SOLAR MONITORING, FORECASTING, AND VARIABILITY ASSESSMENT AT SMUD

James Bing
NEO Virtus Engineering Inc.
410 Great Rd., B-6
Littleton, MA 01460
Email: jbing@neovirtus.com

Obadiah Bartholomy
Thomas Vargas
Sacramento Municipal Utility District
6201 S St
Sacramento, CA 95817
Email: Obadiah.Bartholomy@smud.org
Thomas.Vargas@smud.org

Tom Hoff
Clean Power Research
1700 Soscol Ave., Ste. 22
Napa, CA 94559
tomhoff@cleanpower.com

Pramod Krishnani
Belectric Inc.
8076 Central Avenue
Newark CA 94560
Email: krishnani.pramod@gmail.com
pramod.krishnani@belectric-usa.com

Richard Perez
ASRC, the University at Albany
251 Fuller Rd.
Albany, NY, 12203
Email: perez@asrc.cestm.albany.edu

ABSTRACT

The paper summarizes the deployment of a 71 station solar monitoring network in Sacramento, California, and its use in validating variability relationships as well as satellite based irradiance datasets. The data cleanup methods are described for eliminating shading artifacts in the ground-based solar monitoring data. The cleaned data is then evaluated to confirm theoretical relationships of spatial correlation between PV plants developed by Hoff and Perez. The relationships are confirmed for 1 minute, 5 minute, and 10 minute timeframes. Additionally, the ground-based datasets are compared to satellite datasets for determining error. Possible sources of error are discussed, and results show that for a half hour timeframe, error or difference in GHI is between 6 and 11%. For DNI, errors range from 17 – 22%. A portion of the errors can be attributed to bias, with GHI bias ranging from between -1 and -7% indicating satellite estimated slightly greater GHI resource and DNI bias ranging from between -1 and 11%, indicating generally that the ground-based RSR's measured slightly greater values than the satellite datasets.

1. INTRODUCTION

As solar penetrations increase rapidly, SMUD and utilities around the country have begun to focus on questions of how the variability of those systems and the industry's ability to forecast their output will impact other resources on our system. Forecasting and resource measurement are growing rapidly both within utilities and in the private sector. To better understand the state of forecasting, variability, and future impacts to our system, SMUD and NEO Virtus Engineering have deployed a network of 71 solar monitoring devices covering most of SMUD's 2330 square kilometer service territory. The project started in June, 2010 and will continue for 2 years. The devices were installed to validate solar forecasts and solar resource variability for high penetrations of solar on SMUD's grid. This study will provide statistical analysis of solar radiation variability across Sacramento County, further validation of the Hoff-Perez PV variability correlation theory, and provide validation of satellite based GHI and DNI irradiance data,

Solar monitoring began in May of 2011 and will last for at least 14 months. In this paper the data are being used to validate theoretical relationships developed by Hoff and Perez to represent variability between distributed PV systems.

In addition to five Rotating Shadowband Radiometer's (RSR) supplied under this project SMUD's three existing RSRs bring to eight the number of devices capable of measuring direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI) data. These data are compared with the 1km and 10 km grid satellite DNI irradiance data. Data from 66 ground-mounted GHI sensors are compared to satellite based GHI values.

2. REGIONAL IRRADIANCE GROUND TRUTH DATA SET

The purpose of the grant from the CPUC in 2010 was to develop and deploy hardware and software tools to model and mitigate impacts of high penetrations of PV on the distribution network. SMUD's grant partners and subcontractors on the project include Hawaiian Electric Company (HECO), BEW Engineering, Sunpower Corporation, and NEO Virtus Engineering. The full scope includes modeling and measuring high PV penetration circuits, developing utility interfaces to enhance the understanding of the performance of intermittent resources, developing methods to utilize the smart meters to communicate with PV inverters, and finally this project, to deploy a network of irradiance sensors to monitor and validate solar forecasting approaches. Overall, these efforts will benefit the utilities involved as well as all California ratepayers by identifying solutions to integrating increasing amounts of PV onto the distribution grid.

