proprietary and commercial, the underlying high-frequency power data are publicly available for the National Renewable Energy Laboratory (NREL) [15] and National Institute of Standards and Technology (NIST) [16], [17] systems. The public names for these two systems are "[1283] NREL Research Support Facility II" and "NIST Ground Mount Array", respectively.

C. Empirical Subhourly Clipping Bias

Evaluating the magnitude of subhourly clipping bias from measured data is not straightforward. In principle it could be modeled using a sufficiently accurate performance model and high-resolution weather data, essentially by taking the difference in the model's predictions when running at hourly and (approximately) instantaneous scales. This difference in modeling predictions has been referred to as "clip, then average" versus "average, then clip" to indicate the timescale at which inverter clipping is applied to the "unclipped" array output [18]–[20].

However, it is not clear that current modeling approaches are sufficiently accurate at such a high resolution to generate the "unclipped" power signal, especially for the purpose of predicting the effects of irradiance variability: spatial nonuniformity of irradiance, thermal transience, and other short-duration effects complicate modeling efforts and few model validation studies are done at the short timescales relevant here. Additionally, satellite-based data often represent instantaneous measurements over some geographic area [21], and hourly instantaneous measurements have been demonstrated to result in less bias than hourly averaged measurements [4], further complicating this issue.

We propose an alternative approach: instead of attempting to recreate an "unclipped" array output via conventional weather-to-power modeling, we instead use real power measurements from an inverter with low loading ratio. Because the inverter has low ILR, it rarely if ever clips, meaning its power output is a perfectly realistic "unclipped" signal we can then use with the "clip, then average" and "average, then clip" approach of estimating subhourly clipping bias at any loading ratio of interest. Crucially, this lets us estimate an "empirical" subhourly clipping bias that varies with ILR without fear of model error and while holding all other system parameters constant. This empirical bias can then be compared with the actual bias of an hourly model, again controlling for all effects except ILR. As mentioned above, the empirical subhourly clipping bias is calculated using the "average, then clip" vs "clip, then average" method:

$$\text{bias} = \frac{E_{\text{AtC}} - E_{\text{CtA}}}{E_{\text{CtA}}} \tag{1}$$

where E_{AtC} is total energy calculated by averaging measured data to 60 minute intervals and then artificially clipping (analogous to what a conventional hourly simulation model would calculate) and E_{CtA} is energy calculated by first clipping at 1 minute intervals and then averaging to 60 minutes (analogous to what a real system would do). For each system, this bias is evaluated at ILRs of 1.2, 1.3, 1.4, and 1.5 (with the exception of Commercial Plant 1, which does not have low enough nominal ILR for 1.2 to be relevant), corresponding to the typical range of ILRs seen in new systems today [1].

Although this approach of calculating a "true" subhourly clipping bias avoids the majority of possible model bias, the artificial clipping is still a potential source of error. The artificial clipping is applied using a straightforward numerical threshold where power values are clamped to not exceed the desired AC capacity. Although this is a reasonable approximation of the clipping behavior of many real-world inverters, it does ignore secondary effects like thermal throttling and dynamic plant control that cause the inverter's clipping limit to vary with time. We consider this approximation acceptable, as thermal throttling is considered unusual for most inverters, and the systems analyzed here did not have notable overbuilds of inverter capacity relative to interconnection limits, and therefore did not reflect impacts of dynamic plant controllers.

To eliminate the effect of other modeling biases (discussed more in the next section), each system's "unclipped" power signal was rescaled to have zero bias with respect to its

TABLE I
SUMMARY OF PV SYSTEM CONFIGURATIONS

System	Size [kW$_{dc}$]	ILR	Rack	Location	Year	SFACZ[1]	SVZ[2]
NIST	271.0	1.04	Fixed Tilt	Gaithersburg, Maryland	2018	7	Moderate (low)
Commercial Plant 1	4609.2	1.15	Single-Axis Tracking	Southeast US	2018	6	Moderate (low)
Commercial Plant 2	594.4	1.19	Single-Axis Tracking	Southeast US	2020	7	Low
NREL	204.1	0.82	Fixed Tilt	Golden, Colorado	2020	4	Moderate (high)
SSRC 1-Axis	2.4	0.80	Single-Axis Tracking	Birmingham, Alabama	2019	6	Moderate (low)
SSRC 30S	2.4	0.80	Fixed Tilt	Birmingham, Alabama	2019	6	Moderate (low)

[1] Solar Forecast Arbiter Climate Zone [13]
[2] Solar Variability Zone [14]

nominal PVsyst model output. By doing this we can view the rescaled "unclipped" power data as the 1 minute analog of the hourly PVsyst model. The rescaling coefficients for each system are as follows, where a coefficient of 1.0 indicates no difference: 1.03 (NIST), 0.96 (Commercial Plant 1), 0.97 (Commercial Plant 2), 1.10 (NREL), 1.02 (SSRC 1-Axis), 1.03 (SSRC 30S). It is unclear why the NREL system data required such a large rescaling to match the output of its nominal PVsyst model.

