California Solar Initiative

RD&D: Research, Development, Demonstration and Deployment Program

GO solar CALIFORNIA

Final Project Report:

Improving Economics of Solar Power through Resource Analysis, Forecasting, and Dynamic System Modeling

Grantee:

University of California, San Diego

November 2013

www.CalSolarResearch.ca.gov

PREPARED BY

University of California, San Diego

5998 Alcalá Park
San Diego, CA 92110

Principal Investigator:
Jan Kleissl
kleissl@ucsd.edu
858-534-8087

Project Partners:
Clean Power Research
Enernex
California Independent System Operator

PREPARED FOR

California Public Utilities Commission
California Solar Initiative: Research, Development, Demonstration, and Deployment Program

CSI RD&D PROGRAM MANAGER

Program Manager:
Ann Peterson
Ann.Peterson@itron.com

Project Manager:
Stephan Barsun
Stephan.Barsun@itron.com

Additional information and links to project related documents can be found at
http://www.calsolarresearch.ca.gov/Funded-Projects/

Preface

The goal of the California Solar Initiative (CSI) Research, Development, Demonstration, and Deployment (RD&D) Program is to foster a sustainable and self-supporting customer-sited solar market. To achieve this, the California Legislature authorized the California Public Utilities Commission (CPUC) to allocate **$50 million** of the CSI budget to an RD&D program. Strategically, the RD&D program seeks to leverage cost-sharing funds from other state, federal and private research entities, and targets activities across these four stages:

- Grid integration, storage, and metering: 50-65%
- Production technologies: 10-25%
- Business development and deployment: 10-20%
- Integration of energy efficiency, demand response, and storage with photovoltaics (PV)

There are seven key principles that guide the CSI RD&D Program:

1. **Improve the economics of solar technologies** by reducing technology costs and increasing system performance;
2. **Focus on issues that directly benefit California**, and that may not be funded by others;
3. **Fill knowledge gaps** to enable successful, wide-scale deployment of solar distributed generation technologies;
4. **Overcome significant barriers** to technology adoption;
5. **Take advantage of California's wealth of data** from past, current, and future installations to fulfill the above;
6. **Provide bridge funding** to help promising solar technologies transition from a pre-commercial state to full commercial viability; and
7. **Support efforts to address the integration of distributed solar power into the grid** in order to maximize its value to California ratepayers.

For more information about the CSI RD&D Program, please visit the program web site at www.calsolarresearch.ca.gov.

Abstract

The University of California, San Diego (UC San Diego) and its partners developed and validated solar resource models and forecasting tools to support the goals of the California Public Utilities Commission California Solar Initiative (CSI) Research, Development, Demonstration and Deployment (RD&D) Program. High-resolution solar resource models developed by CSI grantee Clean Power Research were validated against ground measurements and found to be slightly biased high, but generally of unprecedented and high accuracy. Additional solar resource products were developed that allow accounting for effects of topographic shading and soiling on solar photovoltaic (PV) power output, but both impacts were found to be small.

An integration study on a distribution feeder in SDG&E territory with high solar penetration was carried out using highly temporally and spatially resolved solar resource data. It was demonstrated that energy losses decrease, but voltage impacts increase with greater PV penetration. Spatially disaggregated (and more realistic) PV power output data reduced the impacts on the distribution system indicating that previous integration studies that used aggregated data may have overstated the impacts of PV. Considering the need for highly resolved data, a high resolution PV output dataset for over 115 PV sites spread across California was developed to facilitate other integration studies.

A PV performance model was developed that can be applied to satellite solar resource data to simulate PV power output taking into account local weather conditions. Using measured distributed PV power output from the CSI Performance-Based Incentive program aggregated across the California Independent System Operator (CAISO) territory, ramping was most significant during very predictable sunrise and sunset periods. Weather induced ramps were small over the CAISO level, but can be significant for smaller utilities such as San Diego Gas & Electric. Generally, these findings showed that for load balancing over large areas the impact of distributed PV generation are smaller than previously thought and often highly predictable.

Finally, for solar forecasting, satellite forecast models were compared against numerical weather prediction to examine how to best integrate the different models into seamless solar forecast products. For very short term forecasting, cloud speed sensors and algorithms were developed that can enhance the ability to predict and mitigate short-term fluctuations in solar power output.

Contents

Executive Summary

The University of California, San Diego (UC San Diego) and its partners developed and validated solar resource models and forecasting tools to support the goals of the California Public Utilities Commission California Solar Initiative (CSI). This final report provides a summary of work done under the CSI Research, Development, Demonstration and Deployment (RD&D) grant and more detailed results can be found in the individual deliverables.

A 1 km high resolution solar resource dataset (SolarAnywhere, SAW) based on satellite data was developed by Clean Power Research under another CSI RD&D grant. The research team compared the SolarAnywhere estimates to solar irradiation measurements at 53 ground stations throughout California. At a single high quality ground site, SolarAnywhere was unbiased, but it overestimated ground measured irradiation by 3.7 +- 0.9% at the other sites (Fig. ES1). A larger and more consistent overestimation at all sites was observed during May – July and also in clear conditions, indicating that improvements to the clear sky model would remove at least some of the difference. SAW random errors were found to be small. Overall SAW is the most accurate publicly available solar resource dataset.

The SolarAnywhere irradiances were also compared against 2010 measured power output at 192 PV systems in California obtained from the CSI performance-based incentive program (PBI) database. Clear sky indices were calculated from both datasets to allow comparing power output against irradiance. SAW was found to overestimate power output by 2.1% +- 0.2% (95% confidence interval) throughout the year. Since the main differences occurred in non-clear conditions (relative mean bias errors rMBE of 4.06%), the smaller conversion efficiency of PV systems at lower irradiation could explain some of the differences. The small relative Mean Absolute Error (rMAE) of SAW (4.22%) indicates its suitability as input for PV variability studies from scales of a feeder to a balancing area.

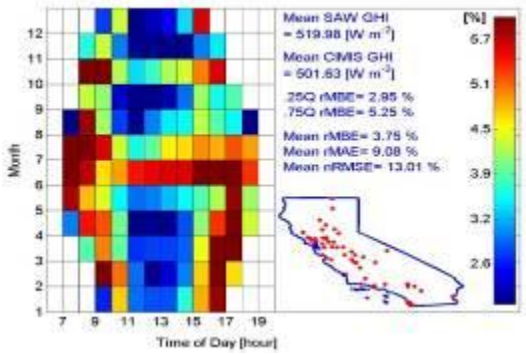


Fig. ES1a: Accuracy of SolarAnywhere: Calendar of the relative Mean Bias Error (MBE) by month and time-of-day for CIMIS ground stations versus SAW averaged over 52 stations. The caption indicates annual mean SAW and CIMIS GHI, 0.25 and 0.75 rMBE quantiles, annual mean rMBE, rMAE, and rRMSE. All relative errors are obtained by dividing by mean(GHI_{CIMIS}).

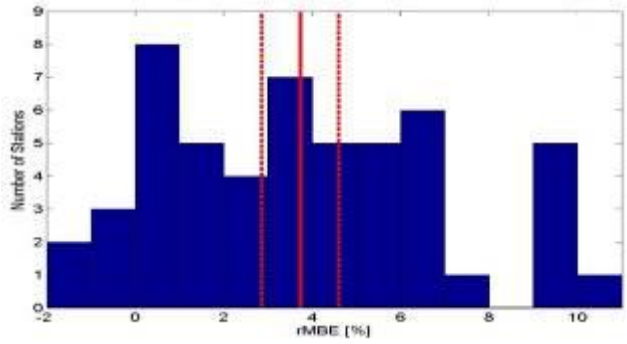


Fig. ES1b: Histogram of $rMBE_{year}$ for 52 high quality CIMIS stations. The red line shows the averaged rMBE and dashed red lines show the 95% confidence level.

Additional solar resource products were developed that allow accounting for effects of topographic shading and soiling on solar photovoltaic (PV) power output, but both impacts were found to be small. The effect of topography on the solar resource of Southern California and other urban areas with high PV penetration across the state was analyzed using the SolarAnywhere database and a Digital Elevation Model (DEM). The DEM allows calculating the altitude of the horizon from every point, with a spatial resolution of 1 arc-second (~30m). From the horizon the effect of topographic shading on the daily irradiation can be analyzed, effectively downscaling the satellite-derived irradiation database from 1km to 30m. All horizons were stored in a public database which can be easily integrated into PV performance models (Fig. ES2).

Topographic horizon calculations have been completed for all areas of expected high PV penetration and where horizon issues may exist (i.e., hilly terrain), in particular, all of Southern California below a latitude of 35^{o} (including Los Angeles, Orange County, and San Diego), and the San Francisco Bay area. This analysis does not consider buildings or other non-natural obstacles.

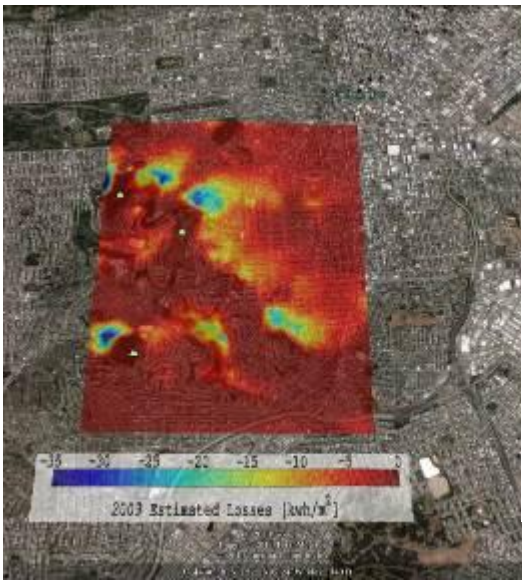


Fig. ES2: Yearly irradiation losses for an example area in San Francisco for a PV panel at south orientation and latitude tilt.

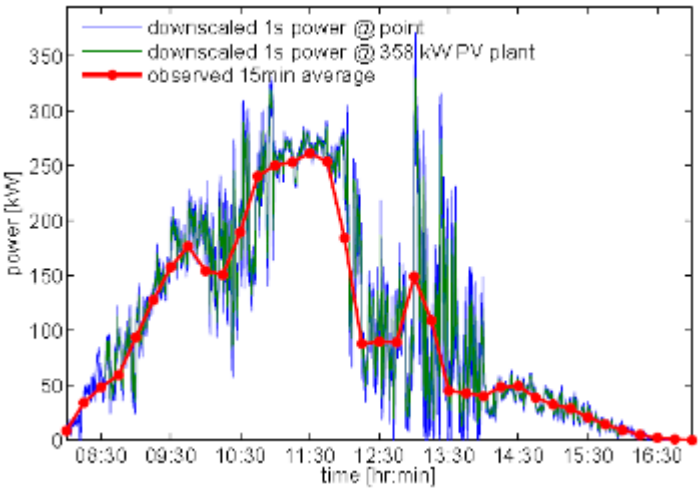


Fig. ES3: Downscaled and 15 min observed power output data for a 358 kW PV plant.

Soiling is the accumulation of dust on solar panels that causes a decrease in the PV system's efficiency. The changes in conversion efficiency of 186 residential and commercial PV sites were quantified during dry periods over the course of 2010 with respect to rain events observed at nearby weather stations and using SolarAnywhere solar resource data. Soiling losses averaged 0.051% per day overall and 26% of the sites had losses greater than 0.1% per day. Sites with small tilt angles (>5^{o}) had larger soiling losses (Fig. ES4) while differences by location were not statistically significant. Soiling was found to be smaller than in previous analysis, because more widely spread data was available through the CSI program. The soiling, topography, and validation work filled knowledge gaps to enable successful, wide-scale deployment of solar distributed generation technologies.

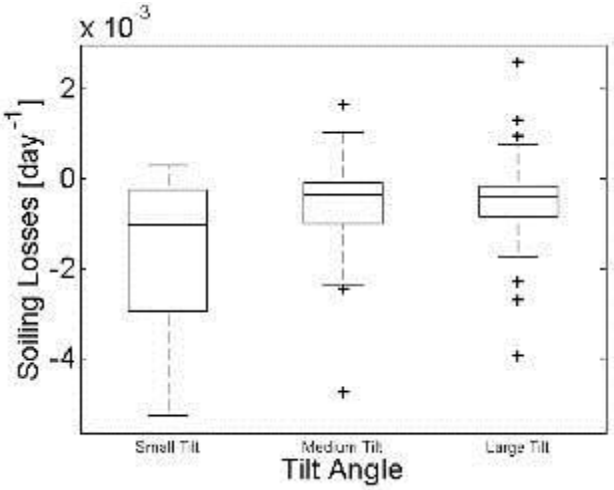


Fig. ES4 Soiling Losses: Box-Whisker plot of the distribution of soiling loss for different tilt angles.

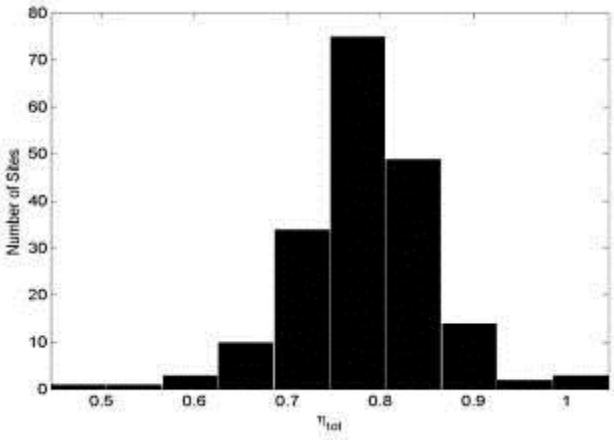


Fig. ES5: Histogram of relative PV solar conversion efficiency (η_{Tot}) for all 186 PV sites.

An integration study on a distribution feeder in SDG&E territory with high solar penetration was carried out using highly temporally and spatially resolved solar resource data. It was demonstrated that energy losses decrease, but voltage impacts increase with greater PV penetration. Spatially disaggregated (and more realistic) PV power output data decreased the impacts on the distribution system indicating that previous integration studies that used aggregated data may have overstated the impacts of PV. Considering this need for highly resolved data, a high resolution PV output dataset for over 115 PV sites spread across California was developed to facilitate other integration studies (Fig. ES3). In this way, the research supported efforts to address the integration of distributed solar power into the grid in order to maximize its value to California ratepayers.