As part of this research and to validate forecast accuracy, irradiance measurements are being made using a combination of eight RSRs (primary stations) and sixty six global horizontal (GHI) measurements systems (secondary stations). This combination of primary and secondary monitoring stations has been deployed on the same five kilometer square grid as used by the National Digital Forecast Database (NDFD) for their skycover (cloud cover) forecasts. The monitored area spans almost 1775 square kilometers within SMUD's service territory. This pairing of primary and secondary stations mimics the format of the National Solar Resource Database (NSRDB). All secondary stations take measurements every two seconds and record one minute averages. Data is being retrieved nightly. At each download the logger station clocks are synchronized with the server clock, which is kept at GMT, if difference between logger time and the server is greater than 3 seconds.

The monitoring stations have been located in the "nominal centroid" of each 25 square kilometer NDFD cell. The funded research will first establish the irradiance forecasts over the monitored region and then will quantify the error between measured and forecast irradiance over the term of the experiment.

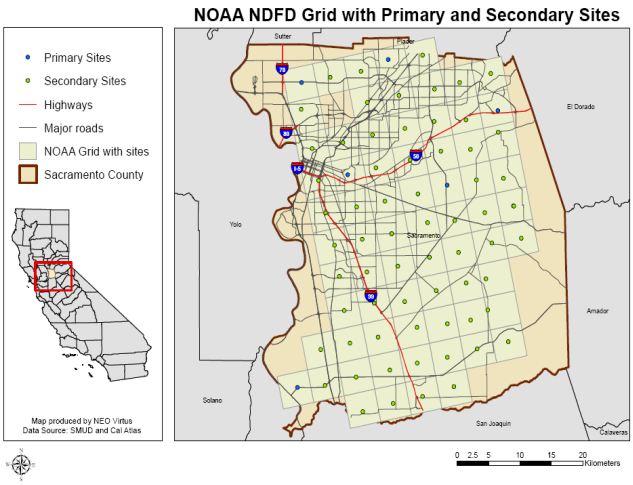


Fig. 1: Map of SMUD Service Territory, NDFD Grid cells and primary and secondary sites.

The intent of the data set is to provide body of ground truth data covering a geographic area of a size which could incorporate both distributed generation, of the residential and commercial scale, as well as central PV plant generation of the utility scale, to validate emerging irradiance forecast methodologies. A primary deliverable of this research will be the database of GHI and temperature measurements from the secondary stations and DNI, DHI, GHI and temperature from the primary stations. The monitoring network will be deployed for approximately 14 months so that the database will cover a full 12 months with all 71 stations deployed. Once completed, this database will be made available to researchers in the field of solar energy forecasting.

3. IRRADIANCE DATA QUALITY MANAGEMENT

The 5km grid spacing of the measurement network in this project was designed to match the geographic spacing of the National Digital Forecast Database grid. The secondary stations which monitor global horizontal irradiance and ambient temperature have been installed as close as possible to the centroids of the NDFD grid cells. This placement was made possible by using SMUD utility poles as installation locations. With nearly 140,000 distribution poles in the service territory, suitable locations near NDFD grid cell centroids were identified for nearly all grid cells. The population of poles was surveyed to eliminate those with significant shading obstructions and the nearest pole to the centroid was selected. Specific pole selection was done by visually inspecting candidate poles near a centroid to determine whether there were any shading obstructions, and whether there was adequate climbing space on another

quadrant of the pole for SMUD linemen. Portions of SMUD's service territory are in rural areas and in some cases the density of poles did not permit placement close to the centroids. In other cases undergrounded utilities, which now make up nearly 65% of SMUD's circuit miles meant there were relatively few poles in certain urban areas. In all, 59 of the units are within 500m or less of the centroids. However of the remaining 12 units, five are over 1km from the centroids.

