Validation of Subhourly Clipping Loss Error Corrections

Preprint

Abhishek Parikh,[1] Kirsten Perry,[2] Kevin Anderson,[2] William B. Hobbs,[3] Rounak Kharait,[1] and Mark A. Mikofski[1]

1 DNV
2 National Renewable Energy Laboratory
3 Southern Company

Validation of Subhourly Clipping Loss Error Corrections

Preprint

Abhishek Parikh,[1] Kirsten Perry,[2] Kevin Anderson,[2]
William B. Hobbs,[3] Rounak Kharait,[1] and Mark A. Mikofski[1]

1 DNV
2 National Renewable Energy Laboratory
3 Southern Company

NOTICE

Validation of Subhourly Clipping Loss Error Corrections

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

Abstract—**Under-performance of solar PV systems is an important issue that increases risks for stakeholders, including developers, investors and operators. Recently some attention has focused on underestimation of inverter clipping losses as a possible source of over-prediction where sub-hourly solar variability is high. Several models and data sets have been analyzed over the past few years, with the aim of quantifying, predicting, and correcting underestimated clipping loss errors for systems with high DC/AC ratio and solar variability. In this research, we apply a machine learning model developed at NREL to two physical PV systems, to correct for subhourly clipping losses. For each system, we compare overall AC power output for the model taken at 1-minute intervals to AC power output taken at 1-hour intervals with the addition of the subhourly clipping correction. Our findings consistently show that the addition of the clipping loss correction lead to a reduction in mean bias error of 0.8% and 1.2% for systems A and B, respectively, with no additional filtering applied. When examining high solar variability periods where clipping is more pronounced, system A and B experienced a 1.8% and 2.7% reduction in mean bias error, respectively, when the clipping correction was applied.**

Index Terms—**inverter, clipping, solar, irradiance, variability, performance, modeling, TMY**

I. Introduction

In its report "2020 Solar Risk Assessment" [1], kWh Analytics warned that systematic underproduction across the industry exposes investors to increased risk. DNV found that a sample of 39 projects from 2019 were under-performing compared to their pre-construction energy assessments by 3% on average, as shown in Fig. 1. NextEra estimates that sub-hourly solar resource variability can affect actual energy production by approximately 1-4% [2]. Cormode, *et al.*, explained that energy assessments based on hourly weather, like typical meteorological year (TMY) data sets, underestimate inverter clipping losses by up to 5% for projects at an annual level with intra-hourly solar variability, especially for high DC/AC ratios [3]. Allen, *et al.*, used measured 1-minute weather data to model clipping loss underestimation from using hourly data, in the Southeast US [4] and at ten sites across the country [5], for a range of DC/AC ratios and, for a subset of sites, across multiple years. They estimated similar errors, with some sites seeing errors of 2 to over 4% for DC/AC ratios of 1.4 to 2.0, respectively. This is demonstrated in Fig. 2, which shows irradiance and power output simulated at 1-minute and 1-hour resolution in the top and bottom panels, respectively. On

July 10th, there is little intra-hour variability, indicative of a mostly clear day where hourly and sub-hourly simulations are in close agreement. In contrast, July 13th has much higher solar variability, but no clipping is observed in the 1-hour output. However, the 1-minute resolution shows intermittent clipping all day long, indicating that not all of the available irradiance is used by the system, and the hourly simulations underestimate inverter clipping [6].
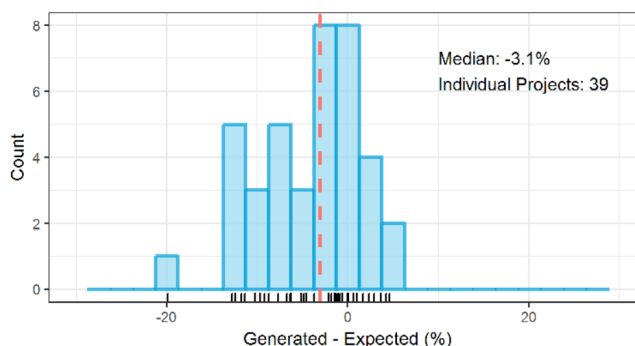
Fig. 1. Project-average validation results for solar energy assessments. Each project-year was adjusted for interannual variability by scaling production by the ratio of TGY to historical monthly insolation.