Additionally, a PV performance model was developed that can be applied to satellite solar resource data to simulate PV power output taking into account local weather conditions. The model using SolarAnywhere data was compared to measured power output from 192 PV systems over a year. An average PV efficiency derate of 79% (loss of 21%) describes losses due to panel temperature, AC conversion, maximum power point tracking, and annual calibration (Fig. ES5). The bias error between modeled and measured power output was found to be less than 3% except near sunrise and sunset and mean absolute errors for 30 min data were less than 5%, which compared favorably to NREL's PVWATTS tool. The model is provided in MATLAB and can facilitate power conversion modeling for large datasets for variability or forecasting applications.

Using measured distributed PV power output from the CSI Performance-Based Incentive program aggregated across the California Independent System Operator (CAISO) territory, ramping was found to be most significant during very predictable sunrise and sunset periods (Fig. ES6 left). The goal was to quantify the largest aggregate ramp rates and evaluate how much on-line metering and telemetry of PV systems is necessary to track output of distributed generation for resource-adequacy applications. Over one year the largest hourly aggregate absolute ramp was a 30% increase and hourly ramps over 23% occurred only about once per day (all ramps are expressed as a fraction of Performance Test Conditions (PTC) rating). By investigating ramp rates of aggregate irradiance at clear sky conditions, the largest absolute ramp rates were found to be predominantly related to the rising and setting sun, and are therefore predictable. Weather-induced ramp rates with reference to a 30-day average of diurnal power output, on the other hand, can elucidate unexpected variations. The largest weather-induced ramp was

20% per hour and occurred due to a widespread decrease in cloud cover on a winter afternoon following the passage of a storm front (Fig. ES6 right). Other weather-induced ramps were 16% or less per hour. Due to large geographic diversity, weather induced ramps were small when aggregated over the CAISO territory, but can be significant for smaller utility areas such as San Diego Gas & Electric, where ramps up to 60% per hour were observed (Table ES1). The findings also confirm that satellite solar models such as SolarAnywhere can very accurately track the distributed solar generation in near-real time. This ability facilitates the accurate estimation of behind-the-meter generation and net load to improve operator visibility to solar ramps and possible risk conditions as they develop.

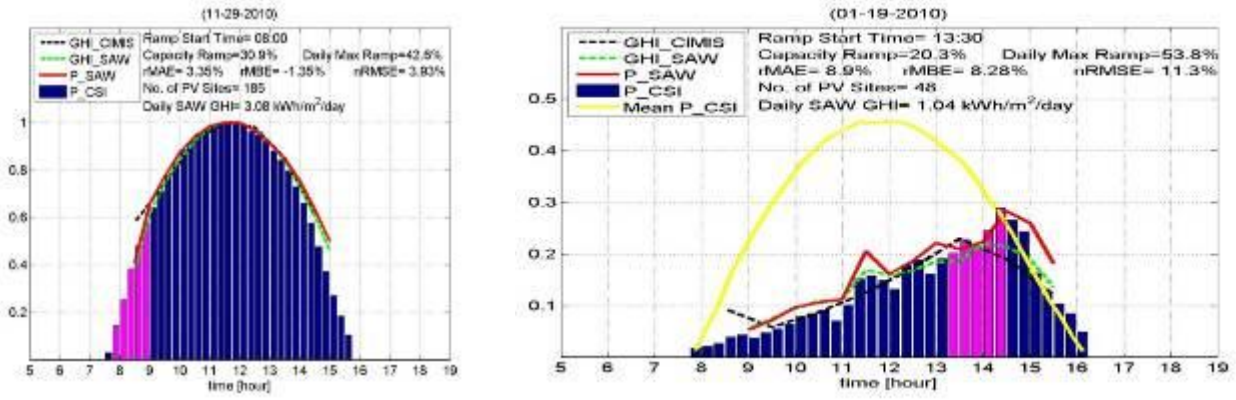


Fig. ES6: Days with largest absolute (left) and weather-induced (right) ramps: Normalized aggregate 15-minute PV output from all 192 PV sites (bars) for the days with the largest 1-hour ramp in 2010 (magenta bars show the timing of the largest ramp). Normalized aggregate SolarAnywhere and measured GHI of 39 weather stations (black) are also shown. See Section 10 for details on the legend and caption. (left) Nov. 29, 2010, (right) Jan. 19, 2010.

Table ES1: Comparison of the largest ramp rates for different utility territories.

	SDG&E	SCE	PG&E	CAISO
Largest absolute ramp	60.4%	31.4%	30.2%	30.9%
Largest weather induced ramp	55.5%	29.8%	28.1%	20.3%

Finally for solar forecasting, satellite forecast models were compared against numerical weather prediction to examine how to best integrate the different models into seamless solar forecast products. A method for solar forecasting using cloud motion vectors from satellite imagery with the ability to characterize forecast uncertainty has been developed. On average, the root mean square error (RMSE) for cloud motion vector (CMV) forecast increases with increasing forecast horizon and becomes larger than the North American Model (NAM, a numerical weather prediction model) forecast error at between 6 hours and 1 day. Consequently, satellite cloud motion vectors forecasts are superior for short time horizons and they are currently used as the model of choice in SolarAnywhere forecasts up to a 6-hour horizon. However, the forecast horizon at which the 'crossover' between cloud motion vectors and NAM occurs is dynamic (as short as 2 hours) and could be adjusted if the cloud motion vectors forecast certainty was known at forecast issue time. The RMSE of cloud motion vectors forecasts was most related to satellite image entropy and uniformity. The relative performance of the satellite compared to NAM could be predicted fairly accurately with two different modeling techniques. The models and metrics developed in this project can be applied to choose the optimal forecast model and

reduce solar forecast errors especially for hour-ahead forecasts. Accurate solar forecasting models will improve the economics of solar technologies.

For very short term forecasting, a cloud speed sensor and algorithms were developed that can enhance the ability to predict and mitigate short-term fluctuations in solar power output. Clouds are the dominant source of solar power variability and their velocity is a principal input to most short-term forecast models. Two methods are presented to estimate cloud speed using radiometric measurements from eight global horizontal irradiance sensors. Our analysis showed good agreement between both methods and nearby weather observations. Both methods require high variability in the input radiation as provided only in partly cloudy skies. The CSI grant provided bridge funding to help promising solar technologies transition from a pre-commercial state to full commercial viability. A provisional patent has been filed and funding to develop a commercializable prototype is being sought.

Through this project the project team took advantage of California's wealth of data from past, current, and future installations to fulfill the CSI goals. As byproduct of the various variability and forecasting models, several quality-controlled datasets were generated that can be made available to other researchers for validation.

- 2010 CIMIS data were quality controlled and results compiled into a report (see Appendix A).
- CSI Performance-Based Incentive data were quality controlled for effects not reflective of irradiance (see Appendix B).[SKB1]

Also, in summary the following models were developed:

- Horizon database for topographic shading analysis (Section 4).
- PV output data at high temporal resolution downscaled from CSI 15-minute average performance based incentive program data (Section 5.2).
- PV performance model (Section 7).

The PV performance model is posted at http://www.calsolarresearch.ca.gov/ and the horizon database and PV output data are available upon request from jkleissl@ucsd.edu. More details on access are provided in the relevant sections of this report and on the deliverables page on the calsolarresearch website.

Acknowledgements
This work was supported by the California Solar Initiative RD&D program. We are grateful to Stephan Barsun, Itron, for helpful comments, guidance, and for making connections with other CSI researchers. SolarAnywhere data for all of California was provided by Clean Power Research (Thomas Hoff and Skip Dise).

1 Introduction and Key Terms

The University of California, San Diego (UC San Diego) and its partners developed and validated solar resource models and forecasting tools to support the goals of the California Public Utilities Commission California Solar Initiative (CSI). This final report provides a summary of work done under this grant and more detailed results can be found in the individual deliverables.

Since several datasets and error metrics were used throughout different tasks, section 2 describes these data sets and equations. In particular, the SolarAnywhere satellite solar resource product, the California Irrigation Management Information System (CIMIS), and California Solar Initiative Performance Based Incentive Program (PBI) PV power output data are described. Further, the clear sky index and the mean bias error metric and confidence intervals are defined.

In Section 3 the SolarAnywhere satellite solar resource product is validated against the ground data sets described in Section 2. Since quality-controlled ground irradiance or power output data in California is difficult to obtain, this was the first attempt of a comprehensive examination of the ability of this satellite product funded by the CSI RD&D program to produce unbiased long-term solar resource projections across California.

A value-added product to satellite or other gridded solar resource mapping products is then described in Section 4. The topographic horizon database allows estimating solar energy generation losses due to topographic shading throughout Southern California and the San Francisco bay area.

High resolution solar resource data is then applied to a distribution feeder to examine the effects of high PV penetration on load tap changer operations, over-voltages, and losses. Since the high resolution data used here is generally not available, a method to temporally downscale solar irradiance or solar power data is described in Section 5.2 and applied to PBI data.

In Section 6 soiling losses for solar photovoltaic systems throughout California are derived from PBI data and stratified by panel tilt and location. A PV performance model is derived and calibrated with PBI data in Section 7.

A satellite cloud motion vector solar forecast algorithm is presented in Section 8 and the ability to project forecast errors based on features of the original image is analyzed. Cloud motion vectors are derived from ground measurements in Section 9 and the design of a cloud speed sensor is presented.

In Section 10 PBI data is analyzed to determine the largest ramp rates from aggregate solar generation. The ramp rates are classified into weather induced ramps and clear sky ramps, and the ramping behavior for different IOU territories and the CAISO territory are quantified for a whole year.

Acronyms and Key Terms

(see also NREL Glossary at http://rredc.nrel.gov/solar/glossary/)

α	PV panel temperature coefficient.
AC	Alternating current. Typically used to characterize inverter capacity at a PV site.
CIMIS	California Irrigation Management Information System: California-wide network of silicon pyranometer solar radiation sensors.
CMV	Cloud motion vector: cloud speed and direction.
DC	Direct current. Typically used to characterize PV panel capacity at a PV site.
DF	Day Fraction
DNI	Direct normal irradiance
η	Efficiency.
GHI	Global Horizontal Irradiance: sum of direct and diffuse irradiance on a horizontal surface.
GOES	Geostationary Operational Environmental Satellite
IOU	Investor-owned utilities (SDG&E, SCE, PG&E)
ISIS	Integrated Surface Irradiance Study: NOAA solar resource network with one station in California.
kt	Clear sky index: actual irradiance (or power output) normalized by expected clear sky irradiance or power output (Section 2.2).
MAE	Mean Absolute Error.
MBE	Mean Bias Error (Section 2.3).
METSTAT	The METeorological/STATistical solar irradiance model developed to produce the 1961-1990 National Solar Radiation Database
MPP	Maximum Power Point
NOAA	National Oceanic and Atmospheric Administration
PBI	Performance-based incentive program: Incentive program of the CSI, where payouts are based on actual solar generation.
PG&E	Pacific Gas & Electric
PV	Photovoltaic
PVWATTS	Online solar power performance calculator operated by the National Renewable Energy Laboratory
ρ	Correlation coefficient.
RMSE	Root Mean Square Error
RR	Ramp Rate
SAW	SolarAnywhere satellite-derived solar resource data.
SCE	Southern California Edison
SDG&E	San Diego Gas & Electric
SoDA	Solar Radiation Data available at http://www.soda-is.com.
SZA	Solar Zenith Angle
ToD	Time of Day
UTC	Universal Coordinated Timezone (PST = UTC – 8 hours).

2 Dataset and Error Metrics

Data

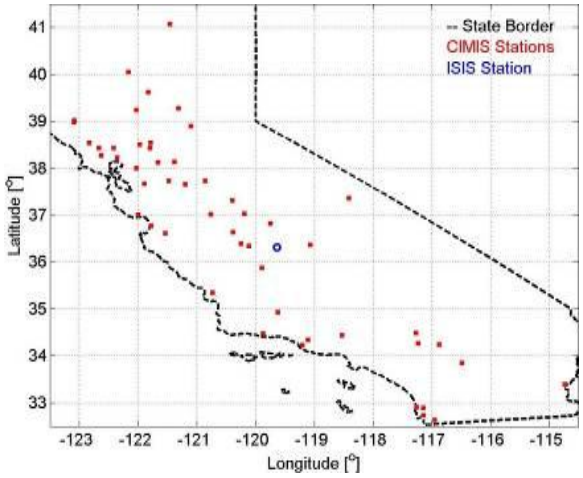
The main characteristics of the data used in various parts of the grant are summarized in Table 2-1.

Table 2-1: Overview of data processing for CSI and SolarAnywhere data to ensure comparability

Parameter	California Solar Initiative PBI	SolarAnywhere (SAW)
Location	Average over PV array	1 km^2 pixel over CSI site
Quantity	Clear sky index from CSI power output and clear sky modeled power	Clear sky index from SAW GHI and clear sky model
Time step	15 min (aggregated when compared to SAW or CIMIS)	30 min (native)
Calibration	Clear sky indices from SAW and CSI are calibrated to match in clear conditions.	

SolarAnywhere Satellite: Clean Power Research's commercially available SolarAnywhere (SAW) provides Global Horizontal Irradiation (GHI) and Direct Normal Irradiation (DNI) derived from Geostationary Operational Environmental Satellite (GOES) visible imagery at 30 minutes temporal and 1 km spatial resolution [1]. To obtain GHI, a cloud index is calculated for each pixel from the reflectance measured by the satellite. Instantaneous, spatially averaged GHI is then calculated by using the cloud index along with a clear sky model that considers local and seasonal effects of turbidity.

Perez et al. found a previous version of the SAW algorithm to have mean bias errors (MBE) between -5 and 15 W m^{-2} and root mean square errors (RMSE, based on hourly averages) ranging from 73-118 W m^{-2} when compared against high quality ground measurements sites across the US. Jamaly et al. validated SAW using ground measurements in 2010 at 52 California Irrigation Management Information System (CIMIS) stations and the NOAA Integrated Surface Irradiance Study (ISIS) network in Hanford, CA. SAW was unbiased compared to the Hanford ISIS data while SAW overestimated the measured GHI at CIMIS stations by 18.07 +- 4.15 W m^{-2} or 3.7% +- 0.9% (95% confidence interval), on average. SAW was also biased large in clear conditions compared to the Ineichen / SoDa clear sky model and the CIMIS measurements.



