# Plant Clipping Bias Due to Spatial Distribution of Inverters and AC Overbuild: Preliminary Investigation

Joseph Ranalli[1] and William B. Hobbs[2]

[1] Penn State Hazleton, Hazleton, PA 18202, USA
[2] Southern Company, Birmingham, AL, 35203, USA

***Abstract* — This study investigated the bias in modeled generation for plants with a level of ac overbuild, that is greater aggregate inverter rated capacity than the plant's grid export limit. The bias was simulated using operational data from two utility scale plants (20 MW and 200 MW). We found daily biases as high as 5% for a single variable day, while clear days experienced virtually no bias. Annual bias for the plants tested is estimated between 0.3-0.7%. In the annual case, the larger 200 MW plant was observed to have a lower bias for comparable overbuilding ratios, which we attribute to the larger inverter footprints for this plant. The use of operational data proved somewhat limiting, as it prevented a fully parameterized investigation of the issues. Further study is warranted to evaluate these effects on plants with different sizes and subject to different climatological conditions.**

## I. Introduction

Developers and owners of photovoltaic (PV) plants rely on data and models to monitor and predict the operation of their facilities. The ability to accurately represent plant production is crucial, as greater uncertainty in models leads to more challenges assessing risk. Research efforts on solar plant modeling aim to improve the ability of models to translate data to plant generation estimates to increase confidence in results and to facilitate optimal decision making.

Issues addressed in the literature often stem from the mismatch between the resolutions of input data and the continuous nature of plant operation. For example, studies have developed corrections for subhourly clipping mismatch, which arises when hourly irradiance data does not accurately represent instantaneous high irradiance events that lead to exceedance of inverter capacity [1], [2], [3]. These temporal mismatch effects cause underestimation of clipping loss (i.e. overestimation of plant generation). Typical satellite-derived input data may have resolutions of 15, 30 or 60 minutes [1], [2], which is not sufficient to reflect the occurrence of cloud-edge-enhancement or other transient events during periods of high irradiance variability. The correction factors for temporal bias described by the literature for practical cases are typically 1-2% percent of annual energy.

In addition to temporal mismatch, mismatches also occur due to differences in spatial resolution. Data acquired from a measurement station represent the irradiance at the scale of the sensor (essentially a point, relative to the large spatial scale of a plant). Using a small sensor to model spatially distributed generation has the potential to lead to overprediction of the losses (underprediction of generation) due to the fact that localized sensor measurements are unable to represent the degree to which spatial aggregation actually reduces variability[4]. Prior investigations have attempted to use models of spatial variability to account for the bias introduced by these effects [5]. As in the temporal mismatch case, the potential exists for the degree of overbuilding to produce variation in results, though in this case, the issue arises with the potential for ac overbuilding (i.e. the cumulative sum of all inverter capacities exceeds the plant's grid export capacity).

This study investigates these spatial scale mismatch effects by considering the difference between inverter-level and plant-level generation, using data from an operating PV plant.

## II. Methodology

We investigated one year (calendar year 2022) of inverter power data from two operational PV plants, both in the United States. One was a 20-MW with 25 inverters rated at 800 kW each, and the second a 200-MW plant with 112 inverters rated around 1800 kW. The 20 MW plant is relatively small compared to many plants being built today, which are frequently 200 MW or larger [13]. Data from the 20 MW plant is not publicly available, but was used in prior work by the authors [6]. The 200 MW plant is publicly available from the American-Made Solar Data Bounty Prize dataset (identifier 7333) [7]. Data for both plants corresponds to the 2022 calendar year.

In both cases, data consist of time series of generated power production for each of the inverters within the plant, measured in kilowatts. The sampling of the acquired data was irregular due to methodologies designed to save data storage space, with recordings occurring only when a minimum threshold of change was exceeded, but with a shortest sample interval of 10 seconds. For the purpose of this study, data were interpolated to produce a consistent 10-second resolution time series for each inverter.