Several methods have been proposed to correct clipping loss error in hourly predictions [2], [3], [6]–[9], but only NextEra validated their method with operational power data [2].

In this report, we adapt the NREL machine learning-based model [7] to predict clipping loss corrections and apply these corrections to hourly energy assessments. The adjusted energy assessment is then validated against operational data from the same system and we report on the average and distribution of validation errors. We have obtained data from two operating PV systems in the southeast US, with DC/AC ratios around 1.4. In this paper, we will outline our method and present model validation results for these two systems.

II. Methods

A. Model

The machine learning (ML) model developed at NREL has already been described in detail [7]. This model was adapted for the purpose of this research. A black-box XGBoost

1

Fig. 2. Irradiance input from the NIST test-bed weather station in Gaithersburg, MD, and predicted energy output using SolarFarmer at time-resolution of 1-minute, *top panel*, and 1-hour, *bottom panel*. The black dashed line in both panels shows the inverter-rated power, 260-kW.

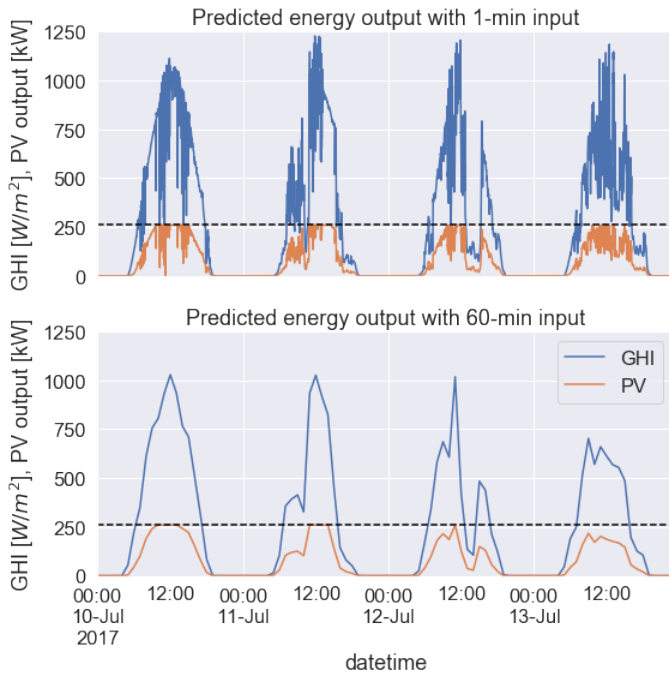|	System A	System B
Array type	Tracking	Tracking
Array axis azimuth	180°	180°
DC capacity	27 MW	43 MW
AC capacity	19 MW	30 MW
DC/AC Ratio	1.42	1.43
Ground coverage ratio	0.3	0.3

model was generated, with satellite data inputs, as well as features derived from these fields. In an extension of the model described in [7], system DC/AC ratio and AC power outputs from the associated SAM model were added as continuous variable features. Specific model input parameters included the following:

- Min-max normalized POA, GHI, clearsky POA, clearsky GHI, and cell temperature
- First-order differenced normalized POA
- Difference between Clearsky POA and POA
- System DC/AC ratio
- Modeled AC power (taken from the SAM simulation)

The model was trained using high quality irradiance measurements from data sources including SURFRAD, SOLRAD, UOSRML, MIDC, and BSRN [10] [11] and 1-minute to 30-minute clipping loss errors predicted using SAM [12]. The resulting trained model was then used to make predictions at the operational systems.

B. Systems

For the validation, two operational systems in Georgia were chosen. The system parameters of the two solar systems used for validation are provided in Table I. These systems were selected for their high DC/AC ratio and 1-minute data sampling frequency.