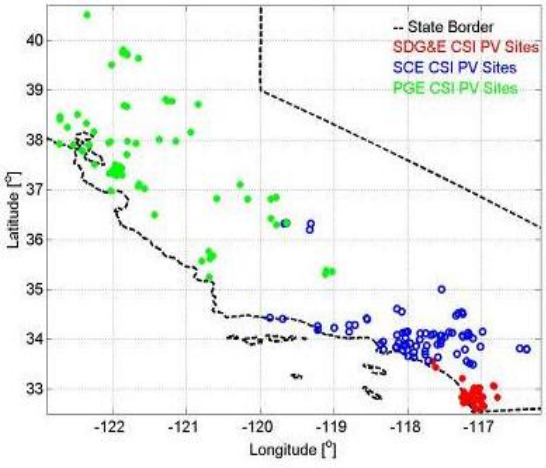


Fig. 2-1: Map of the (a) ISIS and 52 CIMIS stations in California and (b) 192 PV systems in SDG&E, SCE, and PGE territories

California Irrigation Management Information System (CIMIS): For comparison, ground measured irradiation data are analyzed; the California Irrigation Management Information System (CIMIS) with 124 active weather stations [2] and the NOAA Integrated Surface Irradiance Study (ISIS) network with one station in Hanford, CA [3].

CIMIS is operated by the Department of Water Resources (DWR) and has been operational since 1982 (CIMIS, 2009a). Each CIMIS station is equipped with a Li-Cor LI200S photodiode pyranometer, accurate under typical conditions to ±5% (CIMIS, 2009b). GHI is reported as an hourly average of 60 independent measurements within the hour [2].

The measured GHI of the ISIS in Hanford, CA and the 124 CIMIS stations are compared with the SAW GHI data of the corresponding pixel in which the stations are located. The analysis is conducted for yearly data in 2009 and 2010.

PV Data: The California Solar Initiative (CSI) rebate program requires a performance-based-incentive (PBI) payout for systems larger than 50 kW and makes it optional for smaller systems. This requires metering and monthly submission of 15 minute energy output to the payout administrator [4]. We have obtained the 2010 CSI measured output - quality controlled for system performance - for 194, 385, and 403 PV power plants in San Diego Gas & Electric (SDG&E), Southern California Edison (SCE), and Pacific Gas & Electric (PGE) territories (referred to as investor-owned utilities, IOUs) respectively.

The CSI database also includes street address and PV system specifications including DC Rating (kW_{DC}) at standard test condition (STC), AC Rating (kW_{AC}) at performance test condition (PTC), module and inverter models, inverter maximum efficiency, panel azimuth and tilt angles, and tracking type. The STC rating is obtained under idealized, controlled conditions of 1000 W m^{-2} plane-of-array irradiance and cell temperature at 25^{o}C while the PTC is developed in an attempt to simulate more realistic conditions at 1000 W m^{-2} plane-of-array irradiance with panel temperature derived from ambient air temperature at 20^{o}C and 1 m s^{-1} wind speed. Given the rapid increase in solar distributed generation (DG) in most coastal urban centers in California which are included in our study (e.g. Los Angeles, San Francisco, and San Diego) this dataset presents important information to PV system owners whether existing solar resource datasets can provide accurate estimates of irradiation (and as a result, power) in these areas.

Data Quality Control: CIMIS provides an initial QC assessment based on procedures described by Meek and Hatfield (1994), issuing flags that allow the user to remove any data that appears faulty or erroneous. These flags, detailed on the CIMIS website [2], restrict any data that contain obvious outliers or unphysical characteristics. CIMIS provides a further description of the QC method in the CIMIS technical manual. Luoma and Kleissl (2012) reviewed data from each CIMIS station individually; All flagged CIMIS data were excluded leaving 124 CIMIS stations with 70% or more available and high quality data. The same quality control is applied to the ISIS data and all flagged ISIS data (as described on NOAA website [3]) are not considered. Only data with solar zenith angle less than 75° are considered to avoid error in sensor cosine response and shading.

$rMBE_{year}$ (for the whole data) and $rMBE_{clear}$ (for the times with clear condition) between SAW modeled GHI and CIMIS and ISIS measured GHI is calculated. To exclude outliers, the CIMIS stations with $rMBE_{year}$

or $rMBE_{clear}$ out of the range of 0.25-0.75 quantiles are excluded. Therefore, 52 CIMIS stations, along with the ISIS station, are considered in this study (Fig. 2-1a).

PV Data: Quality control was used to exclude all CSI sites with at least one of the following characteristics not representative of irradiance: PV systems with hourly averaged (versus 15 min) data, more than 70% missing data (mostly because they were installed during 2010), significant noise or large spikes in power due to recording issues, decrease in power due to soiling, significant clipping of power due to undersized inverters, less than 5 distinct power output for the whole year, or plants divided into sub-arrays with different panel tilt and azimuth angles (Appendix A). Therefore, a final set of 192 PV systems are analyzed (Fig. 2-1b & Table 2-2). To avoid errors due to sensor cosine response and shading by nearby obstructions (not considered by SAW), only data for solar zenith angles less than 75° are considered.

Table 2-2: Statistics of CSI-PBI PV systems in SDG&E, SCE, and PGE territories that pass quality control (see Appendix A)

IOU	No. of PV systems	Total PTC rated capacity(MW)	Mean PTC rated (kW)	Median PTC rated (kW)
SDG&E	45	4.73	105.1	46.43
SCE	81	17.48	215.8	192.9
PGE	66	16.29	229.4	165.8

Clear Sky Index

Since PV power output is a function of many parameters not related to irradiation, a normalized metric that quantifies the local solar resource is desirable. For this reason, clear sky index (*kt*) is used here to intercompare irradiation (SAW) and power (CSI) data. *kt* is defined as

$$kt = GHI \,/\, GHI_{CSI} \qquad \text{Eq. (1-1)},$$

where *GHI* is the Global Horizontal Irradiance and GHI_{CS} is the *GHI* in clear sky conditions. According to the site longitude and latitude, 1-min GHI_{CS} is calculated based on the Ineichen model with Linke Turbidity from the SoDa database. GHI_{CS} is averaged over the corresponding CSI and SAW time intervals (15-min and 30-min respectively).

Error Metrics

Mean Bias Error (MBE) describes persistent differences between kt_{SAW} and kt_{CSI} and is calculated as

$$MBE = \frac{1}{N}\sum_{n=1}^{N}(kt_{SAW} - kt_{CSI}) \qquad \text{Eq. (1-2)},$$

where *N* is the number of samples. Also, the relative MBE is calculated as

$$rMBE = \frac{MBE}{mean(kt_{CSI})} * 100\% \qquad \text{Eq. (1-3)}.$$

The confidence Interval (CI) is calculated to determine the significance level of the difference between kt_{SAW} and kt_{CSI} as

$$CI = \widehat{MBE} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \quad \text{Eq. (1-4),}$$

where $\widehat{MBE}$ is the average kt_{SAW} bias error over all PV systems (kt_{CSI}), $Z_{\alpha/2}$ is the confidence level coefficient ($Z_{\alpha/2} = 1.96$ for 95% confidence level or α=0.05), σ is the standard deviation of MBE for all PV systems, and *n* is the total number of PV systems.

To illustrate diurnal or seasonal patterns in the data, bias errors are averaged for each time of day (ToD) and each month separately to yield MBE_{MT} as the difference between $kt_{CSI}(m, ToD)$ and $kt_{SAW}(m, ToD)$ for *m* = 1,…, 12 (months) and ToD = 1, …, 48 (30 min segments).

$$rMBE_{MT}(m, ToD) = \frac{MBE_{MT}(m,ToD)}{mean[kt_{CSI}(m,ToD)]} * 100\% \quad \text{Eq. (1-5).}$$

3 Validation of SolarAnywhere Enhanced Resolution Irradiation in California

Comparison of California-wide GHI Averages across the Year

The daily average of CIMIS, SAW, and clear sky GHI (based on the Ineichen model with Linke Turbidity from the SoDa database) averaged over 52 stations is computed for the year 2010. In Fig. 3-1, time series of daily averages through 2010 are shown. Generally, SolarAnywhere overestimates CIMIS irradiation data by 17 to 20 W m^{-2} or 3 to 5% throughout the year. The bias is somewhat larger for the clear periods at 21 to 31 W m^{-2} (not shown). During clear periods, SAW is larger than CIMIS measurements.

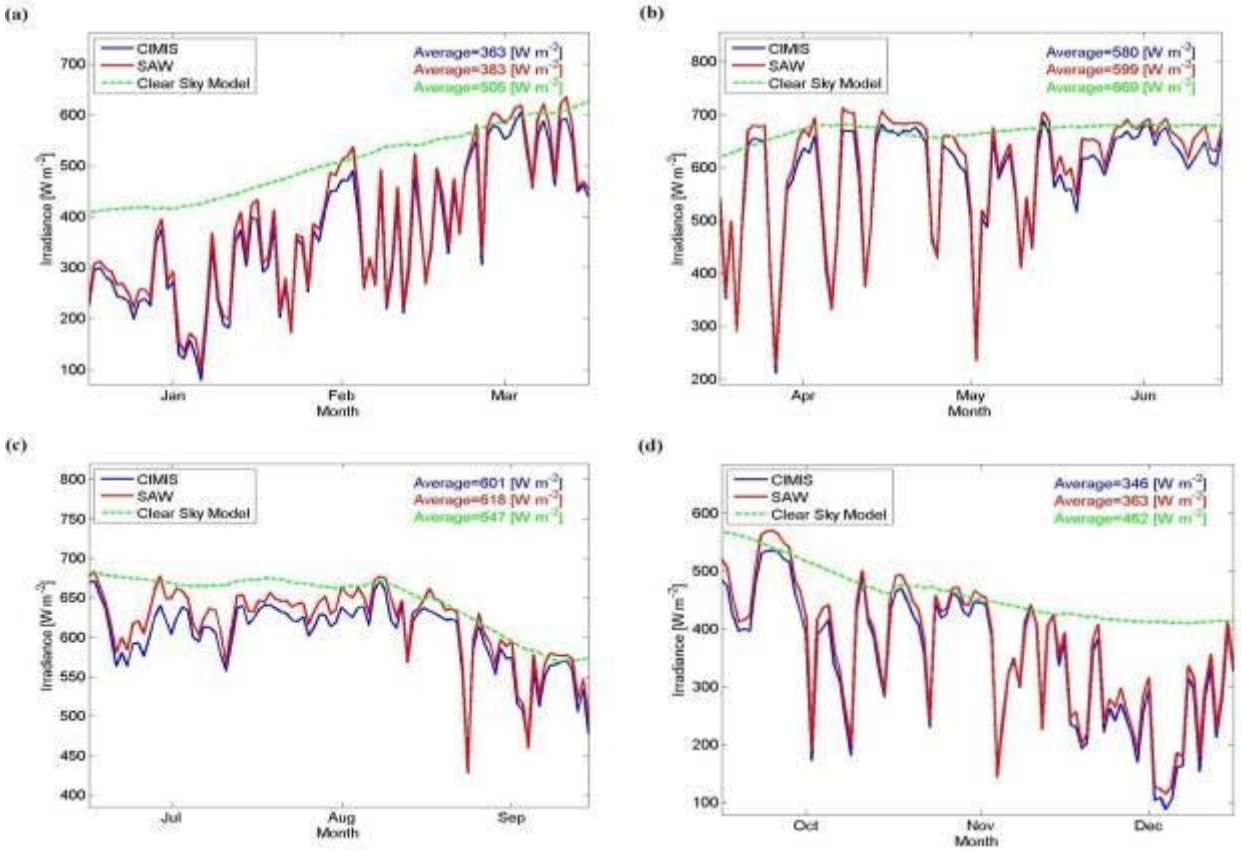


Fig. 3-1, Daily average CIMIS versus SAW: Mean daily (for SZA<75^{o}) irradiation averaged over 52 CIMIS stations and the collocated SAW pixels in 2010 broken out by season. The upper envelope of the CIMIS and SAW lines is expected to be similar to the clear sky model that is shown for reference. The day-to-day variability in the clear sky model is due to missing CIMIS data causing different CIMIS sites/ time steps to be chosen for each day.

The ISIS site in Hanford, CA is of special interest as data quality is expected to be higher and since SAW was calibrated and validated at this site before. The daily average of ISIS GHI and SAW GHI for the whole year and are shown in Fig. 3-2. The daily averages are calculated for the same available time steps of all

the datasets. There is excellent agreement between SAW and ISIS with seasonal biases of less than 2%. The only exceptions are July and August when consistent over- and underestimates of 2 to 5% are observed.