### *A. Quality Control*

The data were filtered using a preliminary quality control workflow, utilizing the *pvanalytics* python package [8] and based upon their example PV Fleets pipeline [9].

First, each individual inverter time series was masked to remove erroneous data. Negative values and low generation (less than about 6% of rated inverter power, 50 kW and 125 kW respectively) periods were removed, along with those occurring overnight. Stale data was removed using a 50-point rolling window. Daily and hourly periods with a mean less than 10% of the annual mean were removed. Finally, the *pvanalytics zscore* function was used to remove outliers on a statistical basis. After these quality control checks, inverter data was removed for whole hours during which an inverter failed quality control for any part of that hour. Following the inverter-level quality control, a plant-level quality check was applied, and any remaining hour periods during which fewer inverters than a threshold (40%, or 10 inverters and 45 inverters for the 20 MW and 200 MW respective plants) were removed.

Normalization was applied to the remaining data such that all inverters were scaled to have a common upper bound value (in this case, the same 95$^{th}$ percentile of generation) for each day of measurements. While the inverters all had approximately equal capacity, this step ensured that any variations in individual inverter performance throughout the year were eliminated.

### *B. Representing Spatially Distribution Impacts*

We represented the impact of spatial variability on plant-level clipping by comparing the daily energy produced by the plant in two configurations, one where clipping occurred on the inverters and one where clipping occurred only at the plant level.

Power for configuration A, including the effects of spatial smoothing, was simulated by performing clipping at the plant level. We summed the measured power outputs of all inverters, linearly scaling up to account for any inverters with invalid data. The clipping was applied to this aggregate signal. The daily energy, $E_A$, for day $j$ was obtained by integrating the power over the course of the day.

Configuration B was intended to represent data that was not subject to spatial smoothing. Thus, the power for each individual inverter $i$ was clipped at a level corresponding to 1/25$^{th}$ or 1/112$^{th}$ of the overall plant level (i.e. the inverse number of inverters for that plant). As in the case of configuration A, the power from each clipped inverter was summed and the result was integrated over the day to yield the daily energy, $E_B$.

Discrepancies between the energy production under configurations A and B represent the degree to which spatial variability in the generation results in differential response to clipping. We represent this as bias as the percent difference between $E_A$ and $E_B$. Ideally, since configuration A represents the aggregate of the individual inverters, under conditions where no inverter clipping occurs, we expect the bias between $E_A$ and $E_B$ across all inverters to be zero. On the other hand, a positive value for the bias demonstrates the degree to which a plant with ac overbuild experiences higher generation by virtue of allowing individual inverters to exceed their share of the plant's generation on a transient basis.

$$bias = \frac{E_A - E_B}{E_A} \quad (1)$$

Clipping was simulated by applying an artificial maximum value to the instantaneous values of the individual time series $P_i(t)$ or the aggregate $\sum_{i=1}^{n} P_i(t)$ for calculation of $E_B$ and $E_A$ respectively. This approach is conceptually similar to the artificial clipping applied to inverter-level data in [10] to explore subhourly inverter clipping and related phenomena. When clipping is introduced here, the magnitude of the bias can be interpreted to understand the spatial aggregation effects. A positive bias would indicate that the output from the plant is higher when clipping the aggregated signal than when clipping the individual inverter signals and is indicative of smoothing of the aggregate time series due to spatial variability. Note that this bias impact is opposite from that seen in some work related to inverter-level clipping due to temporal mismatch, such as [3] and [10], where a positive bias implies the plant's underperformance relative to the higher resolution case.

During time periods where the plant is subject to intermittent clouds, while a single scaled inverter experiencing clear sky might alone exceed its proportional share of the plant generation at any given time, which would result in the inverter signal being clipped without any ac overbuild. However, it is unlikely all inverters experience the clear sky simultaneously. Thus, when aggregating the plant together, the amplitude or aggregate fluctuations are likely to be less extreme [4] implying that a higher mean could be achieved without significantly increasing the probability of clipping, resulting in a higher daily ratio of energy production.