C. Validation

A challenge in validating the clipping loss correction model is the lack of quality high-frequency irradiance measurements

that are co-located with operational PV systems. Therefore this model differs from traditional approaches for model training and validation because it relies on high-quality ground station data measurements and SAM-simulated data to train the model, but uses operational data from different systems for validation. So no holdout methods are required in validating this model because the training and validation data sets are already independent. The downside of this method is that any systematic biases in SAM are introduced into the clipping loss correction model. In future work, if a large population of high-frequency irradiance and co-located operational data are obtained, the model can be retrained holding out a subset of the data for validation. This procedure would have the benefit of incorporating any physical effects that are not captured by SAM.

Quality of operational datasets is vital to avoid disruption of validation. Data points in the operational dataset related to plant sub-performance instances due to sources other than subhourly clipping error are considered to be outliers from the validation's standpoint and need to be removed. However, detecting sub-performance from inverter output is challenging. For this validation effort, the bias change is presented for both unfiltered and filtered data points. For filtering, data points that met the following criteria were retained:

- The difference between the measured, on-site POA and the modeled POA calculated using the NSRDB [13] data is less than 5%
- The difference between actual power and simulated power of a particular timestep is less than 25%.

For system A and system B, 60.5% and 55.1% of the data is retained after applying the data filter, respectively.

In addition, high clipping errors occur during periods where the plant observes DC clipping (usually corresponding to higher irradiance), and high fluctuations in the irradiance. For the purpose of this validation, the fluctuations are quantified by calculating the variability index based on [14]. The variability index is normally defined on GHI. However, since only on-site measured POA was available, the variability index for this study was calculated and aggregated at 30-min intervals using the on-site POA measured at a 1-min frequency. The clearsky POA was calculated using the Perez implementation avaiable in pvlib [15]. Since the ML model should be correcting only those periods with high fluctuations, the change is split into sets of low and high variability index. Periods are divided into three sets based on their variability index value:

2

- Low variability index (≤ 10)
- Medium variability index (> 10 & ≤ 50).
- High variability index (> 50)

Table. II shows the composition of the on-site data based on the variability index. For this validation, two complete years of operational data from system A are used, and one complete year of operational data from system B is used.

	System A	System B
Low variability index (≤ 10)	62%	63%
Medium variability index (> 10 & ≤ 50)	21%	20%
High variability index (> 50)	17%	17%

The bias in model results is used to measure the reduction in clipping loss errors from the hourly energy assessment, and to identify if there are any unknown factors that are systematically affecting the results. Bias and mean bias error (MBE), the two metrics used to evaluate model performance in this study, are described in detail below.

- **Bias**: Delta between the corrected energy assessment predictions and the measured operational data. Positive bias means over-predicted energy output:

$$\Delta E = E_{\text{predicted}} - E_{\text{measured}} \quad (1)$$

where E is the energy output.

- **Mean bias error (MBE)**: Average of the bias:

$$MBE = \frac{\sum_{n=1}^{N} \Delta E_n}{N} \quad (2)$$

Where n is a single timestep, and N is the total number of time intervals. Please note that a low MBE might hide seasonal or diurnal bias

III. RESULTS

30-minute NSRDB data from each system was fed into the ML model to make clipping error predictions. Model input parameters are previously described in the Methods section.

Time steps with a higher variability index should have higher clipping errors. Our results, shown in Fig. 3, support this statement. Fig. 3 displays a heat map of average variability index, as well as the average clipping loss error correction predicted for system A. Generally, higher variability periods coincide with higher clipping loss corrections, usually occurring around the middle of the day, and during summer months.

Fig. 4 shows the actual power of an inverter against the actual POA. At a higher variability index, the power is lower and corresponds to higher model bias.