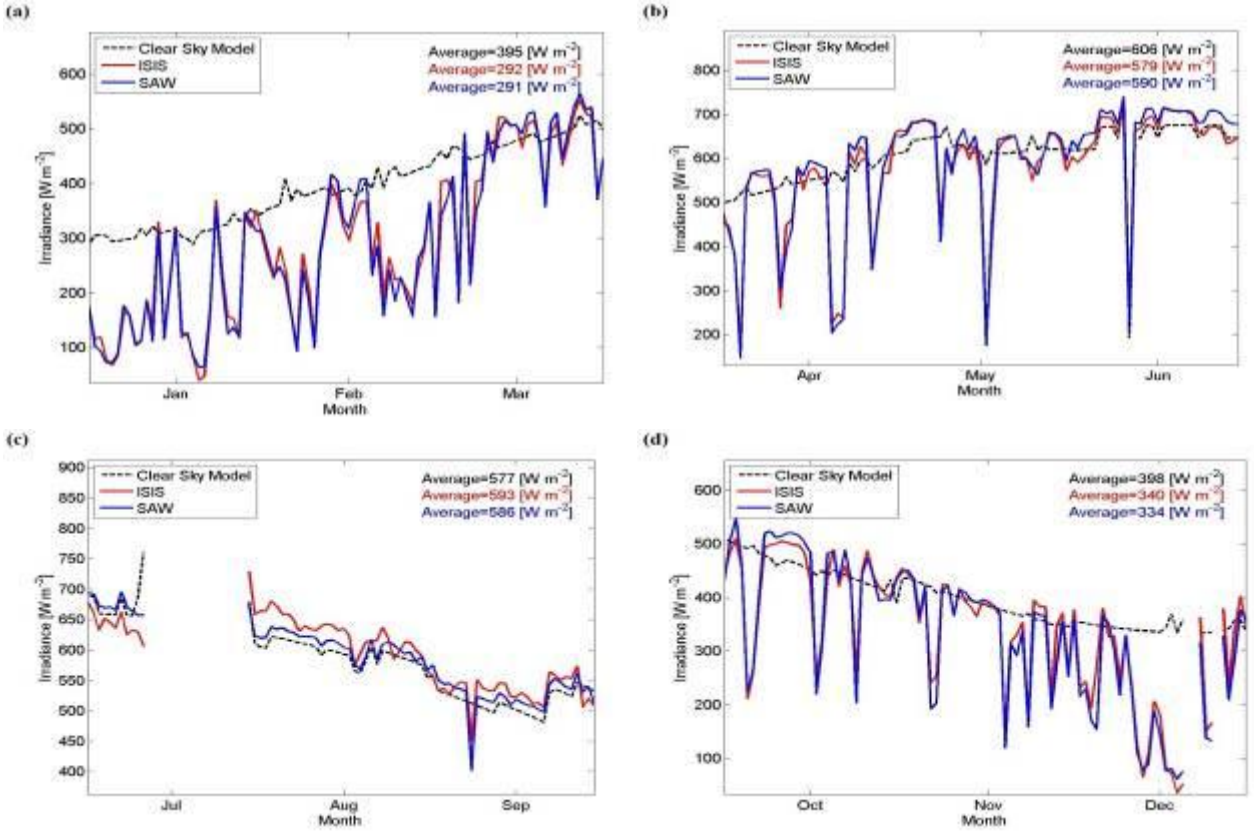
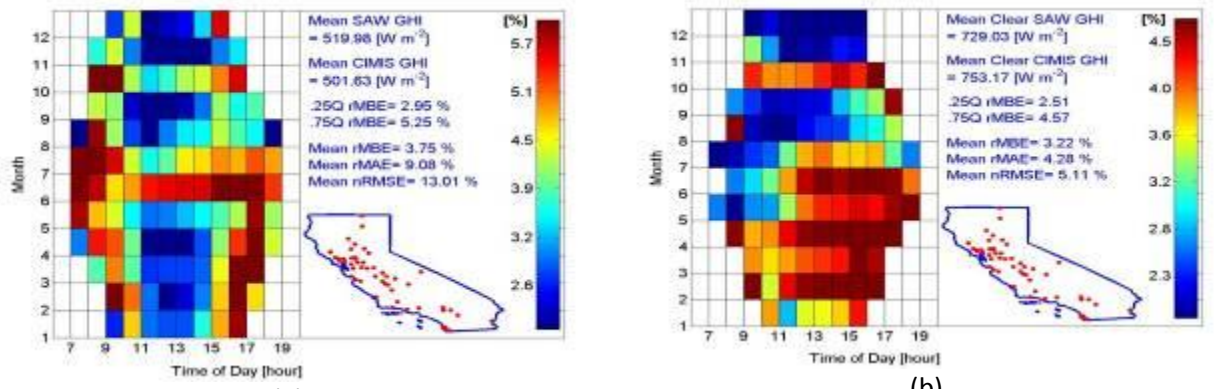


Fig. 3-2, Daily average Hanford ISIS versus SAW: Mean daily irradiation at the ISIS ground station in Hanford, CA for 2010. The day-to-day variability in the clear sky model is due to missing ISIS data causing different ISIS time steps to be chosen for each day.

Climatologies of Bias Errors by Month and Time-of-day

Averaged 2010 $rMBE_{MT}$ of all 52 CIMIS stations is shown in Fig. 3-3 (for both the whole dataset and the clear sky conditions). Overall bias error (SAW overestimates) of 18 W m^{-2} or 3.7% are observed, consistent with Fig. 3-1. MBE is 24 W m^{-2} or 3.2% in clear conditions. The biases are largest in June and to a lesser extent in May and July. For the rest of the year biases are less than 2% during midday, but larger in the (less important) morning and evenings. Clear sky biases are largest from February until June. The same trends are observed for 2009.

$rMBE_{MT}$ at the ISIS station in Hanford, CA (not shown) is qualitatively consistent with the overall trends. There is no bias on average over the year, but SAW over predicts in June and July by 3 to 5%.



(a) (b)

Fig. 3-3, rMBE by month and time-of-day for CIMIS versus SAW for (a) all and (b) clear conditions: $rMBE_{MT}$ of 2010 SAW and CIMIS data (averaged over 52 stations). (a) all data, (b) data in clear sky conditions. The caption indicates annual mean SAW and CIMIS GHI, 0.25 and 0.75 quantiles of $rMBE_{MT}$, annual mean rMBE, rMAE, and rRMSE. All relative errors are obtained from hourly data by dividing annual MBE, MAE, and RMSE by mean(GHI_{CIMIS}).

Averaged $MBE_{MT,rel}$ of the costal stations are shown in Fig. 3-4. The coastal and inland regions were also examined subdivided into northern and southern sub-regions, but no significant north-south difference was observed (not shown). The rMBE is larger at coastal stations with 5% on average. Largest rMBE is still observed in May and June, but the difference to the other months is less pronounced than in the California-wide data (Fig. 3-3).

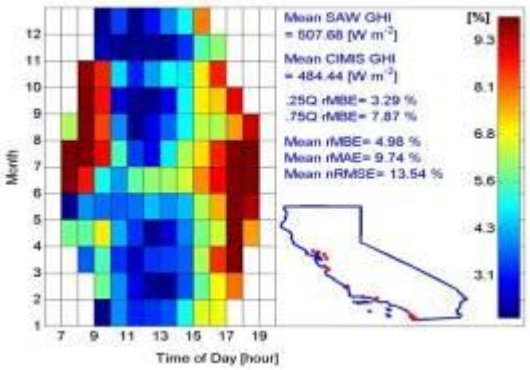


Fig. 3-4: Same as Fig. 3-3 but for coastal stations only (14 stations total).

SAW Validation against CSI PBI PV Sites

The monthly kt_{SAW} and kt_{CSI} averaged for all three IOUs is presented in Fig. 3-5. rMBE, rMAE, nRMSE between kt_{SAW} and kt_{CSI} for each month is shown in Table 3-1. The months with smaller average *kt* (the more cloud months of Jan., Feb., Oct., and Dec.) have larger random errors (rMAE and nRMSE). This is expected as cloud optical depth is more difficult to forecast and small phase-shifts in timing of cloud formation or evaporation can cause large random errors. The same result was observed for bias errors. This is related to the fact that the clear sky indices are calibrated to identical in clear conditions, on average. Therefore, the month with smaller average *kt* should have larger bias errors.

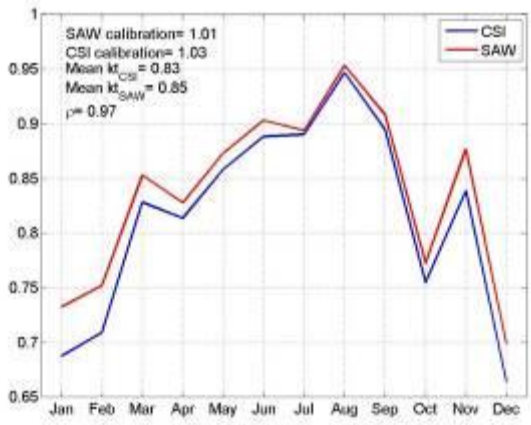


Fig. 3-5, Monthly average kt_{SAW} versus kt_{CSI}: Mean monthly (for SZA<75°) clear sky indices in 2010 for all three IOUs (averaged over 192 PV sites). The caption shows SAW and CSI calibration factor (C_{SAW} and C_{CSI}), average of kt_{SAW} and kt_{CSI}, as well as the correlation (ρ) averaged over 192 PV sites.

Table 3-1: Statistics of monthly average kt_{SAW} versus kt_{CSI} (averaged over 192 PV sites)

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
rMBE (%)	6.5	5.7	3.1	1.8	1.9	1.7	0.5	0.8	1.7	2.5	43	5.0
rMAE (%)	7.9	6.5	4.6	4.3	3.4	3.0	3.3	2.8	3.0	4.9	4.7	8.0
nRMSE (%)	10.2	9.2	6.1	6.3	5.3	5.1	4.8	3.8	4.4	6.8	6.2	10.3

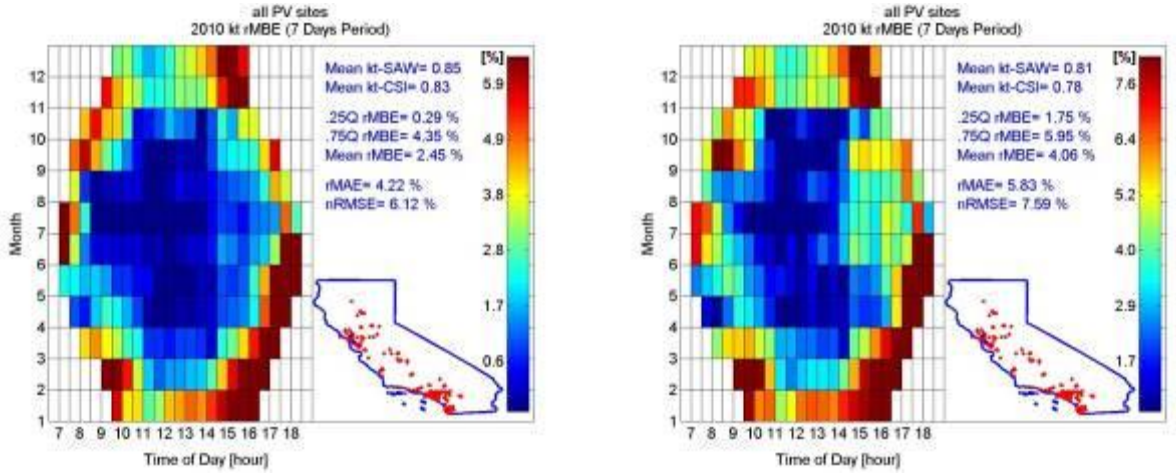


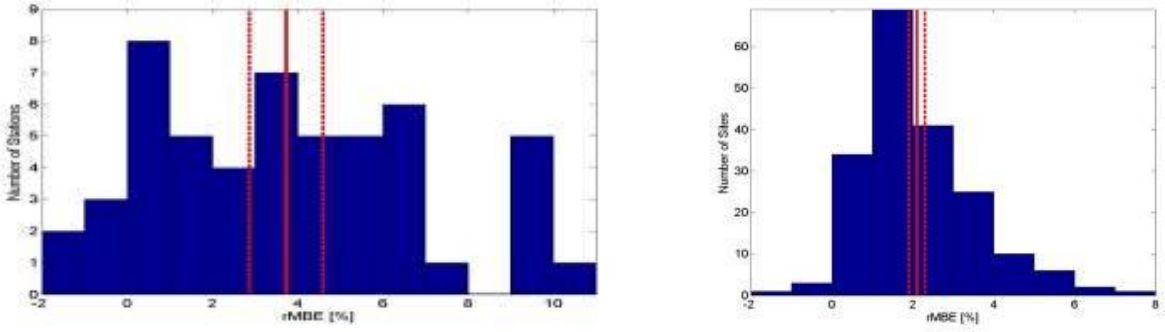
Fig. 3-6, rMBE by month and time-of-day for kt_{SAW} versus kt_{CSI} (averaged over 192 PV sites) for (a) all data and (b) non-clear conditions only. The caption indicates annual mean kt_{SAW} and kt_{CSI}, 0.25 and 0.75 quantiles of $rMBE_{MT}$, annual mean rMBE, rMAE, and nRMSE. All relative errors are obtained from hourly data by dividing annual MBE, MAE, and RMSE by mean(kt_{CSI}).

Averaged $rMBE_{MT}$ of all PV systems is shown in Fig. 3-6 (for both the whole dataset and the cloudy conditions); the trends are the same for all IOUs. The biases are less than 3% during midday, but larger in the (less important) morning and evenings. Fig. 3-6 suggests that the bias error is mostly independent of month (at least during midday) with small increases in January, November, and December.

Overall Bias Errors and Confidence Intervals

MBE_{year}, $rMBE_{year}$, MBE_{clear} and $rMBE_{clear}$ (averaged over the 52 high quality CIMIS stations) are 18.07 W m^{-2}, 3.74%, 24.38 W m^{-2}, and 3.86% with the corresponding confidence intervals of +-4.15 W m^{-2}, +-0.87%, 4.92 W m^{-2}, and 0.68%, respectively. Fig. 3-7 shows histograms of $rMBE_{year}$ and $rMBE_{clear}$ for all the 52 CIMIS stations, which confirms that the $rMBE_{year}$ and $rMBE_{clear}$ for most of the CIMIS stations are close to the respective averaged values. When all 124 CIMIS stations are considered, the errors and confidence intervals are only slightly larger than the ones for 52 stations.

Fig. 3-7b shows histograms of $rMBE_{year}$ and $rMBE_{clear}$ for all PV systems. The rMBE averaged over the year and all PV systems is 2.11% with a 95% corresponding confidence intervals of +-0.2%. The $rMBE_{year}$ and $rMBE_{clear}$ for most of the PV systems are close to the respective averaged values. No significant regional differences by IOU were observed (not shown).



(a) (b)

Fig. 3-7: Histogram of $rMBE_{year}$ for (a) 52 high quality CIMIS stations and (b) all 192 PV sites. The red line shows the averaged respective rMBE and dashed red lines show the confidence interval for the 95% confidence level.

Conclusions

Validations of the new state-of-the-art solar resource model for California (SolarAnywhere, SAW) were conducted using ground measurements. SAW is unbiased compared to the Hanford ISIS data (not surprising since the irradiation versus cloud index relationship was calibrated there). SAW overestimates the measured GHI at CIMIS stations by 18.07 +- 4.15 W m^{-2} or 3.7% +- 0.9% (95% confidence interval), on average. SAW is also biased large in clear conditions compared to the Ineichen / SoDa clear sky model and the CIMIS measurements.

Despite careful quality control by the authors, CIMIS stations have inferior sensors and are generally less well maintained than high quality solar resource sites such as ISIS. That may suggest that the differences between SAW & CIMIS are at least partially related to CIMIS measurement errors. Especially soiling of the ground sites due to infrequent cleaning likely explains some of the bias. However, persistent trends over the year likely indicate some underlying bias in SAW. Also for PV performance applications one could argue that soiling of PV panels will be even larger than for CIMIS sensors, so CIMIS measurements may be more reflective of performance of solar power plants. From our analysis the following recommendations emerge:

- The relative mean bias error ($rMBE_{MT}$, averaged over all 52 stations, Fig. 3-6) is largest in May through July. This is the most significant finding as it holds both for clear data and all data, 2009 and 2010, and for CIMIS and ISIS. The cause of this difference, however, is unclear.
- The SoDa turbidity climatology appears to be more accurate than the METSTAT turbidity database used in SAW, but the average observed CIMIS clear sky data lie in between (Fig. 3-1).
- $rMBE_{MT,}$ of the coastal stations is slightly larger than for inland stations. For the coastal stations the largest differences occurred in both morning (Jun.-Nov.) and evening (Mar.-Jul.). The morning differences could be related to marine layer clouds, while the evening differences show correlation to the clear sky model. However, overall the coastal differences are much smaller than those found for the National Solar Radiation Database indicating an improvement in SAW compared to previous versions.