D. Hourly Performance Models

In the authors' experience, the commercial simulation software PVsyst [22] drives the majority of utility-scale PV system financing. In typical usage a PVsyst model describing the system configuration is applied to an hourly weather dataset and produces a corresponding hourly production time series. We do the same here to mimic typical usage.

We used PVsyst version 7.1.4 to create models of each PV system in this study. We used PAN module and OND inverter files supplied by PVsyst when available and created PAN and OND files from component specifications when needed. We created a "nominal" model for each system that accurately describes the as-built system, including layout for shading. To simulate higher ILRs, we created new variants of the nominal system. For most systems, we increased the ILR by increasing the number of strings in parallel until the ILR was within 0.01 of the desired ILR. For the smaller SSRC systems with only one string of 10 modules, we instead increased the number of modules in series to reach an ILR close to but above the desired ILR (e.g. 1.36 for a desired 1.30), and then adjusted the module quality factor loss parameter to match the desired ILR. All other losses were left at their default settings. Several variants produced a voltage or current higher than the inverter specifications and OND file allow for. In these cases we increased the maximum inverter voltage or current such that PVsyst allowed the model to run and the inverter over-current and over-voltage losses were 0.0%. This represents a minor deviation from the as-built system and does not jeopardize the conclusions of this work. Finally, although these systems likely experienced gradual performance loss in the field for several years prior to the time period used in this analysis, we did not explicitly include the effect of this cumulative performance degradation in the PVsyst models. Instead, we rely on the rescaling procedure described in Section II-C to account for this capacity loss.

Analogous to Eq. 1, the model bias is calculated as:

$$\text{bias} = \frac{E_{\text{PVsyst}} - E_{\text{CtA}}}{E_{\text{CtA}}} \tag{2}$$

where E_{PVsyst} is PVsyst's modeled output using hourly satellite data and the ILR corresponding to E_{CtA}.

We also used a PVWatts-style [23] model in pvlib [24], [25] with both vendor and PSM3 [12] weather data for a single site, NIST. This relatively naive model, with two sets of weather data, was selected to demonstrate that this bias issue is not unique a specific performance model or hourly satellite-derived weather data source.

III. Results

Fig. 3 compares the PVsyst bias and empirical subhourly clipping bias for each system and ILR. For completeness, the biases are also listed in tabular form in Table II. The two biases have a roughly 1:1 relationship, with root mean squared error of 0.80% and mean bias error of 0.44%. Note that these error statistics are in the original bias units (percent of annual production) and reflect absolute difference of the two biases, not relative difference.

Fig. 4 and Table III show the same results for the NIST site from Fig. 3, with the addition of model bias values from the PVWatts-style pvlib model using both PSM and commercial vendor weather data. The simple pvlib model's bias varies similarly to the PVsyst and empirical biases, although this pvlib model's bias is somewhat higher, even when using the same weather dataset as PVsyst.

IV. Discussion

The strong positive correlation between model bias and ILR in Fig. 3, as well as the rough range of model bias (0–4%) for these ILRs, are consistent with predictions from simulation-based studies [18]–[20]. Fig. 4 uses the NIST site to demonstrate that this model bias is not unique to PVsyst or the particular commercial vendor's weather data.

Contrary to what one might expect, these results indicate that subhourly clipping bias is not restricted to humid climates with high variability like those of the Southeast US; even the semi-arid climate of the NREL system (approximately