3.1. Near Field Shading & Data Filtering

In addition to the issue of proximity to the NDFD grid centroids the experimental design has other practical limitations and compromises. To eliminate shading, mounting the devices on the tops of the distribution poles would have been ideal. However, access to the tops of the poles is difficult and would have required specialized equipment and trained crews, increasing the safety risks and costs of the installation significantly. Instead, as a compromise, the monitoring units were located at the approximate midpoint of the poles. Consequently during portions of the year when the solar zenith angle is relatively small the overhead wires and cross arms briefly cast shadows on the pyranometer. These shadow events present as very discrete and repeatable anomalies on clear sky days, however through the year they move in somewhat difficult to predict patterns depending on the nature of the specific shade element. For instance, some shade elements are distribution wires with differing levels of tautness, where the sun crosses at a different point on the wire each day, creating shifts in time and slight shifts in shade width as the relative distance of the shade object changes throughout the seasons.

Figures 2 and 3 are images of global irradiance and temperature measured by unit 64 on two clear sky days two weeks apart. Note the characteristic shading events of the same magnitude at the same times of day. These shading anomalies, when viewed on clear sky days, represent a kind of signature for the monitoring unit in its unique location. Figure 4 is a Google Street View image of monitoring unit #64.

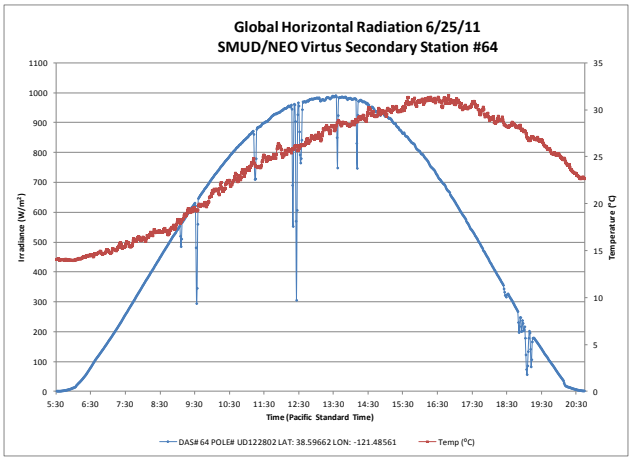


Fig. 2: Clear sky day 6/25/2011 station #64

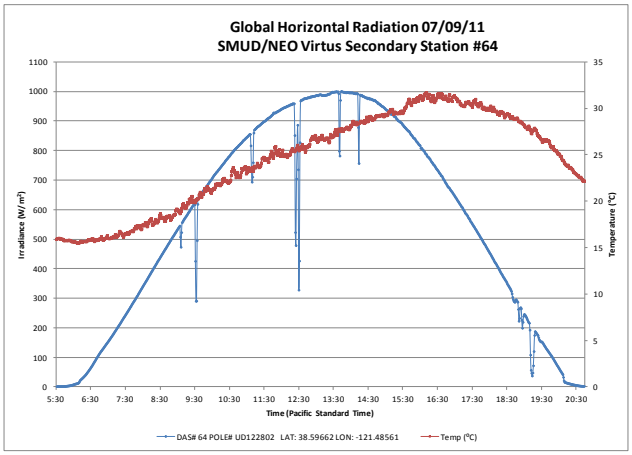


Fig. 3: Clear sky day 7/9/2011 station #64

Figure 4: Station #64 with overhead shading

On days with intermittent clouds these anomalies cannot be distinguished from real clouds. The locations of the anomalies during the course of the day and their magnitude change over time. Throughout the course of the year these events move and change in size. As the sun path becomes lower on the horizon many of these events disappear.