Random differences were not a focus of this study. For completeness we report that typical rMAEs (for hourly data and normalized by average irradiation) were 9% and typical RMSEs were 13%. Overall the SAW solar resource data are very accurate both in bias and random error.

Additional validations of SAW were conducted using measured power output of PV systems throughout California (in SDG&E, SCE, and PGE territories). Clear sky indices were calculated using both datasets (kt_{SAW} and kt_{CSI}). kt_{SAW} was found to overestimate kt_{CSI} by 2.11% +-0.2% (95% confidence interval) throughout the year, on average. The rMAE for 30 min data was 4.22%. Since the kt_{SAW} and kt_{CSI} are calibrated based on the clearest days, the main differences between kt_{SAW} & kt_{CSI} occurred in non-clear conditions with a rMBE of 4.06%. In these conditions, the lower overall conversion efficiency of PV plants at lower irradiation could contribute to the observed differences. On the other hand, higher efficiency is expected at lower irradiation since the temperature is lower. This suggests that PV systems do not operate at maximum power point (MPP) condition at lower irradiation. Also, despite careful quality control by the authors, some of the PV systems are not well maintained. Consequently, some of the differences between kt_{SAW} & kt_{CSI} are related to reduced power output caused by soiling or topographic shading.

4 Downscaling Satellite Derived Irradiation using Topographic Shading

Horizon Calculation

Topographic shading causes reduced power output for PV systems in the mornings and/or evenings based on the sun's altitude being lower than local obstructions such as hills. The horizon (obstruction angle) at a given direction is calculated by finding the maximum elevation angle between the location and all ground points along that direction. The maximum angle found in a given direction determines the horizon in that direction. Once the horizon is calculated the shading analysis can be performed for a given day, by overlapping the sun path and the horizon (Fig. 4-1). From the horizon calculations, we obtain the actual local hours of sunrise and sunset for each day.

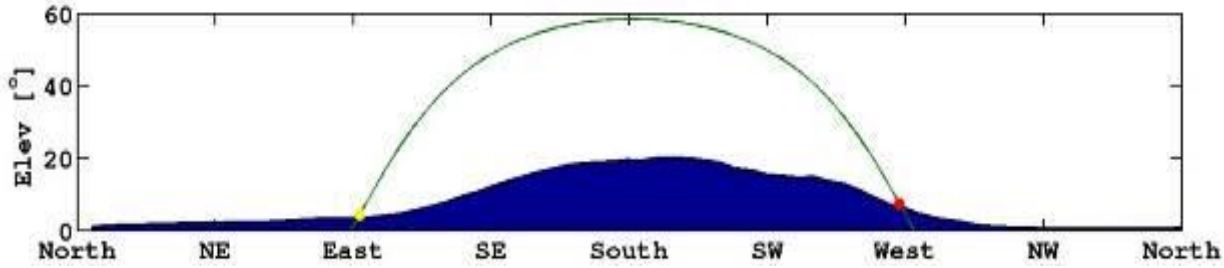


Fig. 4-1: Sunpath and horizon for an inland site in San Diego, CA (Latitude 33°, Longitude -117°) for March 25th. The actual local sunrise (yellow circle) is later and actual sunset (red circle) is earlier than if no obstructions existed.

Global irradiation on the PV panel can be calculated using the Solar Anywhere irradiation database as the sum of direct and diffuse irradiance incident on the plane-of-array. While the sun is obstructed by the horizon, the direct irradiance is zero, which constitutes a loss in power production due the topographic shading. Losses in diffuse irradiance are typically small and are ignored here. Fig. 4-2 shows an example of irradiation losses for a case with PV tilt equal to latitude and south orientation. For this day the shading is responsible for losses of 7.6% (1.3 MJ).

Solar Resource Mapping Example for San Francisco

The topographical shading correction allows a more accurate assessment of the solar resource and PV power production especially in hilly or mountainous areas. As an example, this method has been applied to an area in San Francisco's Twin Peaks neighborhood. Horizons and irradiation were computed for a ~4X4 km² area around Twin Peaks at ~30m resolution resulting in 20736 points.

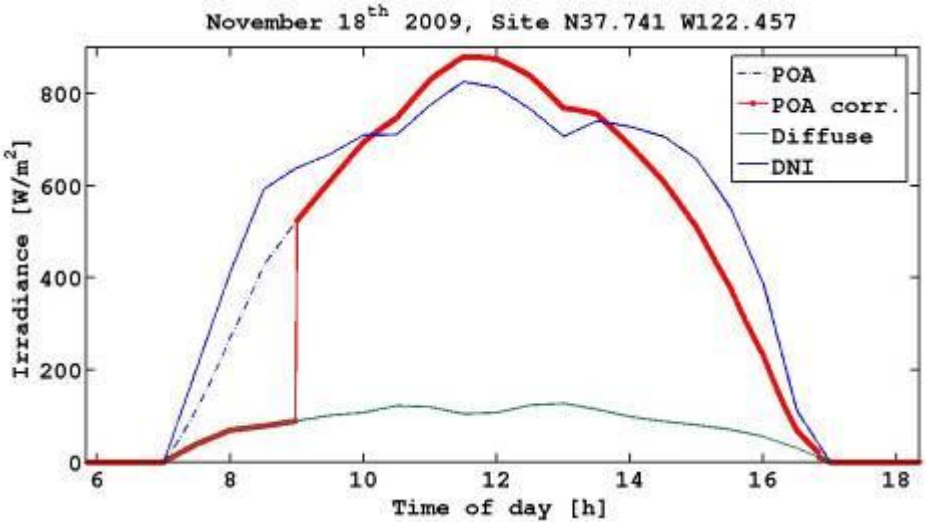


Fig. 4-2: Correction due to obstruction of direct irradiation by topography for a site in San Francisco (coordinates are given in the title). The red line shows the corrected plane-of-array (POA) irradiance and the dashed blue line shows the ideal POA irradiance in absence of topographic shading.

The *Day Fraction (DF)* is the proportion of daylight time that the sun is above the local horizon. For example, if for a given latitude May 25th would have 10 hours of sun without obstacles (e.g. over the open ocean), but the sun is only above the local horizon for 8 hours, then DF = 0.8. DF allows a first general and relative estimation of the horizon effects on irradiation without having to consider the orientation of a specific PV system. Fig. 4-3 shows the day fraction for the study area for January 1st. The areas with lowest DF (lowest DF = 0.7) are located in north-south valleys that block the sun during the morning and/or evenings.

However, DF overestimates the irradiation losses since for most solar systems the majority of the shading occurs in periods of low irradiation. Ultimately the annual irradiation loss for a specific system is the most relevant metric. After applying the topographic shading correction for 2009 (Fig. 4-4) we find maximum daily losses of 2.15 MJ m^{-2} day^{-1} and accumulated losses of 37 kWh m^{-2} $year^{-1}$ corresponding to 2% of yearly irradiation.

Database Description

Computing the horizons for the entire state of California was computationally impossible. About 60,000 processor hours were spent to compute the horizons in areas with the highest expected PV penetration and where topographic shading is expected to be an issue. The area computed covers Southern California (latitudes below 35 degrees including San Diego, Imperial, Orange, Riverside, Los Angeles, Ventura, Santa Barbara. San Bernardino, Kern and San Luis Obispo Counties) completely due to the mountainous area and the large amount of utility scale solar power plants being planned there. In addition, horizons for the San Francisco Bay area were computed. Fig. 4-5 shows the area covered by the database. The database details are described in a report on the californiasolarresearch.ca.gov website.

The horizon files are available on request from jkleissl@ucsd.edu. This horizon at 5 degrees increments can be used as an input to PV performance programs. It can be used to compute the losses due to

topographical shading in absence of any other information on the local obstructions (such as local fish-eye photographs).

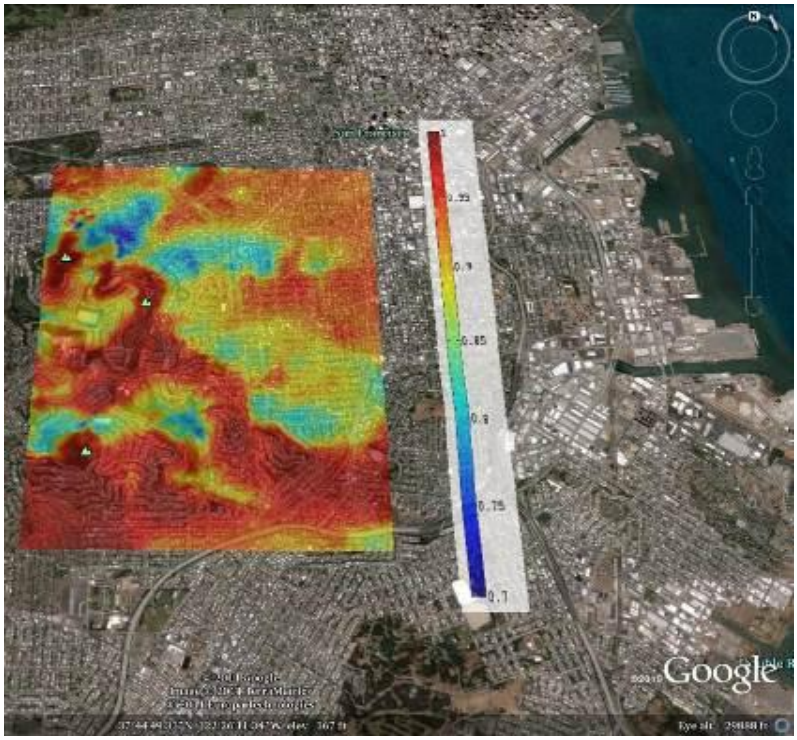


Fig. 4-3: Day fraction for the Twin Peaks area in San Francisco. If no horizon is considered, *DF* = 1. With the horizon day fractions vary between 0.7 and 1. The lowest values are observed near steep topographic features. North is to the top.

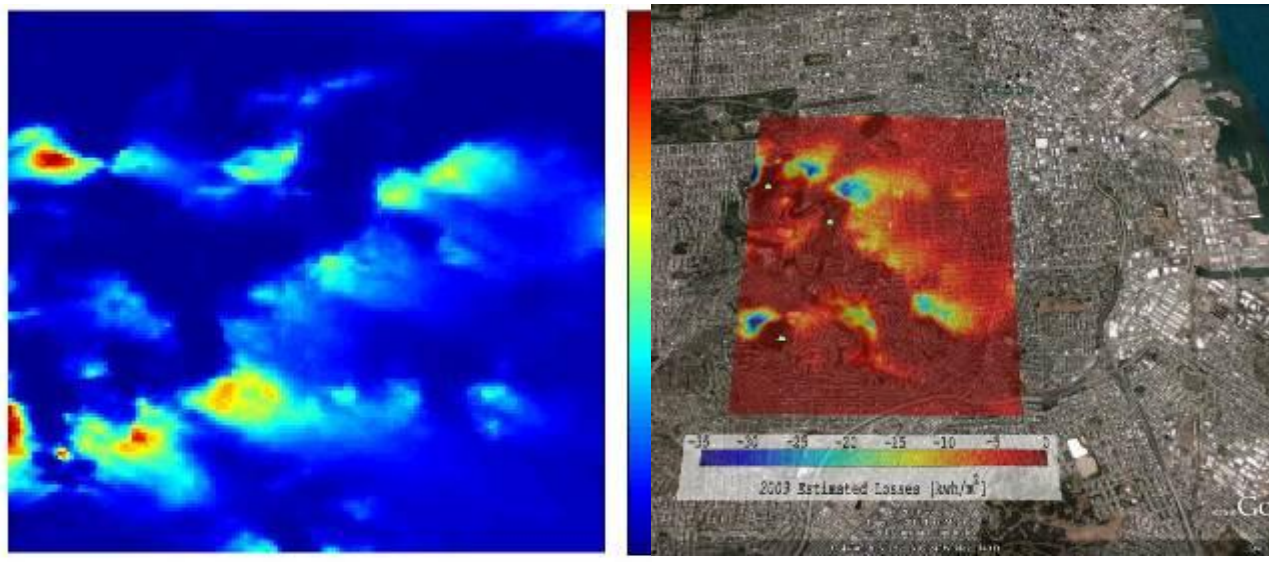


Fig. 4-4: Maximum daily losses (left, MJ m^{-2}) and accumulated yearly irradiation losses for the study area for a PV panel at south orientation and latitude tilt.

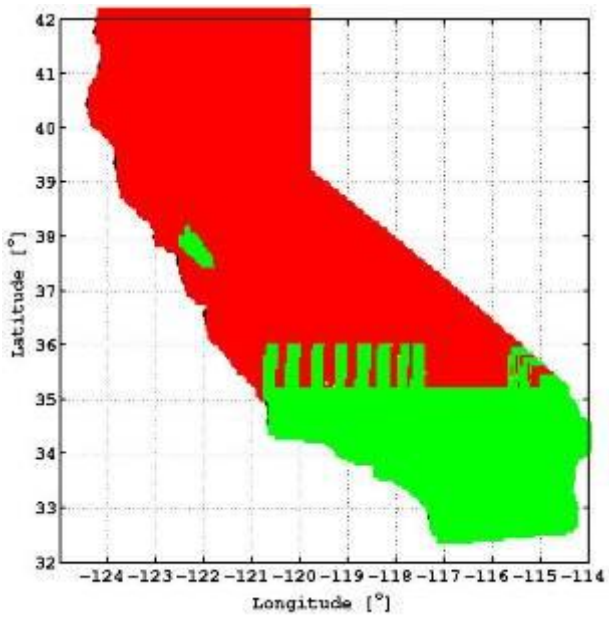


Fig. 4-5: Available horizon files for the state of California. Green indicates sites where horizon files are available.

