



Estimating Subhourly Inverter Clipping Loss From Satellite-Derived Irradiance Data

Preprint

Kevin Anderson and Kirsten Perry

National Renewable Energy Laboratory

*Presented at the 47th IEEE Photovoltaic Specialists Conference (PVSC 47)
June 15 - August 21, 2020*

**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-5K00-76021
June 2020



Estimating Subhourly Inverter Clipping Loss From Satellite-Derived Irradiance Data

Preprint

Kevin Anderson and Kirsten Perry

National Renewable Energy Laboratory

Suggested Citation

Anderson, Kevin and Kirsten Perry. 2020. *Estimating Subhourly Inverter Clipping Loss From Satellite-Derived Irradiance Data: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5K00-76021. <https://www.nrel.gov/docs/fy20osti/76021.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-5K00-76021
June 2020

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

NOTICE

This work was authored 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 Number 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.

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

U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via www.OSTI.gov.

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

NREL prints on paper that contains recycled content.

Estimating Subhourly Inverter Clipping Loss From Satellite-Derived Irradiance Data

Kevin Anderson

National Renewable Energy Laboratory

Golden, USA

kevin.anderson@nrel.gov

Kirsten Perry

National Renewable Energy Laboratory

Golden, USA

kirsten.perry@nrel.gov

Abstract—Photovoltaic system production simulations are conventionally run using hourly weather datasets. Hourly simulations are sufficiently accurate to predict the majority of long-term system behavior but cannot resolve high-frequency effects like inverter clipping caused by short-duration irradiance variability. Direct modeling of this subhourly clipping error is only possible for the few locations with high-resolution irradiance datasets. This paper describes a method of predicting the magnitude of this error using a machine learning regressor ensemble model, comprised of a random forest and an XGBoost model, and 30-minute satellite irradiance data. The method predicts a correction for each 30-minute interval with the potential to roll up into 60-minute corrections to match an hourly energy model. The model is trained and validated at locations where the error can be directly simulated from 1-minute ground data. The validation shows low bias at most ground station locations. The model is also applied to gridded satellite irradiance to produce a heatmap of the estimated clipping error across the United States. Finally, the relative importance of each predictor satellite variable is retrieved from the model and discussed.

Index Terms—photovoltaic, inverter, clipping, saturation, high-frequency, irradiance, variability, modeling, satellite

I. INTRODUCTION

The energy production models driving photovoltaic (PV) system development and financing are usually based on hourly weather files sourced from satellite imagery. Although these satellite-derived datasets are somewhat less accurate than measurements recorded by meteorological stations on the ground, they have the advantage of being available for any location in the satellite’s coverage area, enabling energy models to simulate systems at any location of interest without having to deploy a weather monitoring station ahead of time. The uncertainty associated with using hourly averages instead of higher-frequency measurements has been previously studied [1] and is usually considered acceptable for the production forecasts used in system financing deals. However, as the price of PV modules decline, it is becoming increasingly common for PV system designers to “over build” systems by increasing the size of the PV array relative to the system’s inverter capacity. This capacity difference

is quantified as the “DC/AC” ratio or inverter loading ratio (ILR). Systems with high ILR are more likely to experience inverter saturation where the array’s maximum power point is higher than the inverter’s maximum capacity and the array is operated at a less efficient point on its IV curve to avoid overloading the inverter. This behavior is called “inverter clipping” because it clips the top off the system’s daily power production curve. In practice, modern PV systems with ILRs of around 1.4 can clip on sunny days year-round. The production loss associated with this clipping is offset by the additional production in less sunny periods and made economical by the declining price of PV modules.