For the energy model, 30-minute NSRDB data was used to generate a time series of energy estimates for system A and system B. Then, these energy estimates were corrected using the clipping loss error corrections predicted by the ML model. Table. III summarizes the mean bias errors in the energy models before and after applying the correction, for systems A and B respectively. The overall shift in the bias
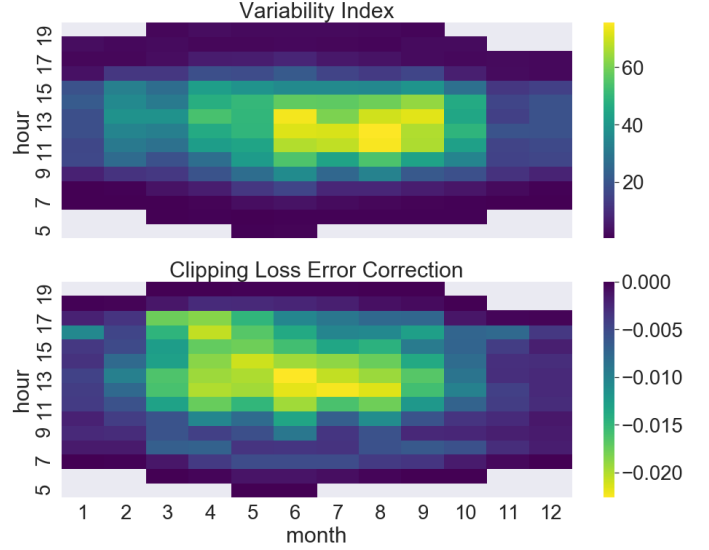

Fig. 3. The first sub-figure shows the average variability index calculated at system A. The variability index is calculated and aggregated at a 30-min interval using on-site POA measured at a 1-min time resolution. In the second sub-figure, the average clipping error correction calculated for system A using the ML model is shown. The model predicts the clipping error correction factor for each 30-min interval.
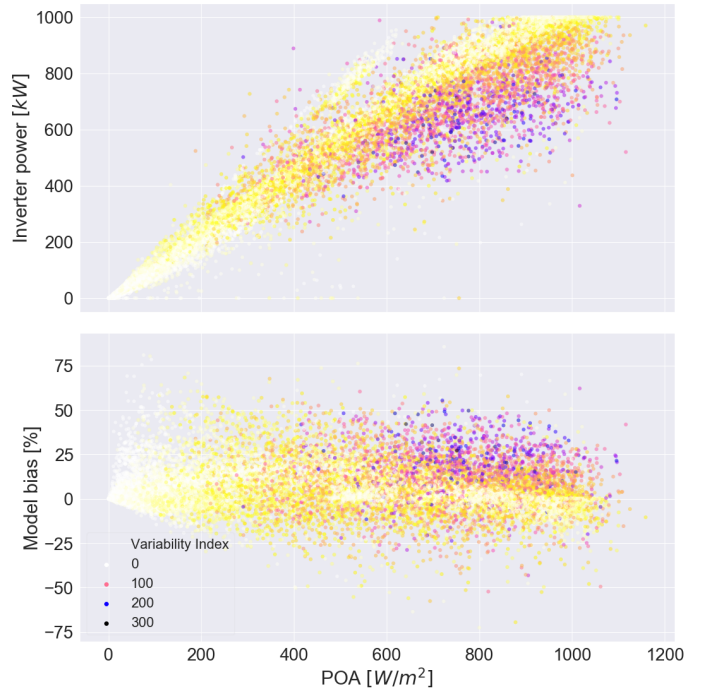

Fig. 4. Inverter power and bias in modeled power for a 1000 kW rated inverter at system A. Lower power and higher model bias are observed at a higher variability index.

3

is small for each time series. As indicated by Fig. 3 and Fig. 4, the magnitude of the clipping loss error correction should be high during time steps with a high variability index and low during time steps with a low variability index. Hence, the bias is explored based on the low, medium and high variability index categories described in the Validation section. Part (A) of Fig. 5 and Fig. 6 shows the cumulative distribution of the overall bias before and after correction. Part (B) of these figures shows the bias broken down by the variability index. Bias is high for the points with a high variability index, and low for points with a low variability index. Promisingly, applying the clipping correction factor for high solar variability periods where clipping is most prominent results in the largest reduction in MBE. For unfiltered system A high VI data, applying the clipping correction factor results in a 1.8% reduction in MBE; similarly, applying the clipping correction factor for the associated system B data results in a 2.7% reduction in MBE.