5 Grid-integration Studies

PV Impacts on Distribution Feeders (study conducted by Enernex)

Small scale renewable generation is being incentivized in California and nationwide. Currently 1% of SDG&E customers have installed distributed solar PV systems. Small numbers of PV generation systems cause few or no problems on the grid, but as the percentage of PV generation grows, a number of issues begin to appear. The electrical impact that large penetrations of PV generators have on the system they are connected to with regards to power quality, power flow, protective relaying, intentional and unintentional islanding, etc. are concerns to utilities that should be assessed in engineering studies. Short-term irradiance fluctuations can cause voltage fluctuations that can trigger automated line equipment (e.g., tap changers) leading to larger maintenance costs for utilities and a reduction of the life-cycle of the stressed equipment. Some utilities report more frequent operation of voltage regulators on their high-PV penetration feeders leading to greater maintenance requirements. This chapter focuses on the integration and economic impacts of Distributed Renewable Generation on grid operation. Specifically, we investigated (1) how the distribution feeder loading changes with PV penetration level, (2) the impact of increased PV penetration on system losses, and (3) the effect of voltage fluctuations due to changing generation levels of PV on voltage regulator operation.

In typical distributed generation scenarios, a large number of PV generators are connected to single phases at many locations within distribution systems that have thousands of buses. Often the analyst is taking “shortcuts” by aggregating multiple generators. Similarly, the complexity of load variability is often ignored in simulations by aggregating thousands of loads and/or by unrealistically (but conveniently) assuming that each load in the system follows the same electrical demand profile. A consequence of the generation and load aggregations is that the flow of power within the distribution feeder is not accurately represented in the simulations, which leads to inaccurate predictions regarding tap changing operations of voltage regulators, line losses, etc.

A unique aspect is that we disaggregated the generation and the loads in the system as explained in the following paragraph. The system impact study was conducted on a distribution feeder that is based on a real-world distribution feeder located in SDG&E’s service territory. We investigated the effect of aggregation by comparing simulation results from simulations with aggregated load/generation with results from highly detailed simulations. For the highly detailed simulations, we explicitly represented the power generated at each PV installation by determining the irradiance at each PV location from experimentally determined shadow maps that have a temporal resolution of 30 seconds and a spatial resolutions ranging from 10 meters to 100 meters. This approach resulted in a realistic representation of the PV variability due to cloud movements. Nevertheless, a single large (2 MW) PV system near the end of the feeder disproportionally contributed to solar variability impacts and so modeling the generation of these large systems accurately is as important as disaggregating each PV generator. Similarly, the loads in our simulations were resolved down to the residential level. We have not been provided with measured load data and, consequently, we assumed that each load varies based on a normal distribution.

We executed a large number of simulations to study the potential issues related to different levels of PV penetration and the importance of high-resolution temporal and spatial variation of the PV and load data. We investigated the following four scenarios:

- Scenario 1, low (actual) penetration of small PV w/ 2 MW PV installation: The number and ratings of the PV generators are based on what is installed in the system today (based on data provided by SDG&E). The total capacity of the PV generators is 2.3 MW (about 20% of the total load MW rating, hereinafter referred to as 'penetration') with 0.3 MW coming from small PVs and 2 MW from two colocated large PV installations near a hospital.
- Scenario 2, low (actual) penetration of small PV w/o 2 MW PV installation: Same as scenario 1 with the exception that we removed the 2 MW PV installation to achieve a more homogenous distribution of the PV that is more representative of a typical distributed generation scenario. This scenario results in a total PV capacity of 0.3 MW (2.5% penetration) with only small PVs in the system.
- Scenario 3, high penetration of small PV w/ 2 MW PV installation: This scenario represents a hypothetical future scenario in which a large number of residential and commercial customers have PV generator operating on their premises. The small PV generators were duplicated nine times and the duplicated PVs were connected randomly to load buses on the feeder. No load bus has more than one PV. All original PV locations from Scenario 1 were retained. The large 2 MW PV installations were retained and not duplicated. This scenario results in a total PV capacity of 5.0 MW (41% penetration) with 3.0 MW coming from small PVs and 2 MW from two large PV installations near a hospital.
- Scenario 4, high penetration of small PV w/o 2 MW PV installation: Same as scenario 2 with the exception that we removed the 2 MW PV installations near the hospital to achieve a more homogenous distribution of the PV that is more representative of a typical distributed generation scenario. This scenario results in a total PV capacity of 3.0 MW (25% penetration) with only small PVs in the system.

We ran seven cases for each of the four scenarios described above (Table 4-1). The first six cases varied with regards to (1) aggregating/disaggregating the loads, (2) the temporal resolution of the input data (1 hour for the low-resolution cases and 30 seconds for the high-resolution cases), and (3) aggregating/disaggregating the PV generation during a cloudy day. The last case was a day with a clear sky with disaggregated loads and high temporal resolution (30 seconds). Aggregation/disaggregation does not make a difference for this clear-sky case as all PV generators are exposed to the same irradiance levels.

Table 5-1: Cases run for the feeder power flow simulations.

Case #	Load Aggregated	PV Aggregated	Resolution	Sky Condition
0	Yes	Yes	1 h	Cloudy to Overcast
1	Yes	Yes	30 sec	Cloudy to Overcast
2	No	Yes	30 sec	Cloudy to Overcast
3	Yes	No	30 sec	Cloudy to Overcast
4	No	No	30 sec	Cloudy to Overcast
5	No	No	1 h	Cloudy to Overcast
6	No	Yes	30 sec	Clear

Based on the simulation results, we draw the following conclusions:

1. Distributed PV vs. concentrated PV: The PV penetration levels in Scenario 1 and Scenario 4 are similar (20% and 25%, respectively). The main difference is that Scenario 1 has two large 1 MW PV system located close together which constitutes nearly 90% of the total PV in the system. On the other hand, Scenario 4 does not contain large PV installations and the numerous smaller PV installations were more-or-less evenly distributed along the feeder. For the concentrated PV Scenario 1, the tab changing operations were slightly more frequent and the line losses were larger compared the distributed PV Scenario 4. It is important to note that this tendency is not necessarily a general conclusion that applies to all feeders, but it shows that the distribution of the PV on the feeder has a significant impact on both the tap changing operations and the losses.

The maximum voltages on the feeder for the no PV scenarios and the PV scenarios without the large PV installations (Scenarios 2 and 4) are essentially identical (1.05 pu). On the other hand, the maximum voltages in the concentrated PV Scenarios 1 and 3 are significantly higher than in the no PV / distributed PV Scenarios (up to 1.08 pu). This indicates that the voltage regulators that exist on the investigated feeder are capable of keeping the voltages along the feeder within permissible limits – even if the number of distributed PV is increased tenfold. On the other hand, concentrated PV on the feeder is much more prone to cause overvoltages – in particular if voltage regulators are not strategically placed to accommodate the PV concentration on the feeder.

2. PV increases operation of tap changers significantly – in particular during cloudy days: Tap changers during cloudy days in the high PV penetration Scenario 3 (41% penetration) are operating about 70 to 80 times each day (depending on the simulation case) which is significantly higher than the number of tap changer operations for the no PV and low PV penetration Scenario 2 (about 13 operation for both scenarios) and the clear-day simulation case 6 (between 12 and 19 operations, depending on the scenario). This indicates that tap changing operations during cloudy days will significantly increase if PV penetration levels are high, which will substantially increase maintenance costs and lifecycle costs of voltage regulators. On the other hand, for low PV penetration levels and during clear days, the tap changing operations is not impacted much.

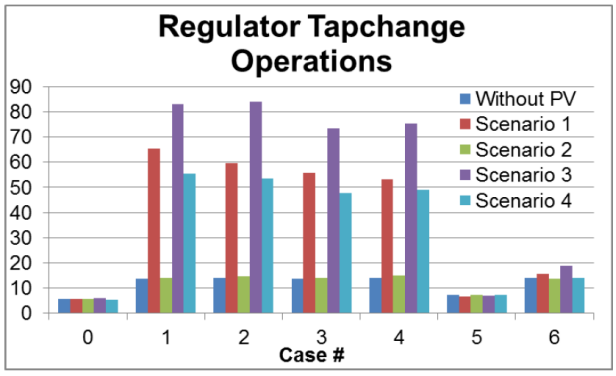


Fig. 5-1: Summary of total regulator events for all scenarios and cases.

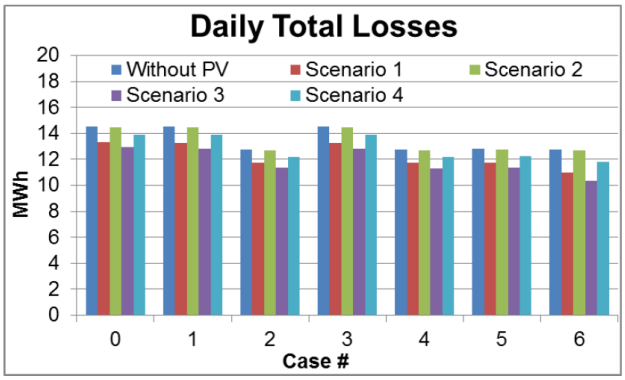


Fig. 5-2: Summary of MWh daily total losses for all scenarios and cases.

3. Low temporal resolution data vs. high temporal resolution data: Using low resolution data underestimates tap changing operations significantly in high PV penetration scenarios. In our simulations, the increase in tap changing operations for high PV penetration levels was only observed when more accurate high resolution data (30 second time step) were used as input to the model. Simulations with low-resolution data (one hour time step) did not show the increase in tap changing operations. This indicates that it is important to use high resolution data when evaluating tap changing operations for high PV penetration scenarios.

4. Effect of disaggregating PV irradiances: For the >20% PV penetration Scenarios 1, 3 and 4, a comparison of the number of tap changing operations for cases 1 and 3 (aggregated/disaggregated PV; loads aggregated in both cases) and cases 2 and 4 (aggregated/disaggregate PV; loads disaggregated in both cases) shows that disaggregation of the PV generation does reduce the number of tab changing operations significantly (30 to 70%). This indicates that aggregating the PV in simulations may exaggerate the actual tap changing operations that would occur on a real life system substantially.

5. PV reduces line losses: The presence of PVs significantly reduces line losses on the feeder. In the low PV penetration Scenario 1, line losses are reduced by about 8% during cloudy days and 14% during a clear day. In the high PV penetration Scenario 3, line losses are reduced by about 11% and 19% during clear days and cloudy days, respectively. The losses are even further reduced when confining the time frame of the comparison to the daylight hours (up to 50% for the high PV penetration Scenario 3 during a clear day). We attribute the reduction of line losses to the fact the PV generation is much closer to the load compared to the no PV scenarios in which the power is supplied from the substation. On the other hand, the line losses for the low PV penetration Scenario 2 are very similar to the line losses for the no PV scenario. Note that our simulations do not account for transmissions losses that occur during the transfer of power from the plant to the substation. Accounting for these losses would increase the difference between losses without PV on the system and losses with PV on the system even more.

6. Net consumption of active power: For the >20% PV penetration Scenarios 1, 3 and 4, the active power net consumption of the distribution feeder, that is, the active power that needs to be supplied from the substation to supply the load demand on the feeder, is significantly reduced. For example, for the high PV penetration Scenario 3, the daily reduction of net consumption due to the presence of PV leads to about 8-15% less energy demand for the feeder, depending on the sky condition. In other words, generation sources outside the feeder need to supply about 8-15% less energy to the feeder (not accounting for losses associated with transferring the power from the plant to the substation). If these

generation sources are fueled by conventional non-renewable resources (fossil and radioactive fuels), then this reduction leads to significant savings of these non-renewable resources.

7. Net consumption of reactive power: Conversely to the previous conclusion, the reactive power net consumption of the distribution feeder is increased, that is, more reactive power is consumed on the feeder, which has to be supplied by generation sources or reactive compensation outside the feeder or at the substation. This is because the PVs, in particular at high penetration levels, increase the voltages on the feeder, which results in more power consumed by the loads on the feeder. The PVs in our simulation operate at unity power factor, that is, they provide active power, but no reactive power. Consequently, the net consumption of reactive power is increased. On the other hand, the net consumption of active power is reduced because the additional active power (due to the voltage increase) is supplied by the PV. Note that this result depends strongly on the load mix on the feeder – we modeled all loads as ZIP loads with 70% active power and 30% reactive power. Our interpretation of this result is that PVs can decrease the operating efficiency of loads if the utility does not make adjustments. However, the increased voltages on the feeder due to the PV gives the utility the opportunity to apply energy saving measures such as conservation voltage reduction, that is, the utility can lower the voltage at the substation during times of PV production and the power generated by the PV will keep the voltages along the feeder within permissible limits. This practice would be particularly effective if (1) the PV is evenly distributed along the feeder as this will avoid localized regions on the feeder that have undervoltage when the voltage at the substation is reduced and (2) during clear days as this would avoid sudden undervoltage conditions due to clouds. Lowering the voltage at the substation will decrease the reactive power net consumption and further reduce the active power that needs to be supplied from outside generation sources.

Temporal Downscaling of Solar Irradiance Data via Hidden Markov Models on Wavelet Coefficients:

Considering the need for highly resolved data demonstrated in Section 4.1, a high resolution PV output dataset for over 115 PV sites spread across California was developed to facilitate other integration studies. 15 minute averaged data such as in the CSI database are not always sufficient. For example, to study power quality issues (‘flicker’) or impacts of PV sites on voltage regulation equipment, data at higher temporal resolution are required. At present, such data only exist at a few sites (e.g. less than five publicly available in California), but is mostly proprietary to power plant owners and system operators. Hence, one objective of the CSI RD&D program was to mimic realistic behavior of solar irradiance observations at a higher temporal resolution than the recording interval of the data. Given merely the actual recordings, this so-called ‘downscaling’ produces realizations at a higher sampling rate that are expected to be *statistically* similar to actual recordings. In this CSI project, observations provided as 15 minute averages are used to generate a sample with one observation per 0.88 seconds. Due to the tree-like structure of hidden Markov models, they are able to reflect certain interscale dependence structures.

Besides clear sky indexes, normalized power output for typical CSI systems is also generated. To estimate the effect of geographic smoothing over a PV site, we apply a wavelet-based variability model (WVM) as described in Lave and Kleissl [5]. Applying the WVM to the downscaled power data does not affect the variability on longer time scales as shown in Fig. 5-3 (left). However, a shorter snapshot (Fig. 5-3, right) reveals some smoothing due to spatially upscaling the point observations to the power plant. In this example, the PV systems are relatively small such that smoothing occurs at short timescales (a

few seconds), only. This effect is nonetheless critical to consider when computing e.g. short-time ramp rates of system power output as it reduces power quality impacts over short time scales.