This clipping behavior is easily captured in production modeling by comparing the array’s power output to the inverter’s maximum power limit. However, as shown in Figure 1, calculating the effect of clipping at hourly scale will overestimate total system production compared to a higher-resolution simulation. This is because irradiance can vary significantly over the course of an hour depending on the nature of the cloud cover over the array. In some cases, irradiance can cycle between overcast ($\sim 300 \text{ W/m}^2$) and full sun ($> 1000 \text{ W/m}^2$) several times in a single hourly interval. Taking the average of these irradiance values might yield an hourly value below the system’s clipping point, making it look like the system was able to capture all of the hour’s insolation when in reality the peaks would all get clipped off. For modeling purposes, this could be described as the difference between “total interval insolation” and “harvestable interval insolation”.

Because this effect manifests during periods with highly variable irradiance (which as noted earlier is dependent on the nature of cloud cover), the modeling error of an hourly simulation is dependent on climate. In locations with highly stable irradiance like southern California, the error should be quite low. In locations with more variable irradiance, the expected error would be larger.

A few case studies have been published on this topic. Analysis of a 1.23 ILR system in San Diego, California reported an increase in estimated clipped energy from less than 1% to over 2.5% when switching from hourly to 1-second irradiance data [2]. Analysis of a system in Freiburg, Germany reported a decrease in inverter

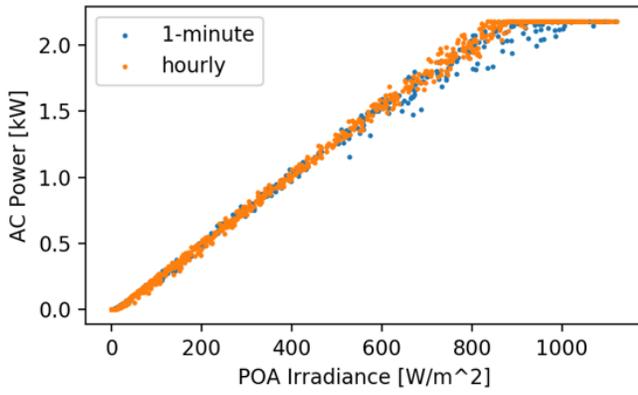


Fig. 1. Simulated system power as a function of plane-of-array (POA) irradiance. The orange dots are from an hourly simulation and show low scatter. The blue points are from a 1-minute simulation and averaged to hourly for the plot. The 1-minute simulation shows some underperformance around the clipping knee because it correctly averages irradiance and the nonlinear power response, but the hourly simulation misses it and overestimates production as a result.

efficiency (including clipping loss) from 95% to 94% by switching from hourly to 1-minute irradiance and a further 0.5% decrease by using 10-second data [3]. Finally, an analysis of irradiance data from Oak Ridge, Tennessee reported using hourly irradiance instead of 1-minute irradiance underestimates clipping loss by 0.4% and 1.5% for ILRs of 1.25 and 1.5, respectively [4]. These studies were able to quantify the error by taking advantage of high-frequency irradiance data from a co-located weather station. However, as mentioned earlier, high-resolution irradiance datasets are rarely available for a given area of interest. To support the modeling of this clipping error in locations without high-resolution ground data, we present a method for predicting the theoretical clipping error from 30-minute satellite-derived weather data from the National Solar Radiation Database (NSRDB) [5].

The method uses an ensemble regressor model, which is comprised of a random forest and an XGBoost model and trained to predict the interval clipping errors simulated from 1-minute ground station data using only low-resolution satellite weather data. Because the model uses only satellite data for its predictions, it can be applied to any location within the coverage area of the NSRDB. The model’s predictions are validated across the ground station network by retraining the model excluding each ground station in turn and comparing its predictions against the ground truth values.