For this paper a subset of the data were used to confirm theoretical relationships developed by Hoff and Perez to represent variability between distributed PV systems. The first six months of data were reviewed and 28 days with high irradiance variability were chosen for the variability analysis and validation of the Hoff/Perez model (see below for list of reference days.) For all of the reference days a filtering algorithm was applied to remove GHI data points which are known to fall within the time windows of the site specific shading anomalies. This was done by mapping the characteristic anomalies from a closely occurring clear sky day to the high variability reference days. The data from the time periods coincident with the anomalies was removed and the missing data were flagged with -999.

For the statistical analysis of the larger dataset, a more refined filtering approach was developed to reduce the amount of buffer around a shade event that was eliminated, and automate the shade object identification process. The filtering algorithm identifies the beginning and end of days. It removes the data during nighttime periods. Using the remaining data, it performs a polynomial regression resulting in a best fit polynomial line. Days that have a high root mean square value, strong correlation within the data set, are identified as clear days. The equation for each clear day is manipulated to create a polynomial boundary that sits slightly below the data set. See Figure 5.

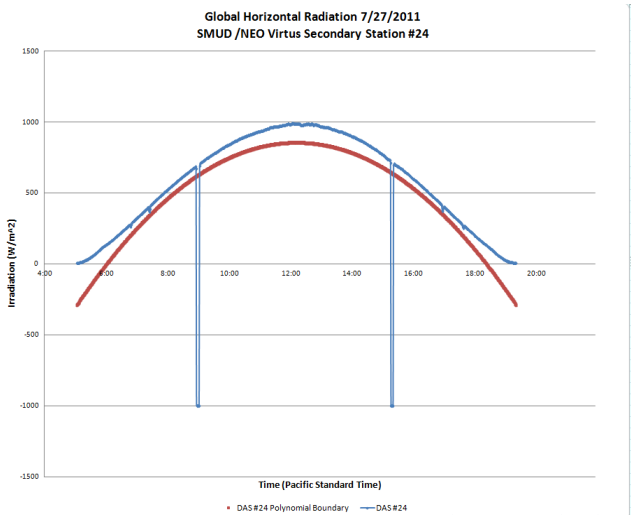


Fig. 5: Reference clear sky day 7/27/2011 Station #24

Any dips that go below the boundary are identified as shadows. The algorithm counts out the number of shadows that occur on the clear days. Sometimes the program with count a little higher or lower than the correct number shadows, in order to compensate, the algorithm will look at the all the occurring shadows for a station and take the most occurring number of dips for reference. The algorithm then will find the closest clear day to a series of cloudy days. It uses that clear day as a reference point for removing shadows from the cloudy days.

The algorithm uses a similar triangles method to pinpoint the predicted locations of the shadows on cloudy days. Due to the complexity involved with developing a complete 3-D model of each individual site, a simple 2D approach was used. Taking conservative approach, three minutes worth of data was removed from each side of the identified shadow to make sure that the complete shadow was removed. A resulting edit of a cloudy can be viewed in Figure 6. Notice that the gaps are larger than in the previous clear day because of the buffering.

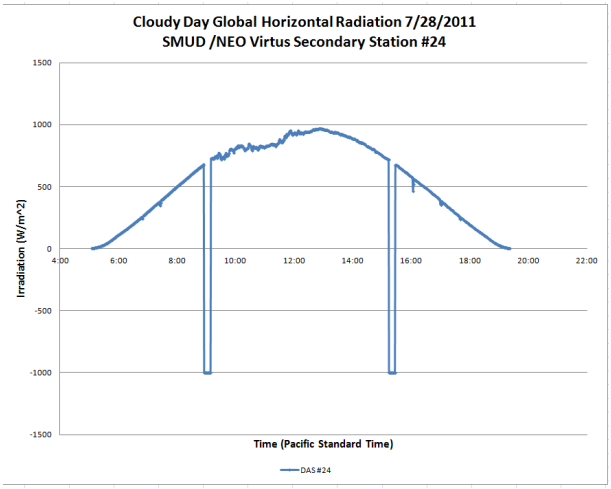


Fig. 6: Edited cloudy day 7/28/2011 Station #24

SMUD will continue to develop algorithms to minimize the amount of lost data associated with these shade objects, however for this analysis, the small buffer approach is adequate.