Due to the large data volume, data are not posted online. Data is available upon request from http://solar.ucsd.edu/datasharing/. More details can be found in a report posted on casolarresearch.ca.gov.

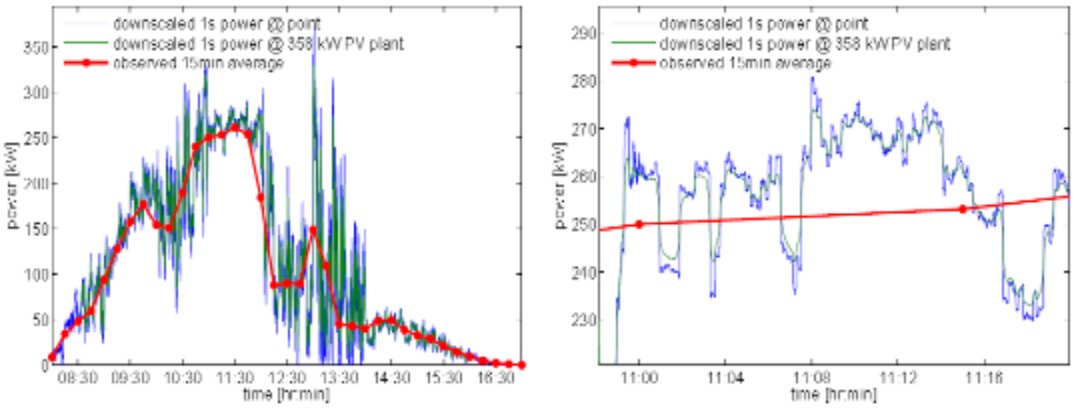


Fig. 5-3, *Left:* Downscaled, downscaled with wavelet-based variability model (WVM), and 15min observed power output. *Right:* Detailed display of 20 minute outtake of left-hand figure. Observations stem from the 358 kW CSI site at 38.0◦ lat, 122.6◦W lon for January 3, 2010.

6 Soiling Losses for Solar Photovoltaic Systems in California

Introduction

With the rapid increase in the use of photovoltaic (PV) power in California, which has 47% of the installed PV capacity in the US, the optimal management and analysis of expected performance of PV sites becomes increasingly important. Soiling can have a large effect on efficiency during long droughts [6], which mainly occur during the summer season coincident with the largest solar resource. Dust from air pollution particles, sea salt, pollen, agricultural activity, construction and other anthropogenic and natural sources accumulates on the panels until it is removed either by rain or washing.

Quantifying Losses due to Soiling

This Section contains a summary of the work presented in [7]. The AC power produced from 194 San Diego Gas and Electric (SDG&E), 385 Southern California Edison (SCE), and 403 Pacific Gas and Electric (PGE) sites were obtained for the year of 2010. These data were then quality controlled one-by-one to eliminate sites that had more than 70 % missing data, large noise, or inverter clipping of power. In this way, 305 sites with high quality data were identified. The CSI database also includes the azimuth and tilt angle of the solar panels.

The 15 minute data from the CSI database was aggregated over a day to obtain more robust efficiency estimates. The estimated solar irradiation from SolarAnywhere (SAW) was used to model the solar resource for each CSI site. Using the daily energy produced from the CSI site (P_{CSI}) and the daily incident solar energy modeled from SAW (P_{SAW}), the daily DC solar conversion efficiency (η) for the solar panels was calculated, controlling for the effects of temperature η_T and inverter η_{AC} efficiency. To be able to compare soiling effects between sites, η was then normalized by its average for the year to obtain a relative performance η_r.

Data from the California Irrigation Management Information Systems (CIMIS) were used to estimate the amount of rain at each CSI site. Two main factors control how much soiling exists on a PV panel: the accumulation of dust which is a function of location and duration of exposure, and the removal of dust through rain. PV panels are naturally cleaned by rain, but the effectiveness of cleaning varies with the amount of rain. Plotting increase in efficiency versus rain amount the team found that efficiency increased with rain amount from 0 to 0.1 in of rain and stabilized at larger rain amounts. This suggests a proportionality relationship for small rain amounts and a threshold of 0.1 in of rain beyond which the cleaning effectiveness does not increase. Consequently, a rain event was defined as a day when more than 0.1 in of rain are observed and is assumed to restore the panel's efficiency to that of a clean panel. Rain storms with multiple consecutive rain event days were combined into one multi-day rain event.

Large soiling impacts were observed at some sites and these soiling effects were particularly strong during the long summer droughts. At the site in Fig. 6-1a, there is a steady decrease in the efficiency of the PV plant after the last rainfall before summer (day 110). The rain events in the fall restore the PV plant to the efficiency observed at the beginning of the year.

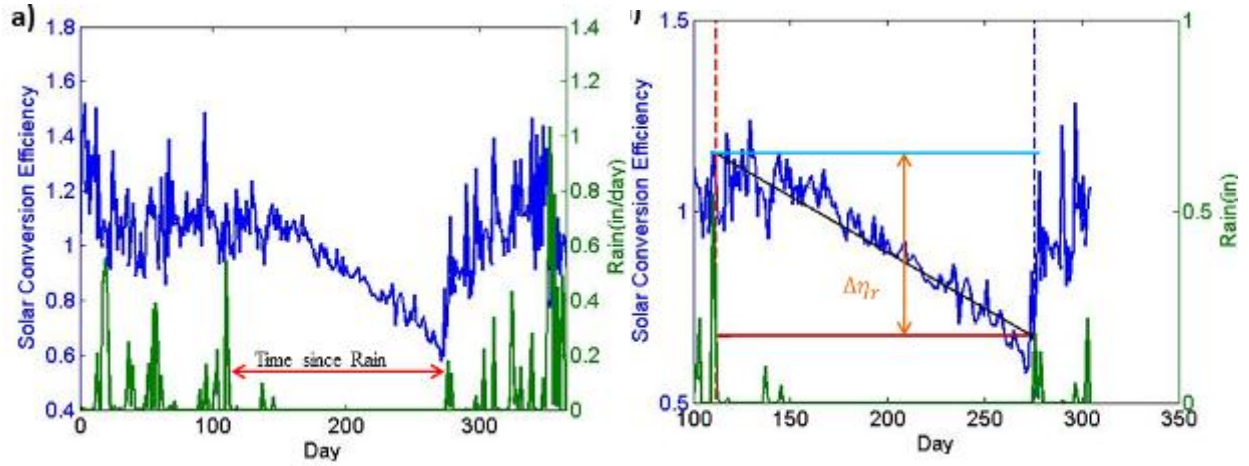


Fig. 6-1: (a) Timeseries of daily solar conversion efficiency $\boldsymbol{\eta_r}$ and daily rainfall for a 554 kW_{dc} PV plant in Hanford Kings, CA in 2010. (b) Soiling losses are quantified through linear regression of efficiency versus number of days since last rain in black. For consistency with the other methods $\boldsymbol{\Delta\eta_r}$ is expressed as the first (blue) minus the last (red) value of the drought period and these data are used in Fig. 6-2.

Soiling losses are calculated by applying a linear regression fit to the entire data during the drought period. The slope of the best fit line is then assumed to be the daily soiling for that drought. For quality control, droughts when more than 20% of the efficiency data were above 1.5 or below 0.5 were excluded. Also efficiency data greater than 1.5 or less than 0.5 were not used in the fit. This method can be observed in Fig. 6-1b, where the soiling losses for that drought period were found to be -0.0029/day.

Results and Discussion

Average soiling losses

Fig. 6-2 demonstrates the soiling losses as change in relative solar conversion efficiency versus time between rain events for all rain events at every site. The slope of the linear regression gives the average daily soiling losses as 0.00051 per day in relative solar conversion efficiency. In other words, if a site had an average efficiency of 15% its efficiency would decrease to 13.89% after a 145 day drought, which is the average of the longest drought period for each site.

To calculate the losses for each site, a linear regression is fit to the scatter plot of the soiling losses and drought period for each site. Some sites have a positive soiling losses (or soiling gains) which indicates that essentially no soiling occurred and that small errors in the solar resource model caused a positive slope in Fig. 6-1b. There is also a possibility that a few sites had automated washing systems (or meticulous owners/operators) which kept the panels continuously clean, but overall these scenarios are unlikely.

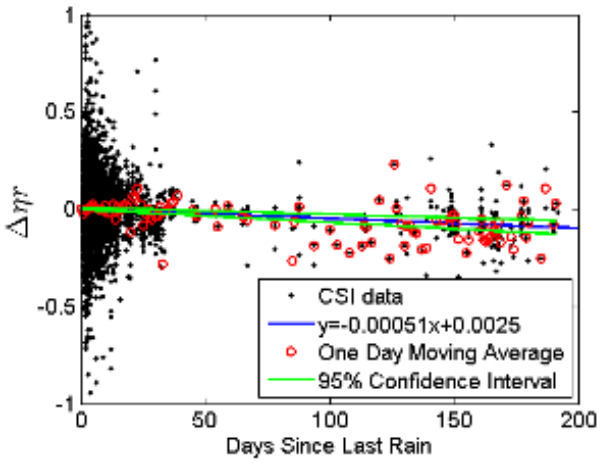


Fig. 6-2: Change in efficiency during a drought period (method 3) versus time since last rain event at all CSI sites. 12 outliers smaller than -1 are not shown. Red circles show the average for each day. A linear regression fit with 95% confidence interval is applied to the data.

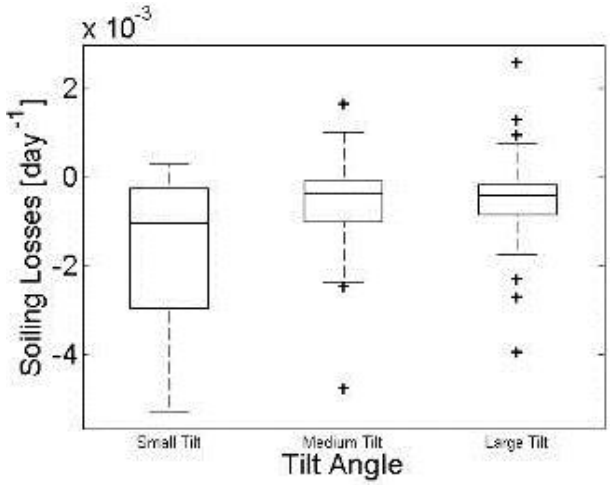


Fig. 6-3: Box-Whisker plot of the distribution of soiling loss (method 3) for different tilt angles. The bin from 0 to 5 ^{o} contains 12 sites, 102 sites have tilts from 6 ^{o} to 19^{o} and 88 sites have tilts equal to or greater than 20^{o}.

Effects of tilt angle and geographical location

48 CSI sites were identified to have losses greater than 0.001. To identify why these sites had larger losses the tilt angle and the location of the sites were investigated. Fig. 6-3 shows the mean soiling losses for tilt angles from 0-5, 6-19, and greater than 20 degrees. The average soiling losses for sites with a tilt angle smaller than 5^{o} is five times that of the rest of the sites as shown in Table 6-1.

A map (Fig. 6-4) and table (Table 6-1) of soiling losses by site was used to identify clustering of large soiling sites to identify patterns due to e.g. air pollution or farming. Large soiling appears to be more prevalent in the Los Angeles Basin and the Central Valley area, but the differences are not statistically significant at the 5% level.

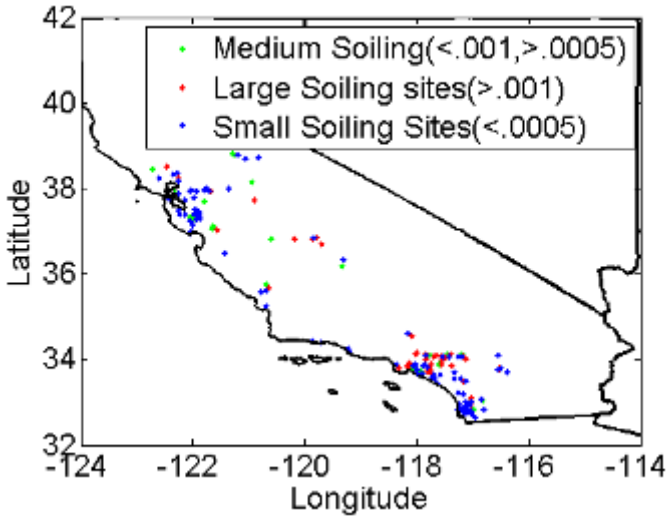


Fig. 6-4: Map of CSI sites and their soiling losses per day.

Table 6-1: Soiling losses stratified by geographical region.

	Number of Sites	Mean Soiling Losses [10^{-4} day^{-1}]	Fraction of sites with soiling > 0.1%/day [%]
San Francisco Bay	47	-4.8	12.8
Central Valley	29	-5.8	24.1
SCE	75	-8.3	41.3
SDG&E	35	-2.7	11.4

Conclusion

Power output data from 186 CSI sites demonstrated how soiling decreases the efficiency of solar PV plants. The accumulated soiling effects were found to be greatly affected by the time since the previous rainfall which is on average 145 days during the summer (Fig. 6-2). This supports previous findings that soiling can be modeled as a linear degradation [6]. On average losses were 0.00051 per day in relative solar conversion efficiency. Over an average summer 145 day drought this results in a 7.4% loss in efficiency which is more than an order of magnitude larger than losses due to cell degradation (typically 0.5% efficiency loss per year or 0.19% in 145 days). With a 15% efficient PV panel soiling losses over a 145 day drought would decrease the efficiency to 13.9%. Using a similar method and in a similar geographical region [6] had found much larger soiling loss of 0.002 per day. We hypothesize that the elimination of sites with R^2 values less than 0.7 in [6] as well as the limited amount of sites examined caused their soiling losses to be biased high. Sites with small soiling losses tend to have smaller R^2 since random errors in the solar resource estimates dominate over the correlation between soiling and time since last rain event.

The distribution of soiling by sites is skewed with a few sites showing very large soiling losses. Of the 186 sites, 48 were found to have soiling losses greater than 0.001 per day. One factor for these large soiling losses was the tilt angle. The data demonstrated that sites with a tilt angle less than 5 degrees had on average 5 times the soiling losses than the other sites. This finding supports that more soiling accumulates in a horizontal panel as previously found for glass plates in. The high variability of the data set suggests that soiling is very site specific and that some sites have reduced soiling at high tilt angles while others do not. For example some sites could have high winds that are able to clean low tilt panels while high tilt panels are better cleaned by gravity. Sites in the Los Angeles basin and the Central Valley Area were found to have larger soiling losses but the differences were not statistically significant to conclusively determine the location as a cause.

