PREDICTING SHORT-TERM VARIABILITY OF HIGH-PENETRATION PV

Thomas E. Hoff
Clean Power Research
Napa, CA 94558
tomhoff@cleanpower.com

Richard Perez
ASRC, 251 Fuller Rd
Albany, NY, 12203
rperez@albany.edu

ABSTRACT

This article evaluates the ability of three operational satellite models (SolarAnywhere® Standard, Enhanced, and High Resolution) to predict ground-based measured irradiance conditions. Results suggest that the performance of the satellite-based monitoring approaches that of well-maintained redundant sensors at a fraction of the operational cost.

SolarAnywhere Enhanced and High Resolution have an *annual* error that is slightly higher than co-located ground sensors when the invalid ground-based measured data are excluded from the analysis. SolarAnywhere High Resolution has 1/3 less error at an hourly time interval compared to SolarAnywhere Standard Resolution. Results also show that SolarAnywhere High Resolution has 10 percent Mean Absolute Error on a one-minute time interval, making it well suited to provide the basis for data required to perform high-penetration PV studies.

The paper also demonstrates that geographically dispersing PV systems results in both lower ***variability*** and less ***prediction error***; this is consistent with recent observations that regional solar resource predictions are considerably more accurate than single site predictions [1].

1. INTRODUCTION

Solar photovoltaic (PV) plant power production variability is one of the critical challenges to greater penetration of PV into the state's electricity system. As illustrated by the list of References, a number of studies have examined the issue of PV output variability [2-12]. A consistent finding of these studies is that variability is reduced when PV systems are geographically dispersed. That is, variability reduces as the number of systems increases across a sufficiently large geographic region.

A second critical challenge to greater penetration of PV is the ability to accurately forecast PV power production variability when it occurs. The California Energy Commission's (CEC) Public Interest Energy Research (PIER) program has embarked on a data validation effort titled, "Demonstration and Validation of PV Output Modeling Approach." A methodology has been developed that uses satellite-derived solar data to forecast PV fleet output and quantify variability given the design attributes and locations of PV systems [13]. The methodology uses advanced algorithms to track cloud patterns and calculate plant correlation coefficients.

The California Independent System Operator (California ISO) sees the potential of using this methodology to calibrate its studies of system operations under alternative renewable energy scenarios, as well as the potential for forecasting PV output. However, before the methodology will be practical and usable in studies and forecasting by the California ISO and others, additional work is needed in data analysis, validation, and system integration.

The California Solar Initiative (CSI) funded the development of an enhanced resolution satellite-based solar resource database for the state of California. It is referred to as SolarAnywhere Enhanced Resolution [14]. The database has a 1 km spatial resolution and ½ hour temporal resolution, using the native spatial and temporal resolution of the US geostationary satellites. This data set has been further expanded to have a 1 km spatial, 1 minute temporal resolution by applying intra-interval short-term forecasting [15]. It is referred to as SolarAnywhere High Resolution [14]. These data sets have the potential to provide the solar resource data required by the methodology described above.

The first step, however, is to quantify the accuracy. This first step constitutes the main objective of this paper. There have been some initial efforts at data validation of SolarAnywhere Enhanced Resolution. For example, Jamaly, Bosch, and Kleissl [16] compared measured output for a fleet of 86 PV systems in San Diego County to simulated

PV fleet output using SolarAnywhere Enhanced Resolution data during high ramping conditions. The authors concluded that "the satellite data were able to closely follow the aggregate power output and detect the timing of the ramps, while 5 irradiance measurement stations [dispersed within the region occupied by the PV systems] stations were not as accurate due to smaller number and non-representative geographical distribution with respect to the PV sites." This useful observation calls for a systematic and quantitative evaluation of prediction accuracy as presented in this paper.

2. METHODS

2.1 Definitions

Accuracy validation often means different things to different people. As such, it is useful to begin with a definition of how accuracy quantification can be performed.
There are three fundamental questions that need to be answered in order to provide a clear definition of how accuracy validation is performed.

1. What is the data source?
2. What are the time attributes?
3. What is the evaluation metric?

2.1.1 Data Source

The first step is to identify the data that is being evaluated. Options include irradiance data or PV power production simulated using irradiance data and other parameters. In addition, the analysis can be performed for individual locations or fleets (i.e., multiple locations). This paper focuses on irradiance data. The analysis is performed for both individual locations and fleets.