3.2. Sensor Error and System Wide Calibration

All of the pyranometers used in this project were Licor 200SZ devices. All were new when deployed and their 2 year calibration spanned the term of the project. However a limitation of the experimental design is the inability to clean

the pyranometers once they are installed on the poles. To address this issue four Eppley Precision Spectral Pyranometers (PSP) were installed on four of the RSR primary stations. The four PSPs are located approximately in the four compass points of the map of the monitored area.

Prior to installation all four of the PSPs were calibrated in the National Renewable Energy Laboratory's (NREL) Broadband Outdoor Radiometer Calibration (BORCAL) process[1].

Before publishing the data we will develop a correction procedure using the PSP data. The procedure will:

- locate a very clear sky day in the data record
- based upon BORCAL information correct known measurement errors inherent in the PSPs.
- average the BORCAL corrected clear sky measurements from the four PSPs and take the averaged value as our reference
- compare the PSP average to each of the 66 GHI values at solar noon and several other points in the day and derive a scale factor which will bring the GHI into agreement with the corrected PSPs
- "adjust" the data set for each GHI unit using the scale factors until the next clear sky day which follows the next scheduled cleaning of the PSPs
- Provide the raw secondary station GHI data, the raw PSP data as well as the BORCAL PSP corrections

Though the PSP data are available at this time this process was not done for the data set used in this paper.

4. VARIABILITY ASSESSMENT

A number of papers have been published by Hoff and Perez examining theoretical relationships between changes in output of dispersed PV systems[2,3,4]. The relationship developed demonstrates that the correlation between two sites decreases as time and distance increase, and can be described as a function of time, distance, and cloud speed. Validation of this theoretical relationship has been performed on a number of datasets for longer and shorter timeframes, though the spacing of this network represents a scale and density not available in previous ground-based monitoring evaluations.

The analysis was performed on the one-minute data for 28 high variability days (listed Table 1) at 61 locations. The analysis was performed as follows:

- Convert each 1-minute global horizontal insolation (GHI) observation at each of the 61 sties to a clearness index by dividing each GHI measurement by the clear sky irradiance for that minute.
- Calculate the change in the clearness index for each minute at each site using a 1 minute time interval.
- Calculate the correlation coefficient of the change in the clearness index using 15 time intervals per time period (i.e., the 1-minute time interval had a 15-minute time period) for each time period for each of the 1,830 unique site/pair combinations.[1]
- Weight the correlation coefficient by GHI variance to calculate a single correlation coefficient for each of the 1,830 site/pair combinations.
- Repeat the analysis using a 5-minute time interval.
- Repeat the analysis using 1-minute data.

The results are presented in Figure 7. The blue circles represent the 1,830 site/pair combinations. The red line is the binned average using a 2 km distance interval. The black line is CPR's patent-pending model using an average cloud speed of 13.8 meters per second.

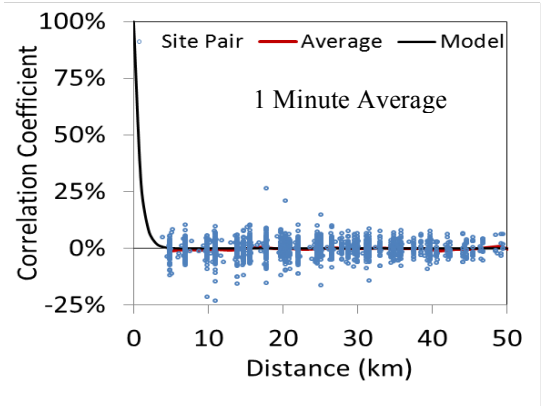
Several observations can be made. First, all 1,830 site/pair combinations follow a pattern of declining correlation as a function of distance. Second, there is no correlation between any two sites when the time interval is 1-minute; there is some correlation at a distance of 10 km when the time interval is 5-minutes. Third, CPR's model matches actual data for this network for time intervals of 1-minute and 5-minutes.