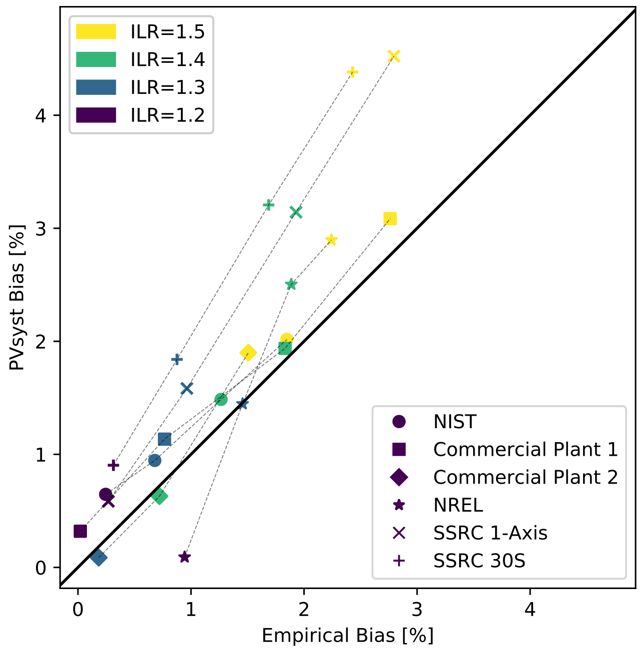

Fig. 3. Comparison of the ILR-dependent PVsyst and empirical biases. Colors indicate the ILR and symbols indicate the system for each pair of biases. The PVsyst ILR-dependent model bias is generally well-predicted by the empirical subhourly clipping bias derived from the measured system power data.

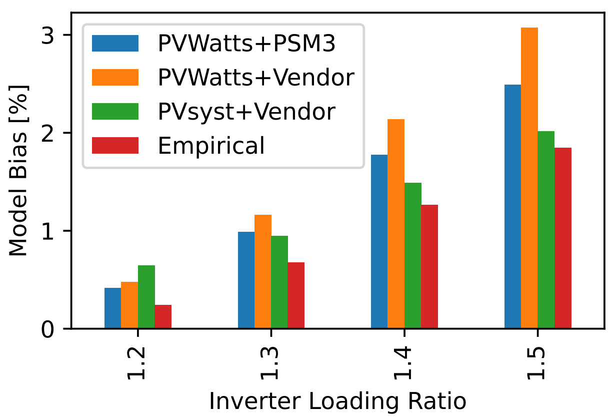

Fig. 4. Comparison of empirical subhourly clipping bias with hourly model bias for three modeling approaches: PVsyst with vendor irradiance, PVWatts with vendor irradiance, and PVWatts with PSM3 irradiance. The two PVWatts models are biased somewhat high compared to the PVsyst model, but follow the same general trend.

1750 meters above sea level) shows 2–3% bias at ILR=1.5. Conceptually, the error only requires that the system operates both below and above the clipping point for a portion of the hour, but the error is independent of the ordering of the points or the number of transitions i.e. the variability. However, we have not characterized the variability at these sites beyond listing their pre-existing zone classifications in Table I. Future work may include a more detailed investigation of the relationship between subhourly irradiance variability and subhourly clipping bias.

TABLE II

RESULTS: ILR-DEPENDENT BIASES (PVSYST+VENDOR)

System	ILR	PVsyst Bias [%]	Empirical Bias [%]
NIST	1.2	0.6	0.2
NIST	1.3	0.9	0.7
NIST	1.4	1.5	1.3
NIST	1.5	2.0	1.8
Commercial Plant 1	1.2	0.3	0.0
Commercial Plant 1	1.3	1.1	0.8
Commercial Plant 1	1.4	1.9	1.8
Commercial Plant 1	1.5	3.1	2.8
Commercial Plant 2	1.3	0.1	0.2
Commercial Plant 2	1.4	0.6	0.7
Commercial Plant 2	1.5	1.9	1.5
NREL	1.2	0.1	0.9
NREL	1.3	1.4	1.5
NREL	1.4	2.5	1.9
NREL	1.5	2.9	2.2
SSRC 1-Axis	1.2	0.6	0.3
SSRC 1-Axis	1.3	1.6	1.0
SSRC 1-Axis	1.4	3.1	1.9
SSRC 1-Axis	1.5	4.5	2.8
SSRC 30S	1.2	0.9	0.3
SSRC 30S	1.3	1.8	0.9
SSRC 30S	1.4	3.2	1.7
SSRC 30S	1.5	4.4	2.4

TABLE III

ILR-DEPENDENT BIASES FOR NIST WITH DIFFERENT MODELS AND WEATHER DATASETS.

ILR	PVWatts+ PSM3 [%]	PVWatts+ Vendor [%]	PVsyst+ Vendor [%]	Empirical [%]
1.2	0.4	0.5	0.6	0.2
1.3	1.0	1.2	0.9	0.7
1.4	1.8	2.1	1.5	1.3
1.5	2.5	3.1	2.0	1.8

It is noteworthy that the largest divergences from a perfect 1:1 relationship between the PVsyst and empirical biases are from the smallest systems, SSRC 1-Axis and SSRC 30S. One possible explanation for this is related to the spatial variability of irradiance: the smaller the array, the less spatial averaging and thus the more output variability it experiences [26]. This suggests that 1 minute averaged data are not able to fully resolve the output variability of these small arrays [26] and the empirical bias calculated here is an underestimate of the true value. It is also possible that the satellite-based weather dataset used for the SSRC systems (they are co-located and covered by the same satellite pixel) has some undiscovered characteristic that disrupts this analysis. In any case, the agreement of the PVsyst and empirical biases is much better for the four larger systems.