II. METHODOLOGY

The approach taken here is to use known clipping errors and the associated NSRDB weather data to train a machine learning model to predict error using only satellite data that is sampled at a lower frequency. In this case, the known clipping errors are generated by simulating the ground-truth clipping error at various locations around the United States where 1-minute ground irradiance data

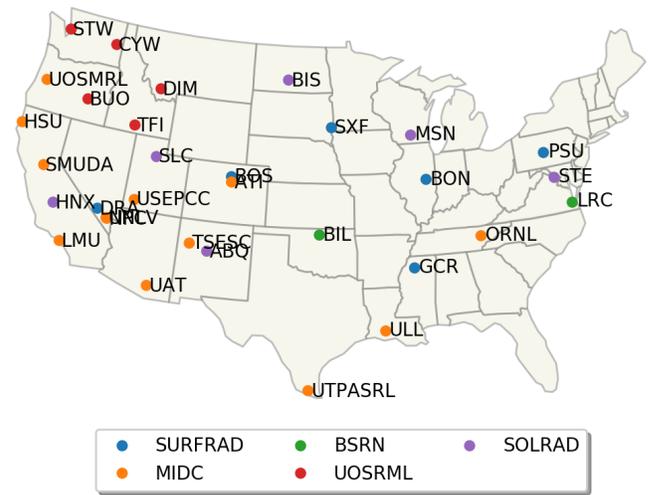


Fig. 2. A map showing the geographic spread of the ground station locations, colored by the source station network. Note that some stations are members of multiple networks, only one of which is shown here.

is available. The 30-minute satellite data associated with each ground station’s location and data set time range is then processed and used to predict the clipping error.

In an effort to make the model as generally-applicable as possible, we combine the data from several ground station networks [6]–[10] with good coverage across various climates in the United States. The geographic distribution of the stations is shown in Figure 2 and specific station details are presented in Table I.

To generate the ground-truth clipping error values, we use the detailed PV model included in the National Renewable Energy Laboratory’s System Advisor Model (SAM) [11] using the PySAM [12] wrapper to enable scripted model runs. SAM is capable of running detailed PV system simulations down to 1-minute resolution. In this case, we use it to predict two array output signals for each ground station: one using the raw 1-minute dataset, and another using the 30-minute average of the 1-minute dataset. The simulations are performed assuming a south-facing fixed-tilt system at 20° with an ILR of 1.4. We do not expect the specific hardware details to be relevant for this calculation, but for completeness, the simulation modeled DC power for a JA Solar JAP6-72-305/3BB polycrystalline module. Other inputs not specific to this simulation (such as wiring losses) were left as the default values. We then post-process the SAM outputs to calculate two 30-minute average timeseries: first, applying a simple clipping model to the 1-minute simulated array power and then calculating the 30-minute average clipped power, and second, clipping the 30-minute average array power. Both result in clipped 30-minute power signals, but the difference is whether clipping is applied at high-frequency scale (as would happen in a real system) or low-frequency scale (as is commonly modeled). The difference between