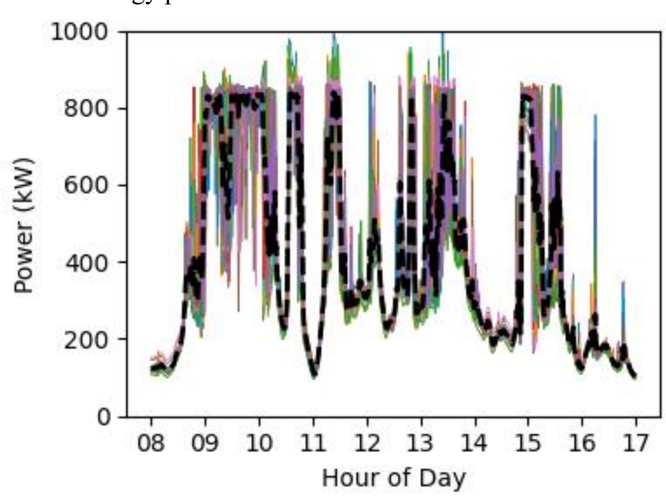


Fig. 1. Time series for the intermittent day for the 20 MW plant. Colored lines show individual inverters, black line shows the mean.

## III. RESULTS & DISCUSSION

We initially investigated the impacts during two days for each plant, one low-variability clear day and a high-variability day experiencing intermittent cloudiness (example shown in Fig. 1). We computed the bias as a function of plant clipping level for each of these days, shown in Fig. 2 for both plants. We assessed the variability using the variability score [11]. Variability score was calculated as the median (across inverters) daily variability score based on the individual inverter time series, with the inverters scaled to a max value of 1000 kW for consistency in comparison. For clear days, variability scores were 0.3 and 0.1 for the 20MW and 200MW plant respectively. For the variable days represented here, the scores were 7.3 and 6.0 for their respective plants.

For both plants, the clear day bias is very low, and exhibits very little variation with clipping level. This is consistent with the hypothesis that under clear conditions there is not a significant influence of spatial variability that could lead to mismatch between clipping individual inverters and the plant as a whole.

Conversely, for the days with high variability, variation in the bias is observed. A maximum bias of around 2.5% occurred for the 20 MW plant at an overbuild clipping ratio of around 1.4. The 200 MW plant observed a higher maximum (around 5.0%) around the same overbuild ratio. Both of these biases imply that when averaging over the plant, there is sufficient lag between cloud shadows passing over different inverter footprints to reduce the incidence of clipping when aggregating inverter signals into the plant output.

Fig. 3 shows a similar visualization on an annual basis for these plants and year of data combination. For the 20 MW plant, the annualized bias is seen as a maximum of around 0.7% with a relatively high ac overbuild ratio of 1.6. The larger plant exhibited a maximum around 0.3% around the 1.6 overbuild, with an additional spike of similar magnitude around an overbuild of 2.5.

## III. DISCUSSION & CONCLUSIONS

In both the cases of individual variable days and annual generation, biases were observed that indicate ac overbuild is capable of leading to increased generation as compared to matching inverter and plant capacity precisely. These biases were not observed for clear days, which is consistent with predictions based on the hypothesized spatial variability basis for the bias. The exact degree of annual bias observed for a plant would be expected to depend on the actual balance of clear and variable days its climate experiences. For example, the 20 MW plant is located in a Solar Variability Zone [12] of “very low”, the lowest on the scale, so the observed annual result could represent something close to a lower bound for impacts.