TABLE III
RESULTS SUMMARY BY SYSTEM

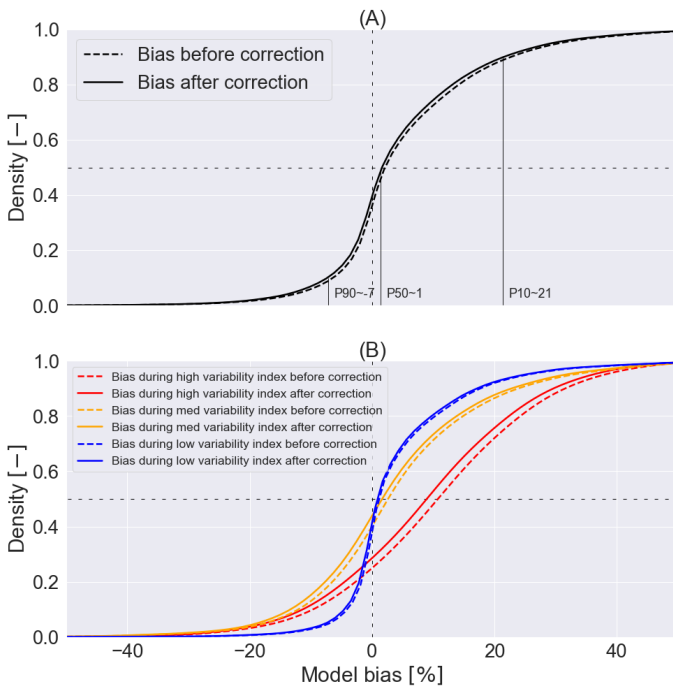| System	Metric	Before correction	After correction	Change [%]
System A	Overall MBE	5.4%	4.6%	-0.8
System A	MBE (High VI)	10.7%	8.9%	-1.8
System A	MBE (Medium VI)	4.3%	3.3%	-1.0
System A	MBE (Low VI)	4.4%	3.9%	-0.5
System B	Overall MBE	7.8%	6.6%	-1.2
System B	MBE (High VI)	10.4%	7.7%	-2.7
System B	MBE (Medium VI)	7.7%	6.1%	-1.6
System B	MBE (Low VI)	7.1%	6.4%	-0.7

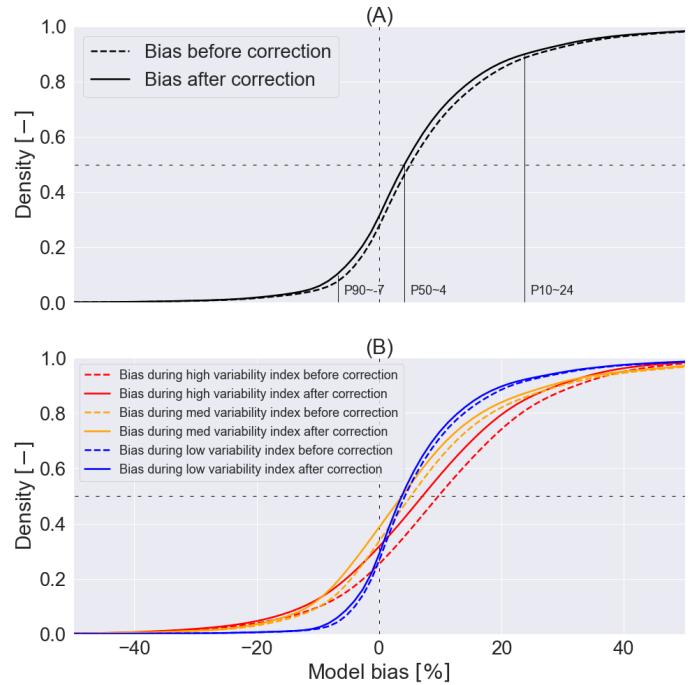Fig. 5. CDF of Model bias for system A before data filtering