2.1.2 Time Attributes

The second step is to specify the required time attributes. These include:

- **Time period**: total amount of data included in the analysis. This can range from a few minutes to many years. This paper focuses on one year of data.
- **Time interval**: how the data in the time period is binned. This can range from a few seconds to annually. For example, if the time period is one year and the time interval is one hour, the time period would be binned into 8,760 increments. This paper examines one-minute to one-year time intervals.
- **Time perspective**: when the predicted observation is reported. This can range from historical, (backward looking) to forecasted a few hours ahead, to forecasted multiple days ahead (forward looking). This paper focuses on historical data.

2.1.3 Evaluation Metric

The third step is to select the evaluation metric. Error quantification metrics used in assessing absolute irradiance model accuracy such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) have been precisely defined [17, 18]. Their relative counterpart (results expressed in percent), however, can be subject to interpretation and may cover a wide range of values for a given set of data depending on reporting practice.

Hoff et al. [19] suggest that the MAE relative to available energy is a good method to measure relative dispersion error. This is the method used in the present analysis. The MAE relative to the average energy available is calculated by summing the absolute error for each time interval over the time period, and then dividing by the total available energy.

$$Relative\ MAE = \frac{\sum_{t=1}^{N}\left|I_t^{test} - I_t^{ref}\right|}{\sum_{t=1}^{N} I_t^{ref}} \qquad (1)$$

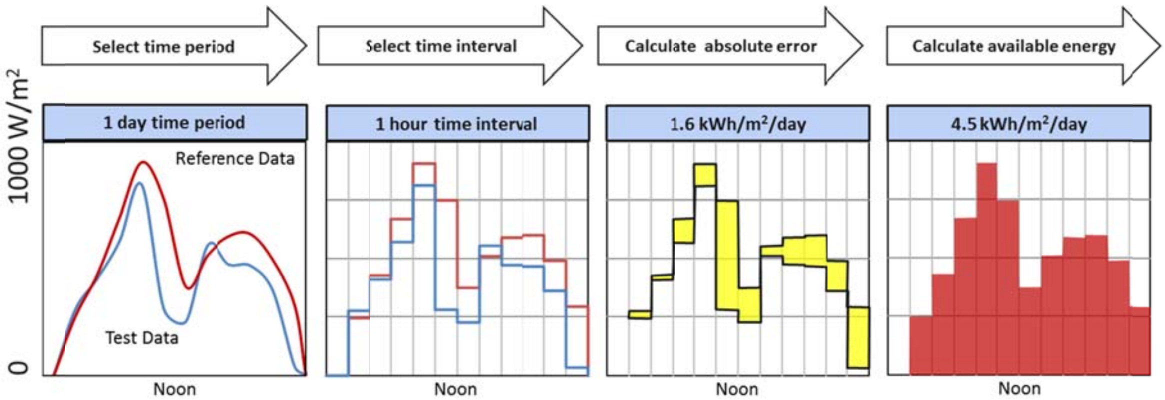
where I_t^{test} is the test irradiance at time *t*, I_t^{ref} is the reference irradiance at time *t*, and *N* is the number of time intervals.

It is useful to provide a hypothetical example of how to calculate the MAE relative to available energy. A short time period (one day) is selected in order to graphically illustrate the calculations; the actual calculations in this paper use a one-year time period.

As presented in Fig. 1, the process is follows:

- Select time period: 1 day.
- Select time interval: 1 hour.
- Calculate absolute error for each hour and sum the result as described in the top part of Equation (1): 1.6 kWh/m^2/day.
- Calculate available energy for each hour from reference data and sum the result as described in the bottom part of Equation (1): 4.5 kWh/m^2/day.
- Calculate Relative Mean Absolute Error: 36% (i.e., 1.6/4.5).

Fig. 1. Mean Absolute Error relative to available energy calculation example.



It is important to note that a more often reported measurement of error is MAE relative to generating capacity. In the above example, however, it is unclear over what time period the generating capacity should be selected. Should it be capacity during daylight hours or capacity over the entire day, including night time hours? MAE relative to daytime capacity is about 13.3% (i.e., 1.6/12) while MAE relative to full day capacity is about 6.6% (i.e., 1.6/24).