We did observe a difference in the daily biases observed between the two plants. Both plants exhibited daily bias associated with variability, though significant variation in maximum bias varied depending on the day chosen. Generally speaking, high variability resulted in observed bias, but there was significant fluctuation observed in the actual maximum bias that was not uniquely correlated with variability score (a positive correlation coefficient between maximum hourly bias and variability score of 0.22 was observed). Intuitively, larger projects with equivalent inverter capacities could be expected to experience a higher bias under variable conditions, because the overall benefit of the post-inverter smoothing would be predicted to increase due to the wider spatial dispersion of the plant. However, confounding variables in the methodology here, including the baseline inverter size and an inability to independently control the operational data prevent the present a conclusive determination relative to this hypothesis.

In the annual results, the larger 200 MW plant consistently experienced lower biases. Generalized interpretation of this result is hampered, both because of climatological differences in the sites and because the two plants had disparate inverter capacities (and thereby spatial footprints). So, a direct comparison between these results is not possible; a larger inverter footprint corresponds to a greater degree of smoothing that occurs prior to the clipping [4], meaning that the occurrence of individual inverter exceedance events may be intrinsically decreased.

While our results indicate that there is a benefit to ac overbuilding that can be observed from operational data, there are several important limitations to understand when interpreting this work. By utilizing real operational data, it was impossible to completely eliminate the occurrence of actual inverter clipping that was present in the data, particularly for the annual result. While the individual clear and variable days were chosen to avoid the incidence of clipping as much as possible, clipping events are certainly present within the annualized representation of data, which might mask the ability to model the actual degree of long-term bias for these plants. Additionally, we observed that overbuild ratio corresponding to the peak bias varied depending on the mean production during a given hour, resulting from the formulation of the bias in Eq. 1. Highly variable periods with a lower mean of generation exhibited high bias at a greater overbuilding ratio, because the clipping level did not begin to impact the plant until reaching a lower overall threshold. While this could in principle be addressed by changing the artificial scaling in the calculation of the bias ratio, doing so could also yield potentially erroneous results by representing an unrealistic scaled ramp rate.

These variables may be able to be eliminated further study utilizing a simulated generation facility with simulated spatial variability to allow for a more consistent parameterization of how these effects result in plant bias. We believe this would allow researchers to develop a more complete understanding of the factors leading to ac overbuilding bias and to allow more concrete recommendations to be made for developers and operators.

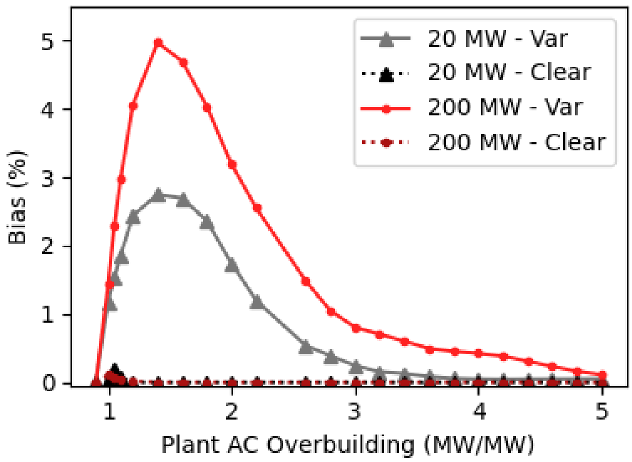


Fig. 2. Variation in bias with plant ac clipping level. Lines compare sample clear and high-variability days for each plant.

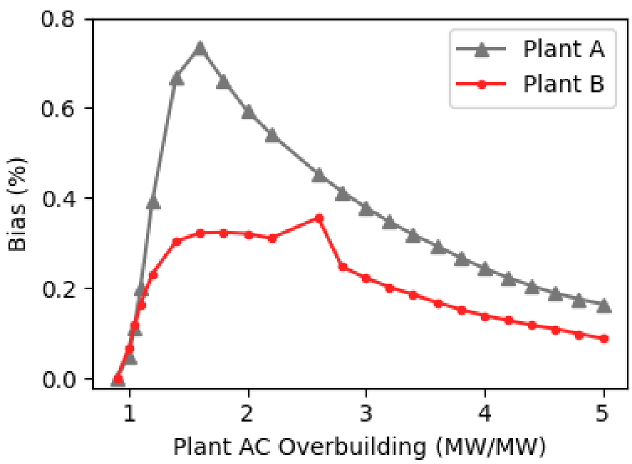


Fig. 3. Variation in bias with plant ac overbuild on annual basis.