Fig. 6. CDF of Model bias for system B before data filtering

The overall model bias in Fig. 5 and Fig. 6 can be attributed to many reasons. Factors like the difference in the resource measured on-site and in the NSRDB dataset, system-specific sub-performance, incorrect loss assumptions in the energy model, etc., impact the bias along with the clipping loss errors. However, for a given data point, it is unrealistic to have a bias of greater than 25% in power when the bias in POA is less than 5%. Consequently, points fitting this logic are removed and the bias change is recalculated. Table. IV summarizes the mean bias errors in the filtered data points of the energy models before and after applying the correction, for systems A and B respectively. Fig. 7 and Fig. 8 show the bias before and after the correction for this filtered dataset. Compared to the CDFs plotted for the unfiltered dataset, these plots show a much tighter distribution as the outliers are removed.

Results for the filtered data mirror the unfiltered data results, except with lower overall mean bias error values. Interestingly, the drop in mean bias error after applying the clipping error correction is generally consistent for both filtered and unfiltered data (see the "% Change" field in Tables III and IV for comparison). For example, the drop in overall MBE for unfiltered data is 0.8% vs. 0.5% for filtered data for system A. Similarly the overall MBE difference is 1.2% for unfiltered data and 0.8% for filtered data for system B.

IV. CONCLUSIONS

Over-prediction occurs in typical energy assessments that use hourly data if the systems have sub-hourly solar variability and high DC/AC, because clipping losses are underestimated. These errors have been quantified using a machine-learning model trained on high-frequency solar irradiance data and

4

System	Metric	Before Correction	After Correction	Change [%]
System A	Overall MBE	2.3%	1.8%	-0.5
System A	MBE (High VI)	10.7%	9.0%	-1.7
System A	MBE (Medium VI)	3.1%	2.2%	-0.9
System A	MBE (Low VI)	1.1%	0.8%	-0.3
System B	Overall MBE	3.6%	2.8%	-0.8
System B	MBE (High VI)	8.7%	5.9%	-2.8
System B	MBE (Medium VI)	4.4%	2.9%	-1.5
System B	MBE (Low VI)	2.8%	2.3%	-0.5

simulated operational data to create clipping loss error corrections. The corrections were applied to a typical hourly energy assessment of an existing operational system and compared to the measured output from the same system. Model bias was calculated before and after the clipping correction was applied, and the difference in MBE values was compared. Overall, the clipping correction outputs from the model reduced bias in the energy assessments by 0.8% and 1.2% for systems A and B, respectively, with no additional data filters applied. However, these bias reductions should be interpreted with two points in consideration: first, clipping loss error decreases with decreasing simulation timestep, so the corrections in this 30-minute validation are somewhat smaller than they would be in an hourly context; and second, based on the CDF comparisons, it seems likely that the model corrections do not fully eliminate the bias. Thus, the bias reductions reported in this paper likely underestimate the bias of a typical energy model. For high solar variability periods where clipping is most pronounced, the difference in MBE before and after applying the clipping correction is more pronounced. For unfiltered data with a high VI index, the MBE difference after applying the correction was 1.8% and 2.7% for systems A and B, respectively.