It is due to this sort of ambiguity, as well as the fact that MAE relative to energy is a much more stringent metric (e.g., in this example, MAE relative to energy is 6 times higher than MAE relative to daily generation capacity), that the MAE relative to energy is selected as the evaluation metric.

2.1 Validation Approach

The present model validation is part of a project whose overall goal is to demonstrate and validate PV power prediction models in collaboration with the California ISO. Two key objectives of this project are: (1) to measure the accuracy of the models for PV sources within the California ISO control area; and (2) to ensure that the data is delivered in a manner compatible with the existing energy and reserve market mechanisms.

This paper focuses on the first objective and quantifies the accuracy of the irradiance data for a one-year time period (2011) and time intervals ranging from one minute to one year using a historical time perspective. The analysis was performed for both individual locations (i.e., single solar systems) and the ensemble of those locations (i.e., a fleet of solar systems).

A total of six test locations were analyzed where PV systems are located within California ISO's control area. The locations are identified as locations A through F for purposes of confidentiality. Each location is equipped with two redundant global horizontal irradiance (GHI) sensors. One of the sensors was used as a reference and compared to four test configurations: the second ground sensor, and the three satellite-derived sources (SolarAnywhere Standard, Enhanced, and High Resolution data sets).

The validation approach involved the following steps:

- Obtain time-series GHI data for 2011 for six locations (see Fig. 2 for an example of one day of data):
  - 4 second data averaged into 1-minute time intervals from two separate sensors at each location (sources: California ISO [20])
  - Satellite based data at the following resolutions (source: SolarAnywhere [14])
    - 1 minute, 1 km grid (High Resolution)
    - ½ hour, 1 km grid (Enhanced Resolution)
    - 1 hour, 10 km grid (Standard Resolution)
- Time-synchronize data sets by converting ground sensor data from Pacific Daylight Time to Pacific Standard Time.
- Evaluate all observations for data quality; exclude data where any one of the data sources has data quality issues.
- Calculate MAE relative to the actual energy available using the ground sensor that minimizes SolarAnywhere error as a reference.
- Calculate MAE relative to the actual energy available using the other ground sensor as a reference.
- Repeat the analysis for fleets of locations.

3. RESULTS

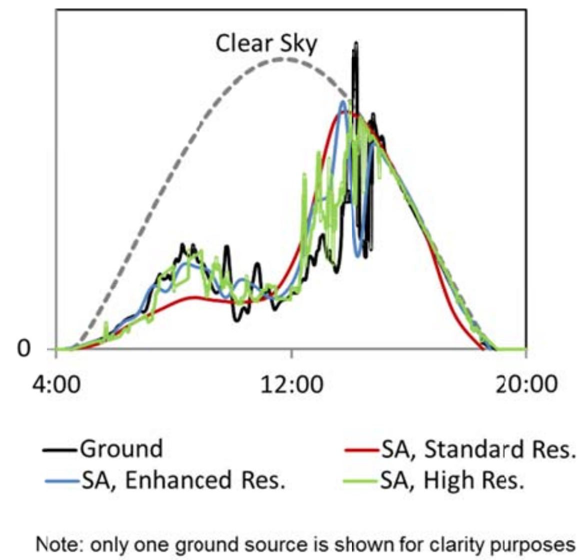
The relative MAE was calculated for individual locations and for a fleet of locations.

3.1 Individual Locations

Fig. 3 presents the average MAE of four individual locations (two locations were eliminated due to insufficient data after data screening). The black line summarizes the error when two ground stations were used (one was the reference and the other was the test). The green, blue, and red regions summarize the error when SolarAnywhere High, Enhanced, and Standard Resolution were compared to the ground sensor. The green, blue, and red areas are regions rather than lines because they compare satellite data to ground data using the two different ground sensors; the top of the region is the comparison using the ground sensor that maximizes error; the bottom of the region is the comparison using the ground sensor that minimizes error.

There are several important things to notice in the figure. First, as expected, error decreases for all data sources as the time interval increases. Second, accuracy improves for each of the three satellite models as the spatial and temporal resolutions are increased. Third, error exists even between two ground sensors that are in almost the same location (i.e., ground sensors have 1 percent annual error). Fourth, SolarAnywhere High Resolution has only 10 percent error over a one-minute time interval, 7 percent error over a one-hour time interval, and 2 to 3 percent error on a one-year time interval.