## REFERENCES


[1] K. Anderson and K. Perry, "Estimating Subhourly Inverter Clipping Loss From Satellite-Derived Irradiance Data," in *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*, Jun. 2020, pp. 1433–1438. doi: 10.1109/PVSC45281.2020.9300750.

[2] M. A. Mikofski, W. F. Holmgren, J. Newmiller, and R. Kharait, "Effects of Solar Resource Sampling Rate and Averaging Interval on Hourly Modeling Errors," *IEEE J. Photovolt.*, vol. 13, no. 2, pp. 202–207, Mar. 2023, doi: 10.1109/JPHOTOV.2023.3238512.

[3] W. B. Hobbs, C. L. Black, W. F. Holmgren, and K. S. Anderson, "Evaluation of Irradiance Variability Adjustments for Subhourly Clipping Correction," in *2023 IEEE 50th Photovoltaic Specialists Conference (PVSC)*, Jun. 2023, pp. 1–4. doi: 10.1109/PVSC48320.2023.10359541.

[4] M. Lave, J. Kleissl, and J. S. Stein, "A Wavelet-Based Variability Model (WVM) for Solar PV Power Plants," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 501–509, Apr. 2013, doi: 10.1109/TSTE.2012.2205716.

[5] C. Hayes, A. Parikh, M. Mikofski, and R. Kharait, "Sensitivity of Sub-Hourly Modeling Error to Project Size," in *2022 IEEE 49th Photovoltaics Specialists Conference (PVSC)*, Jun. 2022, pp. 0858–0858. doi: 10.1109/PVSC48317.2022.9938848.

[6] Joseph Ranalli and William B. Hobbs, "PV Plant Equipment Labels and Layouts can be Validated by Analyzing Cloud Motion in Existing Plant Measurements," *J. Photovolt.*, Under Review.

[7] C. Deline, K. Perry, M. Deceglie, M. Muller, W. Sekulic, and D. Jordan, "Photovoltaic Data Acquisition (PVDAQ) Public Datasets." 2021. doi: 10.25984/1846021.

[8] K. Perry, W. Vining, K. Anderson, M. Muller, and C. Hansen, "PVAnalytics: A Python Package for Automated Processing of Solar Time Series Data," National Renewable Energy Lab. (NREL), Golden, CO (United States), NREL/PR-5K00-83824, Sep. 2022. Accessed: Apr. 11, 2024. [Online]. Available: https://www.osti.gov/biblio/1887283

[9] "PV Fleets QA Process: Power — PVAnalytics documentation." Accessed: Jan. 08, 2025. [Online]. Available: https://pvanalytics.readthedocs.io/en/stable/generated/gallery/pvfleets-qa-pipeline/pvfleets-power-qa.html

[10] K. S. Anderson, W. B. Hobbs, W. F. Holmgren, K. R. Perry, M. A. Mikofski, and R. A. Kharait, "The Effect of Inverter Loading Ratio on Energy Estimate Bias," in *2022 IEEE 49th Photovoltaics Specialists Conference (PVSC)*, Jun. 2022, pp. 0714–0720. doi: 10.1109/PVSC48317.2022.9938632.

[11] M. Lave, M. J. Reno, and R. J. Broderick, "Characterizing local high-frequency solar variability and its impact to distribution studies," *Sol. Energy*, vol. 118, pp. 327–337, 2015.

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

[13] J. Seel *et al.*, "Utility-Scale Solar, 2024 Edition: Data File," Lawrence Berkeley National Laboratory, Oct. 2004. [Online]. Available: https://emp.lbl.gov/publications/utility-scale-solar-2024-edition