TABLE I
GROUND STATION METADATA

	Station Code	Years	Data Source	Latitude	Longitude	Elevation (m)	UTC Offset
0	ABQ	2017-2018	SOLRAD	+35.0380	-106.6221	1617	-7
1	ATI	2006-2016	MIDC	+39.7423	-105.1785	1828	-7
2	BIL	1998-2017	BSRN	+36.6050	-97.5160	317	-6
3	BIS	2016-2018	SOLRAD	+46.7718	-100.7596	503	-6
4	BON	2010-2017	SURFRAD	+40.0519	-88.3731	230	-6
5	BOS	2010-2015	SURFRAD	+40.1250	-105.2368	1689	-7
6	BUO	2012-2018	UOSRML	+43.5192	-119.0216	1265	-8
7	DIM	2012-2018	UOSRML	+45.2100	-112.6400	1590	-7
8	DRA	2011-2017	SURFRAD	+36.6237	-116.0195	1007	-8
9	GCR	2010-2017	SURFRAD	+34.2547	-89.8729	98	-6
10	HNX	2016-2018	SOLRAD	+36.3136	-119.6316	73	-8
11	HSU	2009-2017	MIDC	+40.8760	-124.0800	36	-8
12	LMU	2011-2014	MIDC	+33.9667	-118.4226	27	-8
13	LRC	2015-2018	BSRN	+37.1038	-76.3872	5	-5
14	MSN	2016-2018	SOLRAD	+43.0725	-89.4113	271	-6
15	NPC	2010-2012	MIDC	+36.0858	-115.0519	523	-8
16	ORNL	2008-2018	MIDC	+35.9300	-84.3095	245	-5
17	PSU	2010-2017	SURFRAD	+40.7201	-77.9309	376	-5
18	SLC	2016-2018	SOLRAD	+40.7722	-111.9549	1288	-7
19	SMUDA	2011	MIDC	+38.5459	-121.2403	51	-8
20	STE	2016-2018	SOLRAD	+38.9720	-77.4869	85	-5
21	STW	2016-2018	UOSRML	+47.6800	-122.2500	20	-8
22	SXF	2010-2017	SURFRAD	+43.7340	-96.6233	473	-6
23	TFI	2014-2018	UOSRML	+42.5500	-114.3500	1200	-8
24	TSESC	2012	MIDC	+35.4186	-108.0883	2106	-7
25	UAT	2014-2015	MIDC	+32.2297	-110.9553	786	-7
26	UNLV	2007-2018	MIDC	+36.1070	-115.1425	615	-8
27	USEPCC	2011-2012	MIDC	+37.6960	-113.1648	1675	-7
28	UTPASRL	2012-2018	MIDC	+26.3059	-98.1716	45	-6

these two signals is what we call “1-to-30” clipping error, and represents the amount that a low-frequency simulation overestimates production. For convenience, we normalize the 1-to-30 clipping error by the simulated AC capacity so it spans the range $[-1, 0]$. Note that to avoid the influence of other nonlinear effects (e.g. solar position and cell temperature) from confounding the results, we also calculate the 30-minute difference in the unclipped signals and remove that from the clipping error.

Once the 1-to-30 clipping error is calculated, we pair it with the values used as predictors. In an effort to keep the prediction process as simple as possible, we use only raw NSRDB values and derived values that are easily scripted using the open-source library `pvl-lib-python` [13]. The NSRDB data provides the estimated surface irradiance components global horizontal irradiance (GHI), diffuse horizontal irradiance (DHI), and direct normal irradiance (DNI), the corresponding three clearsky irradiance components, as well as ambient temperature and wind speed. From the irradiance components, plane-of-array (POA) irradiance and clearsky POA irradiance are calculated for a south-facing array at 20° tilt using the isotropic diffuse sky model as implemented in `pvl-lib-python`. Additionally, we calculate estimated cell temperature using `pvl-lib-python`’s Sandia Array Performance Model (SAPM) temperature function with the “open rack/polymer back” thermal parameters.

In addition to these PV variables, two additional feature

variables are derived to aid the model: the “velocity” of POA irradiance, i.e. the first-order difference of the time series, and the difference between POA irradiance and clearsky POA irradiance at each time interval.

These values are collected to form a 30-minute predictor dataset containing:

- GHI and clearsky GHI
- POA irradiance and clearsky POA irradiance
- Cell temperature
- POA velocity
- Difference between POA and Clearsky POA

Finally, each of the predictor variables is max-min normalized by location so that each series spans the range $[0, 1]$. This predictor dataset and the associated 1-to-30 clipping error values were generated for each of the ground stations at thirty minute intervals. Datasets of this form were used for both model training and model validation. Total dataset size, including all 29 locations, is approximately 1.31 million rows, with the size of the data by location varying.

The prediction model used in this study is an ensemble model implemented with the open-source `scikit-learn` package in Python [14], comprising a random forest (RF) regressor model and an extreme gradient boosting (XGBoost, [15]) regressor model. In the model, each 30-minute interval acts independently; consequently, the model will treat each timestamp separately when predicting 1-to-30 clipping error, where the only dependence on additional

time readings is the previously calculated POA velocity parameter. Model hyper-parameters include:

- RF model instantiated with 100 trees.
- Equal fifty-percent weighting for the XGBoost model and the RF model in the finalized ensemble model.