Fig. 2. GHI data from 4 sources (July 4, 2011, Site B).



3.2 Fleet of Locations

As illustrated by the list of References, a number of studies have examined the issue of PV output variability. A consistent finding of these studies is that variability is reduced when PV systems are geographically dispersed. That is, variability is reduced as the number of systems increases across a sufficiently large geographic region. So far, this paper has focused on the error associated with individual locations.

Fig. 3. Average MAE of 4 individual locations.

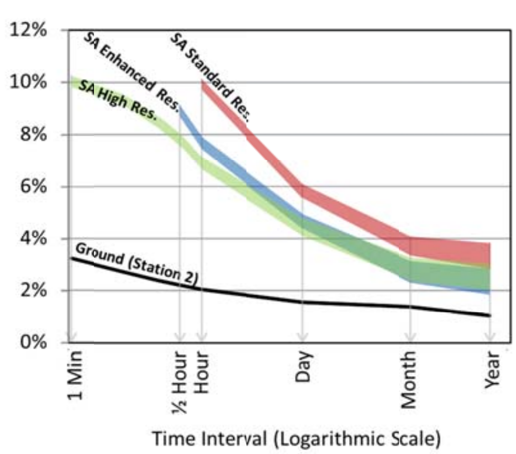
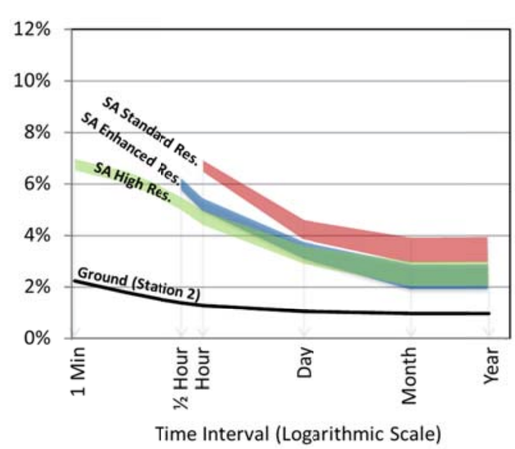


Fig. 4. MAE of 4 locations combined.



While individual locations are of interest in some cases, there are certainly many other cases in the utility industry when users are most interested in the error associated with a set of locations.

The MAE analysis was repeated with the input data being the combined irradiance across four locations. The results are presented in Fig. 4. A clear reduction in error due to combining locations can be seen by comparing Fig. 4 to Fig 3. That is, the effect of geographic dispersion on reducing output variability reduction that has been observed by others is now also observed with regard to prediction accuracy: accuracy improves as a geographically diverse set of independent locations are combined.

4. CONCLUSIONS

Two critical challenges to greater penetration of PV into a state's electricity system are: (1) PV output variability; and (2) ability to accurately predict PV output variability. A number of researchers focusing on the first challenge have demonstrated that PV output variability is reduced by geographic diversity. This paper begins to quantify the accuracy in predicting variability.

Results suggest that, first, satellite-based irradiance has annual error comparable to ground sensors. Thus, satellite data may perform as well as ground data for plant siting at a fraction of the cost, plus the benefit of long-term data streams. It should be noted that even well maintained ground sensors produce considerably more invalid data points than the satellite (a ratio of 100-to-1 in the present study), and that the satellite data were key in detecting these erroneous data points (particularly when both redundant sensors were inaccurate at the same time).

Second, high resolution satellite-based irradiance has 10 percent one-minute error for a single location, making it suited to provide the data required to perform high penetration PV studies.

Third, accuracy improves predictably due to the benefit of geographic dispersion. That is, the effect of geographic dispersion on reducing output variability reduction that has been observed by others is now also observed with regard to prediction accuracy.

5. ACKNOWLDEGEMENTS

This study was funded under a California Energy Commission (CEC) Public Interest Energy Research (PIER) Grant Agreement titled “Demonstration and Validation of PV Output Modeling Approach” with co-funding from the California ISO. Special thanks to Jim Blatchford at the California ISO for providing data and guidance for the analysis. Opinions expressed herein are those of the authors only.