[1] There are a total of 61^2 possible combinations. 61 of the combinations are between the site and itself (i.e., the correlation coefficient always equals 1).

Of the remaining 61^2 – 61 combinations, only half of them are unique (i.e., correlation for the combination of site 1 and site 2 is the same as the correlation for the combination of site 2 and site 1). Thus, there are (61^2 – 61) / 2 = 1830 unique, meaningful combinations.

Table 1. Days included in analysis.

2011-07-13 00:00:00.000
2011-07-16 00:00:00.000
2011-07-17 00:00:00.000
2011-09-11 00:00:00.000
2011-09-12 00:00:00.000
2011-09-24 00:00:00.000
2011-09-25 00:00:00.000
2011-10-03 00:00:00.000
2011-10-05 00:00:00.000
2011-10-06 00:00:00.000
2011-10-11 00:00:00.000
2011-11-03 00:00:00.000
2011-11-04 00:00:00.000
2011-11-06 00:00:00.000
2011-11-07 00:00:00.000
2011-11-17 00:00:00.000
2011-11-18 00:00:00.000
2011-11-19 00:00:00.000
2011-11-20 00:00:00.000
2011-11-26 00:00:00.000
2011-11-28 00:00:00.000
2011-11-30 00:00:00.000
2011-12-15 00:00:00.000
2011-12-19 00:00:00.000
2011-12-23 00:00:00.000
2011-12-28 00:00:00.000
2011-12-30 00:00:00.000
2011-12-31 00:00:00.000



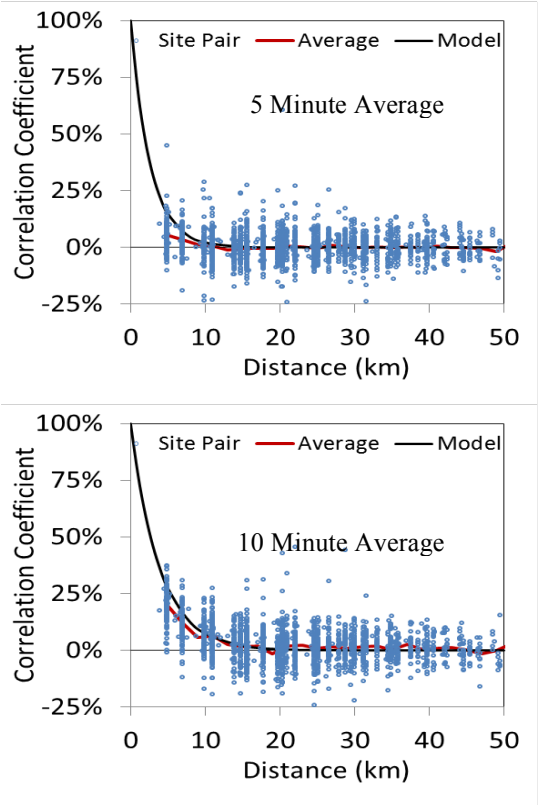


Fig. 7: Change in clearness index correlation coefficient vs. distance.

5. GROUND SATELLITE IRRADIANCE DATA COMPARISON

One of the objectives of deploying the ground-based solar monitoring network was to evaluate the satellite based irradiance data against the ground data to get a better understanding of the differences between the two. While the comparison is presented as % error, it is not clear whether the satellite or the ground record can be taken as the more accurate record for purposes of estimating impacts on a utility scale plant. Ground based sensors have a point perspective while satellite based approaches aggregate irradiance over an area, in this case 1km. The result of the aggregation is that the intensity of the dips due to clouds are substantially mitigated for smaller clouds. Depending on the size of the PV array that is being modeled, point-based data will tend to exaggerate variability relative to satellite based data. This effect is particularly noticeable for shorter time increments, and is mitigated as time increments grow. Beyond this difference, both methods face their own unique biases and errors that are well documented in the literature. With those caveats, we analyze the two datasets against one another to evaluate the differences.