To test the model’s predictive ability at locations it was not trained on, we use a cross-validation technique. By omitting the data associated with one station from the training dataset, it can be used as a validation set to test the model’s accuracy at a novel location. This method was applied to each ground station to quantify the model’s error across different geographies. In all, the model was regenerated 29 times, for each of the 29 locations. Root mean squared error (RMSE) and mean bias error (MBE) were calculated for each location’s predictions using the cross-validation technique, ensuring unbiased model results.

Additionally, we trained an “all-in” model using data from the entire ground station network and used it to predict the clipping error across a grid of points covering the United States. Ensemble model hyper-parameters, including individual model weighting, remained consistent while regenerating models using cross-validation, as well as when generating the “all-in” model.

III. RESULTS

Figure ?? compares clipping error predicted by the ensemble model to the simulated clipping error from ground station data. Each point represents clipping error as a percentage of monthly production, with one point per station per month. The results show low overall bias, although the scatter for individual station-months can be significant. Figure 4 shows the difference between predicted and actual monthly clipping errors by station and the associated error statistics are listed in Table III. The model performs well at most stations (the model’s predictions have $|MBE| < 0.1$ for over half the stations) and it seems much of the scatter in Figure ?? is from only a few stations. It is not clear why the model performs poorly on these stations. However, the overall results from the cross-validated models show that the model typically performs well on novel data sets, although there is room for improvement. We expect the prediction error would be reduced by refining the ensemble model structure and training method.

Table II provides a list of the feature importances for both the XGBoost and Random Forest models which comprise the ensemble model. Feature importance is a value ranging between $[0, 1]$, providing a fractional representation of how much each model variable affects overall model outcome. For both models, normalized clearsky POA has the highest feature importance score, indicating that this variable has the greatest effect on overall model outcome. However, the XGBoost model assigned a feature importance score of 0.495 to the normalized clearsky POA variable, which is significantly higher than the random forest model score of 0.196 for the same variable. With

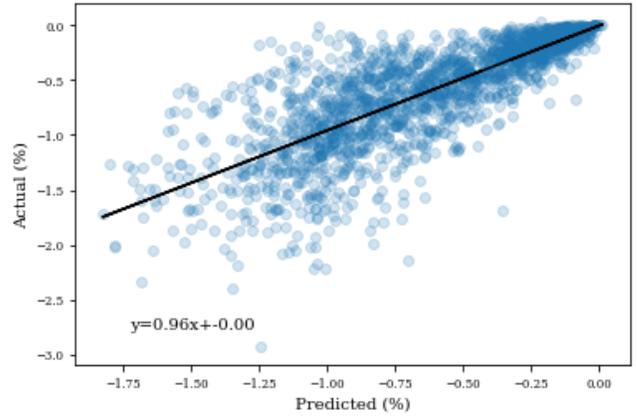


Fig. 3. Comparison of clipping error predicted by the ensemble model and expected loss from ground station data. Units are in percent of monthly production. A least-squares fit is shown. Results are for each location, using cross-validated predictions.