5. REFERENCES

1. Lorenz E., T. Scheidsteger, J. Hurka, D. Heinemann and C. Kurz, (2011): Regional PV power prediction for improved grid integration. Progress in Photovoltaics, 19, 7, 757-771.
2. Hoff, T. E., Perez, R. 2012. “Modeling PV Fleet Output Variability.” Solar Energy, Forthcoming.
3. Hoff, T. E., Perez, R. 2010. “Quantifying PV power Output Variability.” Solar Energy 84 (2010) 1782–1793.
4. Kuszamaul, S., Ellis, A., Stein, J., Johnson, L. 2010. “Lanai High-Density Irradiance Sensor Network for Characterizing Solar Resource Variability of MW-Scale PV System.” 35th Photovoltaic Specialists Conference, Honolulu, HI. June 20-25, 2010.
5. Mills, A., Wiser, R. 2010. Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. Lawrence Berkeley National Laboratory Technical Report LBNL-3884E.
6. Mills, A., Ahlstrom, M., Brower, M., Ellis, A., George, R., Hoff, T., Kroposki, B., Lenox, C., Miller, N., Stein, J., Wan, Y., 2009. Understanding variability and uncertainty of photovoltaics for integration with the electric power system. Lawrence Berkeley National Laboratory Technical Report LBNL-2855E.
7. Navigant, Sandia National Laboratories, and Pacific Northwest National Laboratory. 2011. Large-Scale PV Integration Study. Report.
8. Norris, B. L., Hoff, T. E., 2011. “Determining Storage Reserves for Regulating Solar Variability.” Electrical Energy Storage Applications and Technologies Biennial International Conference 2011.
9. Perez, R., Hoff, T.E. 2011. “Solar Resource Variability: Myth and Fact.” Solar Today, August/September 2011.
10. Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr., C. , Hoff, T. E. 2011. Parameterization of site-specific short-term irradiance variability, Solar Energy 85 (2011) 1343-1353.
11. Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr., C. , Hoff, T. E. 2010b. “Short-term irradiance variability correlation as a function of distance.” Accepted to Solar Energy (Feb. 2011).

12. Woyte, A., Belmans, R., Nijs, J. 2007. “Fluctuations in instantaneous clearness index: Analysis and statistics.” Solar Energy 81 (2), 195-206.
13. Hoff, T. E. (2011). U.S. Patent Applications: Computer-Implemented System and Method for Determining Point-to-Point Correlation of Sky Clearness for Photovoltaic Power Generation Fleet Output Estimation (Application Number 13/190,435), Computer-Implemented System and Method for Estimating Power Data for a Photovoltaic Power Generation Fleet (Application Number 13/190,442), Computer-Implemented System and Method for Efficiently Performing Area-to-Point Conversion of Satellite Imagery for Photovoltaic Power Generation Fleet Output Estimation(Application Number 13/190,449).
14. Solar Anywhere, 2012. Web-Based Service that Provides Hourly, Satellite-Derived Solar Irradiance Data Forecasted 7 days ahead and Archival Data back to January 1, 1998. www.SolarAnywhere.com.
15. Perez, R., T. Hoff and S. Kivalov, (2011): Spatial & temporal characteristics of solar radiation variability. Proc. of International Solar Energy (ISES) World Congress, Kassel, Germany
16. Jamaly, S., Bosch, J., Kleissl, J. 2012. “Aggregate Ramp Rates of Distributed Photovoltaic Systems in San Diego County.” Submitted to IEEE Transactions on Sustainable Energy.
17. Willmott CJ, Ackleson SG, Davi RE, Feddema JJ, Klink KM, Legates DR, O'Donnell J, Rowe CM (1985) “Statistics for the evaluation and comparison of models.” J. Geophys. Res. 90C: 8995-9005
18. IEA Solar Heating and Cooling Programme, Task 36, Subtask A, (2011): Standard Qualification for Solar Resource Products.
19. Hoff, T.E., Perez, R., Kleissl, J., Renne, D. Stein, J. 2012. “Reporting of Relative Irradiance Prediction Dispersion Error.” American Solar Energy Society Annual Conference. May 2012. Denver, Co.
20. California Independent System Operator monitoring program (2011).