Imperfect datasets are another source of uncertainty in the empirical subhourly clipping bias. The two commercial systems have higher loading ratios than the other systems and their power datasets do include some clipping. Similarly to the spatial averaging issue discussed above, this could cause the empirical bias to be somewhat underestimated.

Finally, as mentioned in Section II-C, the simple clipping model we applied to the measured "unclipped" power is an imperfect approximation of how real inverters behave. In particular, the clipping point of the inverter from one of the commercial systems is known to decrease slightly with increasing temperature.

The PVsyst models have some untracked uncertainty as well. Although we rescaled the "unclipped" power to have zero bias compared with the nominal PVsyst model, varying the model's ILR might introduce some small bias from other model nonlinearities. For example, inverter efficiency is generally not constant over the inverter's power range and increasing the ILR will tend to shift the distribution of operating points towards a different efficiency at the higher end of the efficiency curve.

The similar model biases shown in Fig. 4 and listed in Table III indicate that subhourly clipping model bias is not necessarily specific to PVsyst or the commercial weather dataset used in the primary analysis. This is consistent with expectations; any hourly performance model without some kind of subhourly adjustment might be expected to suffer from this bias, and as satellite-based datasets tend to draw from the same underlying data sources, one might expect little difference there as well.

Solar Variability Zones could be an intuitive reference for subhourly clipping error, but the Commercial Plant 2 site serves as a notable counterexample: its solar variability zone classification [14] is "low" variability (the scale includes one lower classification, "very low") but still exhibits a PVsyst bias of 1.9% at an ILR of 1.5. It is also worth noting that the Solar Variability Zones were developed based on hourly, 10 km gridded NSRDB data and only seven sites in the Western continental US (plus two in Hawaii). Regions with different climates (e.g., higher frequencies of clouds that are smaller than 10 km) may not be as well represented in this dataset.

V. Conclusion

We have shown that energy models using the approach that industry developers and financiers use for financing systems suffer from a material positive bias that grows with increasing inverter loading ratio. Especially for large systems, this bias is a close match to an empirical subhourly clipping bias determined from real-world system data, as well as what previous model-based studies have predicted. We also show that a similar bias is recreated using an alternative hourly simulation tool and weather dataset, indicating that this bias is not unique to a specific model or dataset.

Acknowledgment

The authors wish to acknowledge and thank our colleagues at NREL, NIST, EPRI, and our commercial partner for providing the PV datasets used in this work, as well as Johan Kemnitz for providing the baseline model for the NIST system.

This work was authored in part by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Numbers 34348. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

References

[1] M. Bolinger, J. Seel, C. Warner, and D. Robson, "Utility-scale solar, 2021 edition: Empirical trends in deployment, technology, cost, performance, PPA pricing, and value in the United States," Lawrence Berkely National Laboratory, Tech. Rep., 10 2021. [Online]. Available: https://www.osti.gov/biblio/1823604

[2] S. Ransome and P. Funtan, "Why hourly averaged measurement data is insufficient to model PV system performance accurately," in *20th European Photovoltaic Solar Energy Conference*, 2005.

[3] C. W. Hansen, J. S. Stein, and D. Riley, "Effect of time scale on analysis of PV system performance," Sandia National Lab, Albuquerque, NM, Tech. Rep. SAND2012-1099, 2012.

[4] D. A. Bowersox and S. M. MacAlpine, "Predicting subhourly clipping losses for utility-scale PV systems," in *2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)*. IEEE, Jun. 2021. [Online]. Available: https://doi.org/10.1109/pvsc43889.2021.9518956

[5] D. Cormode, N. Croft, R. Hamilton, and S. Kottmer, "A method for error compensation of modeled annual energy production estimates introduced by intra-hour irradiance variability at PV power plants with a high DC to AC ratio," in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*. IEEE, Jun. 2019. [Online]. Available: https://doi.org/10.1109/pvsc40753.2019.8981206

[6] K. Bradford, R. Walker, D. Moon, and M. Ibanez, "A regression model to correct for intra-hourly irradiance variability bias in solar energy models," in *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. IEEE, Jun. 2020. [Online]. Available: https://doi.org/10.1109/pvsc45281.2020.9300613

[7] K. Anderson and K. Perry, "Estimating subhourly inverter clipping loss from satellite-derived irradiance data," in *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. IEEE, Jun. 2020. [Online]. Available: https://doi.org/10.1109/pvsc45281.2020.9300750