For assessing GHI differences, data from 61 sensors that met quality screening criteria were compared using 1-minute data resolution compared to 1 minute, 1 km square High Resolution SolarAnywhere® satellite derived data. The differences were compared for 1 minute, half hour, hourly, daily, and 6 months, after filtering out any data that appeared to be influenced by shadows or poor irradiance sensor performance. Based on the assessment, for GHI, the satellite data was found to have an MAE of approximately 6.5% for 1-minute resolution compared to the ground data. This MAE narrowed as shown in Figure 8 for larger time increments, as the spatial averaging of the satellite data was better matched to the cloud patterns moving in time over the relevant irradiance sensor on the ground. This newly available high-resolution dataset error matches well to previously published Standard Resolution and Enhanced Resolution datasets from SolarAnywhere.

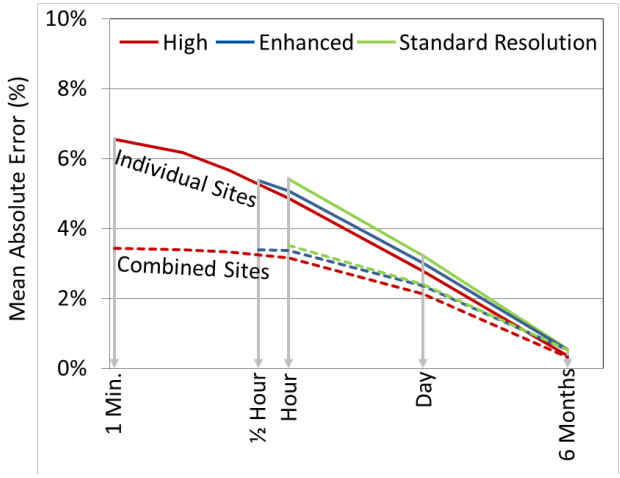


Fig. 8: Comparison of MAE for Global Horizontal Irradiance between satellite and ground measurements at 61 sites for varying time intervals

The use of the MAE % based metric, which uses MAE/average has been suggested by Hoff, et.al. in their draft paper on the topic[5] as a means of most effectively conveying error in irradiance measurement and forecasting to a non-solar research audience. In addition to evaluating GHI error, the presence of 8 Rotating Shadow-band Radiometer sensors in the network allowed the comparison of Direct Normal Irradiance data derived from satellite data by SolarAnywhere against ground measurements. The data were collected between July 1st 2011 and December 31st 2011, though some of the data were removed due to shadowing or errors in the telecom equipment leading to bad values.

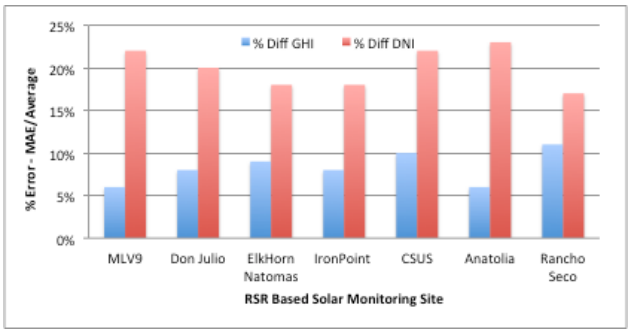


Fig. 9: Percent Error between satellite and ground data, MAE/Average method

As the 7 DNI datasets were evaluated separately from the 61 GHI datasets, global calibration was not applied to the RSR sensors, so instead, average bias is reported in Figure 10 for each of the stations. We expect that adjustment for this bias error in the MAE assessment would have reduced the reported error, however given the different direction of the errors and the intricacies of the DNI calculation, this adjustment was not performed for this paper.