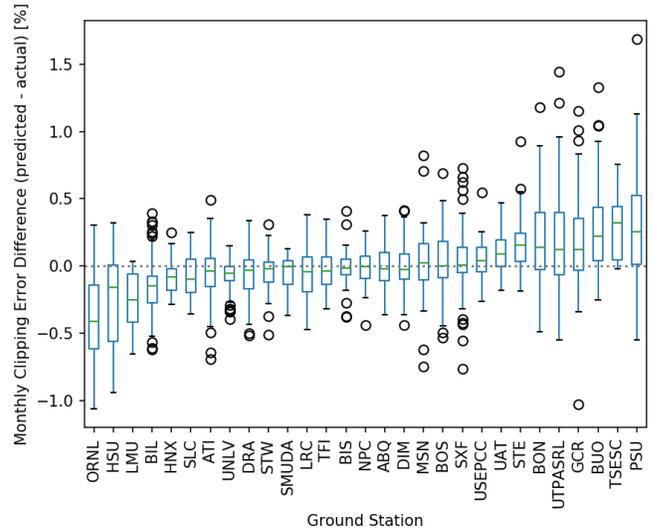


Fig. 4. The distribution of monthly clipping error differences (predicted - actual) by station, ranked from least to greatest overall MBE. Units are percent of monthly production. The model shows minimal prediction bias at most stations, although it does show significant bias and scatter at some stations.

the exception of normalized clearsky POA, feature variable rankings varied based on the model.

Figure 5 shows the model’s prediction at a grid of locations across the United States using the NSRDB dataset for the year 2018. The geographic variation in the map is roughly consistent with an irradiance variability map previously published by researchers at Sandia National Laboratory as it relates to power ramp rates in the context of grid integration [16]. It shows a nation-wide average annual clipping error of roughly 1%, but as shown by the range of monthly error in Figure ??, individual months can be significantly higher or lower. We emphasize that the model used to generate Figure 5 was trained on a certain system configuration and thus its predictions are

TABLE II
ENSEMBLE MODEL FRACTIONAL FEATURE IMPORTANCES

Feature Variable	XGBoost	Random Forest
Normalized Clearsky POA	0.495	0.196
Difference between Clearsky POA and POA, Normalized		
Normalized POA	0.172	0.105
Normalized POA Velocity	0.112	0.151
Normalized Cell Temperature	0.096	0.170
Normalized Clearsky GHI	0.050	0.140
Normalized GHI	0.039	0.128
	0.036	0.110

TABLE III
CROSS-VALIDATION RESULTS BY STATION:
MONTHLY MODEL PREDICTION ERROR

	Station Code	MBE	RMSE
0	ORNL	-0.41	0.50
1	HSU	-0.26	0.42
2	LMU	-0.25	0.32
3	BIL	-0.16	0.23
4	HNX	-0.08	0.15
5	SLC	-0.07	0.18
6	ATI	-0.07	0.20
7	UNLV	-0.07	0.12
8	DRA	-0.06	0.18
9	STW	-0.05	0.16
10	SMUDA	-0.05	0.15
11	LRC	-0.05	0.19
12	TFI	-0.02	0.16
13	BIS	-0.02	0.17
14	NPC	-0.02	0.13
15	ABQ	-0.01	0.15
16	DIM	-0.00	0.17
17	MSN	0.03	0.29
18	BOS	0.03	0.24
19	SXF	0.04	0.23
20	USEPCC	0.05	0.18
21	UAT	0.10	0.19
22	STE	0.17	0.27
23	BON	0.17	0.34
24	UTPASRL	0.19	0.45
25	GCR	0.19	0.38
26	BUO	0.26	0.40
27	TSESC	0.29	0.39
28	PSU	0.29	0.46

specific to that system configuration. Additionally, there is some evidence in the literature that 1-minute data may be insufficient to fully resolve this effect (e.g. [2], [3]) and the true clipping error may be higher. We include the map here not to provide guidance for adjusting energy models to account for the effect, but rather to illustrate the model’s geographic behavior.

A notable feature is the distinct low error region starting in California’s central valley and extending down into Arizona. Because these areas are dominated by clearsky conditions, the low predicted clipping error is consistent with expectation. An additional note is that the outlier pixels in Figure 5 predicting large error appear to be associated with bodies of water: Great Salt Lake (Utah), Salton Sea (California), and Mobile Bay (Alabama).