[8] A. Parikh, K. Perry, K. Anderson, W. B. Hobbs, R. Kharait, and M. A. Mikofski, "Validation of subhourly clipping loss error corrections," in *2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)*. IEEE, Jun. 2021. [Online]. Available: https://doi.org/10.1109/pvsc43889.2021.9518564

[9] R. Kharait, S. Raju, A. Parikh, M. A. Mikofski, and J. Newmiller, "Energy yield and clipping loss corrections for hourly inputs in climates with solar variability," in *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. IEEE, Jun. 2020. [Online]. Available: https://doi.org/10.1109/pvsc45281.2020.9300911

[10] A. Berlinsky, M. Reusser, and S. Vaishnav, "Modeling subhourly clipping losses using available irradiance and power data," in *Proceedings of the PV Reliability Workshop, February 22-26, 2021*, no. NREL/CP-5K00-80055, 6 2021. [Online]. Available: https://www.osti.gov/biblio/1797569

[11] C. Bordonaro, "A study of sub-hourly clipping losses for various solar PV plant designs across the United States," in *Proceedings of the PV Reliability Workshop, February 22-26, 2021*, no. NREL/CP-5K00-80055, 6 2021. [Online]. Available: https://www.osti.gov/biblio/1797569

[12] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, "The national solar radiation data base (NSRDB)," *Renewable and Sustainable Energy Reviews*, vol. 89, pp. 51–60, 2018. [Online]. Available: https://doi.org/10.1016/j.rser.2018.03.003

[13] C. W. Hansen, W. F. Holmgren, A. Tuohy, J. Sharp, A. T. Lorenzo, L. J. Boeman, and A. Golnas, "The solar forecast arbiter: An open source evaluation framework for solar forecasting," in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, 2019, pp. 2452–2457. [Online]. Available: https://doi.org/10.1109/PVSC40753.2019.8980713

[14] M. Lave, R. J. Broderick, and M. J. Reno, "Solar variability zones: Satellite-derived zones that represent high-frequency ground variability," *Solar Energy*, vol. 151, pp. 119–128, 2017. [Online]. Available: https://doi.org/10.1016/j.solener.2017.05.005

[15] C. Deline, K. Perry, M. Deceglie, M. Muller, W. Sekulic, and D. Jordan, "Photovoltaic data acquisition (PVDAQ) public datasets," Dec 2021. [Online]. Available: https://doi.org/10.25984/1846021

[16] M. Boyd, T. Chen, and B. Dougherty. (2017) NIST campus photovoltaic (PV) arrays and weather station data sets. National Institute of Standards and Technology. U.S. Department of Commerce, Washington, D.C. [Online]. Available: https://doi.org/10.18434/M3S67G

[17] M. Boyd, "Performance data from the NIST photovoltaic arrays and weather station," *Journal of Research of the National Institute of Standards and Technology*, vol. 122, Nov. 2017. [Online]. Available: https://doi.org/10.6028/jres.122.040

[18] J. O. Allen and W. B. Hobbs, "Effect of time-averaging on PV production estimates on systems with high DC to AC ratios," Sandia PV Performance Modeling Collaborative, 2015.

[19] J. O. Allen, W. B. Hobbs, and M. Bolen, "The effect of short-term inverter saturation on PV performance modeling," Sandia PV Performance Modeling Collaborative, 2018. [Online]. Available: https://pvpmc.sandia.gov/download/6707/

[20] ——, "Predicting the effect of short-term inverter saturation on PV performance modeling," Sandia PV Performance Modeling Collaborative, 2019. [Online]. Available: https://pvpmc.sandia.gov/download/7415/

[21] S. M. Wilcox, "National solar radiation database 1991-2010 update: User's manual," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2012.

[22] A. Mermoud and B. Wittmer, "PVSYST user's manual," PVsyst SA, 2014.

[23] A. P. Dobos, "PVWatts version 5 manual," National Renewable Energy Laboratory, Golden, CO, Tech. Rep. NREL/TP-6A20-62641, 2014. [Online]. Available: https://doi.org/10.2172/1158421

[24] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib python: a python package for modeling solar energy systems," *Journal of Open Source Software*, vol. 3, no. 29, p. 884, 2018. [Online]. Available: http://doi.org/10.21105/joss.00884

[25] W. Holmgren *et al.*, "pvlib/pvlib-python: v0.9.1," Mar. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6395177

[26] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, "From irradiance to output power fluctuations: the PV plant as a low pass filter," *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 5, pp. 505–510, 2011. [Online]. Available: https://doi.org/10.1002/pip.1063