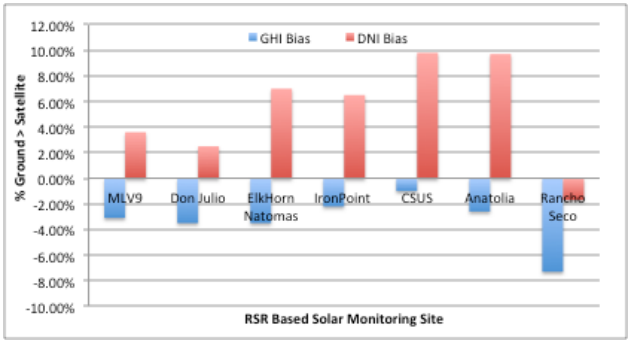


Fig 10: Percent Bias Between Ground and Satellite Irradiance Data, Negative Indicates Ground Higher than Satellite

6. CONCLUSION

The deployment of 71 irradiance sensors covering SMUD's service territory has allowed a unique opportunity for validating satellite based SolarAnywhere irradiance measurements as well as theoretical variability relationships developed to by Hoff and Perez to characterize variability correlation between pairs of dispersed PV systems. The pole-mounted sensors allowed for deployment in a grid-like pattern to closely align with NDFD weather forecasts for future work in validating PV forecasts by NEO Virtus Engineering based on the NDFD forecasts. Further, they

provided a secure location for the devices to operate autonomously with remote cell-modem based data collection. However, they also presented unique challenges in that there were shade objects present at each location resulting from distribution wires, cross-arms, and trees or buildings. Two methods for eliminating these shade objects were developed, and further work in refinement of automated methods continue to ensure the maximum amount of this data is usable.

The results of the validations were that for GHI, the MAE between satellite and ground data was 6.5% after data filtering and calibration of the ground dataset. By comparison, the DNI MAE averaged 18%, but with an average bias of 5% greater irradiance measurement from the ground-based sensors.

The dataset was also used to confirm previous theoretical work done by Hoff and Perez to demonstrate relationships between dispersed PV systems. The fundamental basis for this theoretical relationship between dispersed systems is that correlation the change in output between sites decreases as a function of increased distance between sites, shorter timeframes, and lower cloud speed. In other words, two PV systems change in output will have reduced correlation the further apart they are. The correlation will also tend to increase over longer timeframes as weather systems have the opportunity to impact both systems in a longer time increment (1 hour rather than 1 minute). Finally, due to the relationship between time and distance, the correlation will tend to be higher for shorter time increments if the cloud speed is higher, as again, the same weather conditions are more likely to impact two separated sites in a given time increment if the clouds are moving faster rather than slower. The associated algorithm to describe this relationship was validated using 1 minute, 5 minute, and 10 minute average data, providing validation for a geographic region larger than previous high time resolution validations and at finer time resolution than previous large geographic scale evaluations.

(1) Wilcox, S.M, Myers, D.R. “Evaluation of Radiometers in Full-Time Use at the National Renewable Energy Laboratory Solar Radiation Research Laboratory” December 2008, NREL NREL/TP-550-44627
(2) Tom Hoff, Richard Perez "Quantifying PV Power Output Variability", Solar Energy, Volume 84, Issue 10, October 2010, Pages 1782-1793.
(3) Tom Hoff, Richard Perez "Modeling PV Fleet Output Variability", Solar Energy, Forthcoming. Draft available at http://www.cleanpower.com/Content/Documents/research/capacityvaluation/Modeling%20PV%20Fleet%20Output%20Variability.pdf.
(4) Tom Hoff, Richard Perez "Parameterization of Site-specific Short-term Irradiance Variability", Solar Energy, Volume 85, Issue 7, July 2011, Pages 1343-1353,
(5) Tom Hoff, Richard Perez, Jan Kleissl, Dave Renne, Josh Stein, “Reporting of Relative Irradiance Prediction Dispersion Error” Draft Whitepaper, December 2011