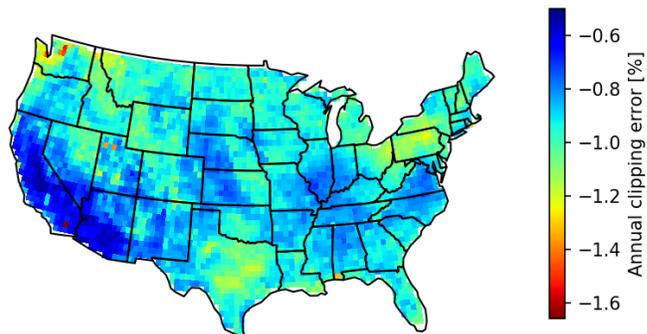


Fig. 5. A map showing the model’s estimated 1-to-30 clipping error on a $0.5^\circ \times 0.5^\circ$ grid across the United States for the year 2018.

IV. CONCLUSIONS

This paper presents a method of using widely-available weather data derived from satellite imagery to predict the effect of local high-frequency irradiance variability on PV system production. The method can be applied to any location within the NSRDB’s coverage area, but is validated only within the United States. The method is trained and validated using data from an extensive network of ground weather stations, showing low bias when considering the validation as a whole. The results of a cross-validation analysis showed that the model generally performs well at novel locations ($|\text{MBE}| < 0.1$ at monthly scale for over half of the stations) but its predictions do have high bias and scatter at some locations. Because the method predicts clipping error at 30-minute intervals, this method has the potential to be used to correct conventional hourly energy estimates. The model’s predictions across a grid of locations covering the United States are qualitatively consistent with existing results in the literature. Outlier results for pixels near large bodies of water suggest that predicted clipping error should be treated with caution in NSRDB pixels containing bodies of water. Overall, these results demonstrate the viability of using satellite data to predict the effects of high-frequency irradiance variability.

V. ACKNOWLEDGEMENT

This work was authored 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 non-exclusive, 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] 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.
- [2] J. Luoma, J. Kleissl, and K. Murray, "Optimal inverter sizing considering cloud enhancement," *Solar Energy*, vol. 86, no. 1, pp. 421–429, Jan. 2012.
- [3] B. Burger and R. R  ther, "Inverter sizing of grid-connected photovoltaic systems in the light of local solar resource distribution characteristics and temperature," *Solar Energy*, vol. 80, no. 1, pp. 32–45, Jan. 2006.
- [4] J. Good and J. X. Johnson, "Impact of inverter loading ratio on solar photovoltaic system performance," *Applied Energy*, vol. 177, pp. 475–486, Sep. 2016.
- [5] M. Sengupta *et al.*, "The national solar radiation data base (NSRDB)," *Renewable and Sustainable Energy Reviews*, vol. 89, pp. 51–60, Jun. 2018.
- [6] 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. 2341–2358, 2000.
- [7] 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](https://doi.org/10.1175/1520-0477(1996)077<2857:TNISIS>2.0.CO;2)
- [8] J. Peterson and F. Vignola, "Structure of a comprehensive solar radiation dataset," *ASES Proceedings*, 2017.
- [9] "NREL MIDC: Measurement and Instrumentation Data Center." [Online]. Available: <https://midcdmz.nrel.gov/>
- [10] A. Driemel *et al.*, "Baseline surface radiation network (BSRN): structure and data description (1992–2017)," *Earth Syst. Sci. Data*, no. 10, pp. 1491–1501, 2018.
- [11] 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)," *NREL Technical Report TP-6A20-70414*, 5 2018.
- [12] "PySAM Version 2.0.2," National Renewable Energy Laboratory. Golden, CO, May 2020. [Online]. Available: <https://github.com/nrel/pysam>
- [13] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlb python: a python package for modeling solar energy systems," *Journal of Open Source Software*, vol. 3, no. 29, p. 884, Sep. 2018. [Online]. Available: <https://doi.org/10.21105/joss.00884>
- [14] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 785–794.
- [16] 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, Jul. 2017. [Online]. Available: <https://doi.org/10.1016/j.solener.2017.05.005>