ACKNOWLEDGMENT

REFERENCES

[1] R. Matsui, J. Moore, C. Nunalee, L. Garcia da Fonseca, N. Vadhavkar, J. Crimmins, S. Dise, J. Ahmad, I. Gregory, J. Fort, and J. Corbitt, "Solar Risk Assessment: 2020 Quantitative Insights from the Industry Experts," kWh Analytics, Tech. Rep., 2020. [Online]. Available: https://www.kwhanalytics.com/solar-risk-assessment
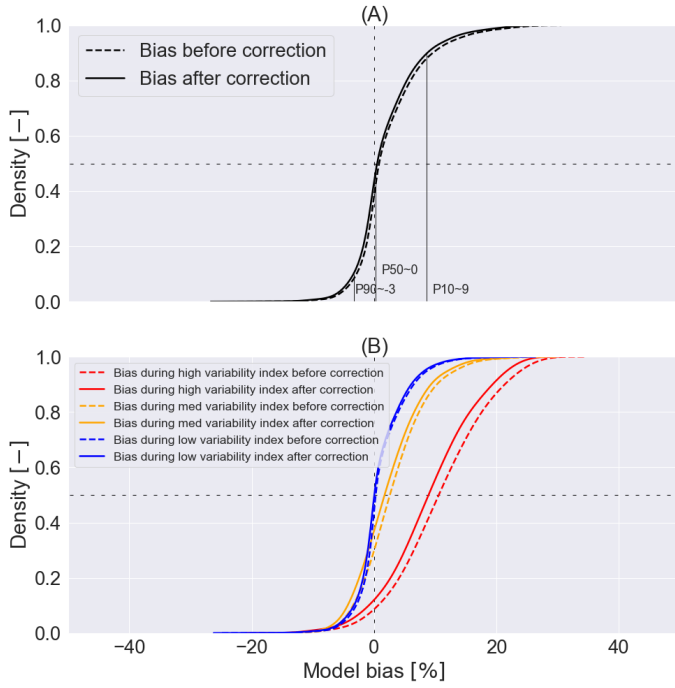
Fig. 7. CDF of Model bias for system A after data filtering

Fig. 8. CDF of Model bias for system B after data filtering

[2] 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, pp. 2679–2682. [Online]. Available: https://ieeexplore.ieee.org/document/9300613/

[3] 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, pp. 2293–2298. [Online]. Available: https://ieeexplore.ieee.org/document/8981206/

[4] 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.

[5] 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/

[6] 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, pp. 1330–1334. [Online]. Available: https://ieeexplore.ieee.org/document/9300911/

[7] K. Anderson and K. Perry, "Estimating Subhourly Inverter Clipping Loss From Satellite-Derived Irradiance Data," in *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*, no. June. IEEE, jun 2020, pp. 1433–1438. [Online]. Available: https://ieeexplore.ieee.org/document/9300750

[8] J. O. Allen, W. B. Hobbs, and M. Bolen, "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/

[9] M. Hofmann, S. Riechelmann, C. Crisosto, R. Mubarak, and G. Seckmeyer, "Improved synthesis of global irradiance with one-minute resolution for pv system simulations," *International Journal of Photoenergy*, vol. 2014, 2014.

[10] J. A. Augustine, J. J. DeLuisi, and C. N. Long, "SURFRAD - A national surface radiation budget network for atmospheric research," *Bulletin of the American Meteorological Society*, vol. 81, no. 10, pp. 2341–2357, 2000.

[11] B. B. Hicks, J. J. DeLuisi, and D. R. Matt, "The NOAA integrated surface irradiance study (ISIS)—a new surface radiation monitoring program," *Bulletin of the American Meteorological Society*, vol. 77, no. 12, pp. 2857–2864, 1996. [Online]. Available: https://doi.org/10.1175/1520-0477(1996)077<2857:TNISIS>2.0.CO;2

[12] J. M. Freeman, N. A. DiOrio, N. J. Blair, T. W. Neises, M. J. Wagner, P. Gilman, and S. Janzou, "System Advisor Model (SAM) General Description (Version 2017.9.5)," National Renewable Energy Laboratory (NREL), Golden, CO (United States), Tech. Rep. NREL/TP-6A20-70414, may 2018. [Online]. Available: https://www.nrel.gov/docs/fy18osti/70414.pdf

[13] M. Sengupta *et al.*, "The national solar radiation data base (NSRDB)," *Renewable and Sustainable Energy Reviews*, vol. 89, pp. 51–60, Jun. 2018.

[14] J. Stein, C. Hansen, and M. Reno, "The variability index: A new and novel metric for quantifying irradiance and pv output variability," in *World Renewable Energy Forum*, 05 2012.

[15] 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: https://doi.org/10.21105/joss.00884