How much additional energy could be harvested through panel washing? Manual washing is expensive and typically only scheduled to occur during the summer drought. On average, the sites would have yielded 0.81% more annual energy if they had been washed halfway through the summer drought period (estimates are based on average soiling losses for each site and the length of the summer drought for each site). The distribution of additional energy production for one annual washing in Fig. 6-5 demonstrates solar energy production increases of up to 4%.

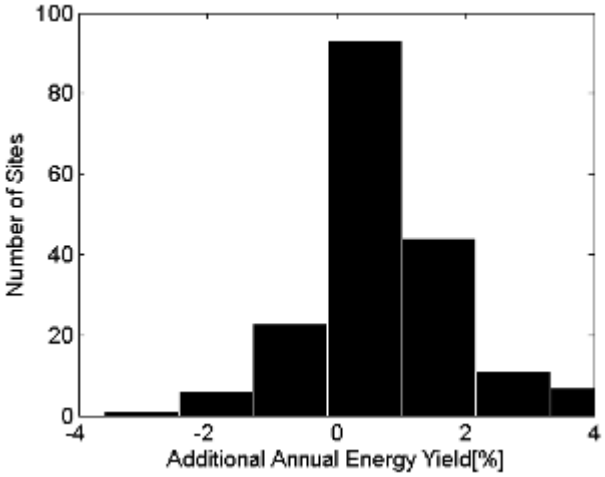


Fig. 6-5: Histogram of the additional annual yield possible for each site by cleaning the panel once per year halfway through the summer drought period.

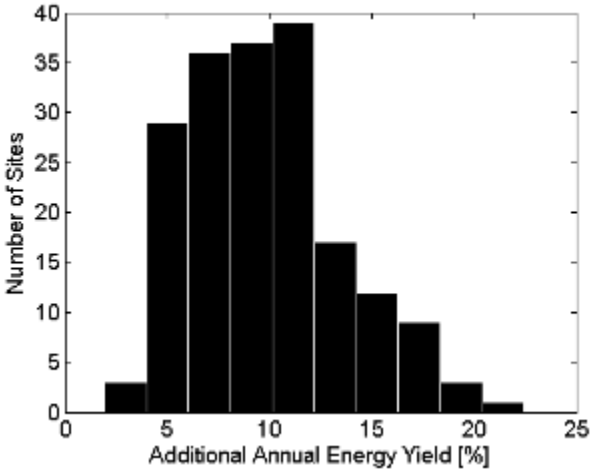


Fig. 6-6: Histogram of the percent additional yield if the panel was always clean.

If an automated cleaning system was installed such that the sites could be maintained clean larger energy gains are possible, on average 9.8% of annual energy. This estimate is calculated by assuming that the annual maximum of the 30 day moving average efficiency describes the maximum possible energy output for a completely clean panel. The extra yield is then calculated as the integral between the efficiencies of the ideal clean panel and the actual observed panel (Fig. 6-6).

The California Solar Initiative database is unique in that production data from a large set of stations is publicly available and soiling losses could be determined without confidential information. While soiling effects in California were found to be relatively small and rarely warrant the additional expense of panel cleaning, other sites with greater anthropogenic air pollution or natural events such as dust storms may experience more significant soiling.

7 Power Conversion Model for Distributed PV Systems in California Using SolarAnywhere Irradiation

Performance Model

The objective of this study is to design and test a PV performance model that can be applied to SAW irradiances to accurately estimate power output at PV systems in California. To design the performance model, using the GHI, DNI and tilt and azimuth angles of the PV panel, plane-of-array Global Irradiance (GI) is calculated using the Page model. To predict cell temperature (T_{cell}) ambient air temperature and wind speed are obtained from measurements at the closest CIMIS station. Then, T_{cell} is calculated using a 1D transient heat transfer model. The temperature efficiency correction is then calculated as

$$\eta_{Temp} = 1 - \alpha(T_{cell} - 25^{\circ}\mathrm{C}) \qquad \text{Eq. (6-1),}$$

where α is the temperature coefficient and calculated such that the highest linear correlation between modeled and measured AC power output are obtained (see Appendix A). The DC power output of the PV system is estimated as

$$P_{SAW,DC} = kW_{DC}.\eta_{Temp}\frac{GI}{1000\ \mathrm{Wm^{-2}}} \qquad \text{Eq. (6-2).}$$

Inverter efficiency is modeled as

$$\eta_{AC} = \frac{pf}{0.007 + 1.009\,pf + (0.0975\,pf)^2} \qquad \text{Eq. (6-3),}$$

where the power factor is $pf = P_{SAW,DC} / kW_{AC}$.

Next maximum power point (MPP) efficiency is considered as

$$\eta_{MPP} = a_1 + a_2 GI + a_3 \ln(GI) \qquad \text{Eq. (6-4).}$$

The MPP efficiency is applied to correct typically observed deviations in modeled output from measurements across different irradiation values using empirically obtained coefficients a_1, a_2, and a_3. In this study, these coefficients are calculated such that the highest nonlinear correlation between modeled and measured AC power output are obtained (see Appendix A). Then, the AC power output of the PV system is estimated as

$$P_{SAW,AC} = \eta_{AC}.\eta_{MPP}.P_{SAW,DC}C_f \qquad \text{Eq. (6-5),}$$

where C_f = $avg(P_{CSI})/avg(P_{SAW,AC})$ accounts for the line losses and soiling by calibrating the modeled performance by the ratio of the annual average CSI measured to modeled power output.

The calibration guarantees that modeled averaged annual performance based on SAW irradiation is consistent (or without bias) with observed performance. Such a 'modeled output statistics' (MOS) correction would typically also be applied in operational forecasting of PV power output.

According to Eqs. (6-5), the total solar irradiance to power conversion efficiency (η_{Tot}) is equal to

$$\eta_{Tot} = \eta_{Temp} . \eta_{AC} . \eta_{MPP} . C_f \qquad \text{Eq. (6-6).}$$

Validation

California-wide power output averages across the year

$P_{SAW,AC}$ and P_{CSI} averaged for all three IOUs together are computed for the year 2010 (Fig. 7-1). The results are essentially unbiased on average due to the calibration (C_f in Eq. 6-5) with small random errors. Typical differences (as measured by the rMAE and nRMSE) between the calibrated modeled and measured performance are 4 to 9% for 30-min averaged data. The average calibration factor of 0.91, C_f in Eq. 6-5, confirms that the performance model generally overestimates before calibration, likely since line losses and soiling of the PV systems that are not considered.

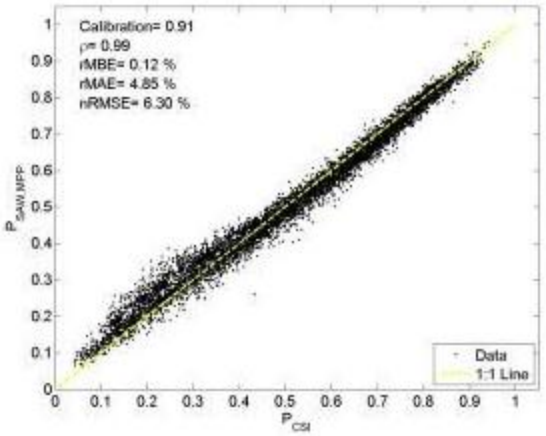


Fig. 7-1, Modeled SolarAnywhere versus measured power output: **30 min power output (for SZA<75°) in 2010 averaged over CAISO territory. The caption indicates the calibration factor (Eq. 6-5), the correlation coefficient (ρ), rMBE, rMAE, nRMSE between $P_{SAW,AC}$ and P_{CSI}.**

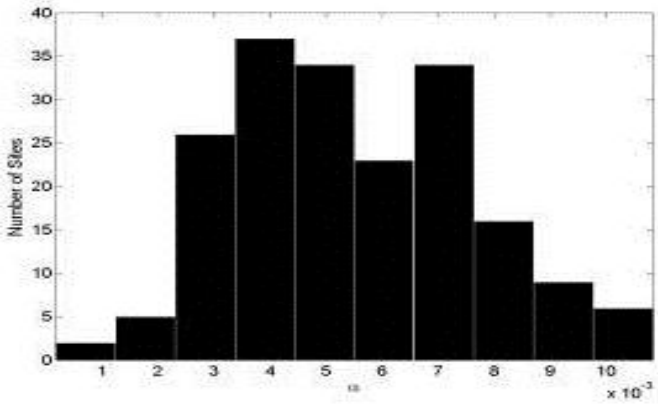


Fig. 7-2: Histogram of temperature coefficient (α) for all 192 PV sites.

The average temperature coefficient (used in Eq. 6-1) was found to be $\alpha = 5.5 \times 10^{-3}$ K^{-1}. Fig. 7-2 shows a histogram of the temperature coefficient for all PV systems. Temperature coefficients are expected to be between 3×10^{-3} K^{-1} for thin film and 5×10^{-3} K^{-1} for silicon solar cells. Since mounting details of individual PV systems was unknown the spread in temperature coefficients is also due to inaccuracies in the temperature model. Especially the spacing of the panel from the roof and roof properties can significantly affect convection heat losses and heat gain due to radiative interactions.

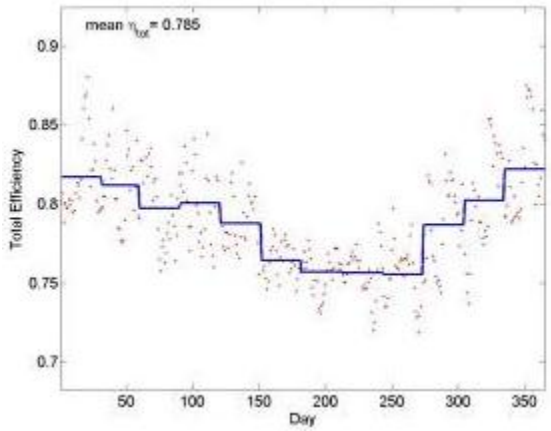


Fig. 7-3: Daily average relative solar conversion efficiency (SCE, η_{Tot}) averaged over 192 PV sites. The blue line represents monthly averages and the caption shows the annual average of η_{Tot}, averaged over 192 PV sites.

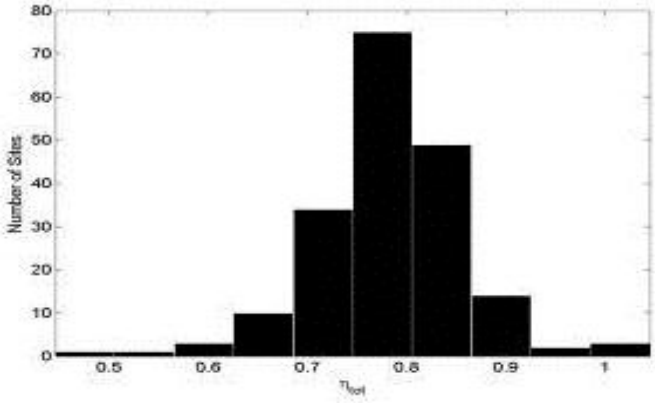


Fig. 7-4: Histogram of relative PV solar conversion efficiency (η_{Tot}) **for all 192 PV sites.**

The daily and monthly average PV solar conversion efficiency (η_{Tot}) averaged over all the PV sites is shown in Fig. 7-3. The annual average η_{Tot} is 78.5%. The histogram of annual average of η_{Tot} for all PV systems is presented in Fig. 7-4. Note that η_{Tot} is not the solar conversion efficiency that would be reported on a specification sheet of a solar module or even the plant solar conversion efficiency. Rather, since the panel area is unknown, only relative real-world efficiencies considering losses due to (likely in this order) temperature, line, soiling, and inverters can be considered. The trends in Fig. 7-3 are dominated by lower temperatures and consequently higher efficiencies during the winter.

Climatologies of bias errors by month and time-of-day

Averaged $rMBE_{MT}$ of all PV systems is shown in Fig. 7-5. Generally, the biases are less than 3% during midday. $P_{SAW,AC}$ underestimates P_{CSI} especially in March through May mornings, while $P_{SAW,AC}$ overestimates for small SZAs.

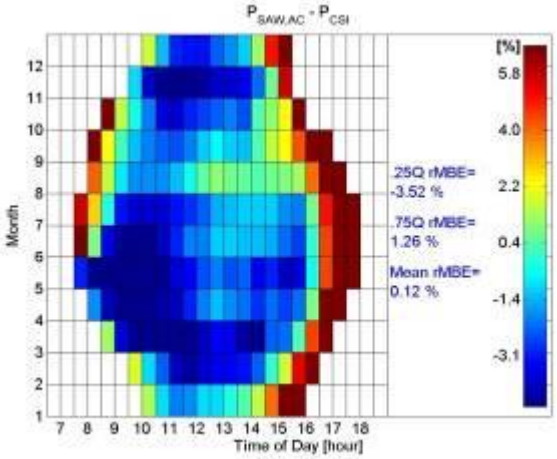


Fig. 7-5: rMBE by month and time-of-day for $P_{SAW,AC}$ versus P_{CSI} (averaged over 192 PV sites) in 2010. The caption indicates mean, 0.25 and 0.75 quantiles of rMBE.

Comparison to PVWATTS

PVWATTS is a standard calculator of PV performance maintained by NREL. For three representative sites, the PVWATTS PV performance was computed by Stephan Barsun (Itron) and compared to our performance model (Table 6-1).

Generally PVWATTS and our model provide similar results. Our model performs slightly better, but the bias was calibrated to be zero in our model which probably also explains some of the improvement in MAE and RMSE. While the performance is not significantly better, one of the big advantages of our MATLAB model is the 'true up' to a significant amount of real metered data as opposed to the defaults in PVWATTS and ability to use different time steps.

Table 6-1: Relative Mean Bias Error (rMBE) and relative mean absolute error (rMAE) for three different CSI PV sites in California. The following datasets are compared: PVWATTS (PVW), the performance model described in this chapter (us), and the measured CSI data.

Error Metric	**Site 1** PVW-us	PVW - CSI	Us - CSI	**Site 2** PVW-us	PVW - CSI	Us - CSI	**Site 3** PVW-us	PVW - CSI	Us - CSI
rMBE [%]	-1.6	-2.15	-0.6	-2.4	-2.9	-0.6	-4.4	-4.8	-0.5
rMAE [%]	3.7	9.8	9.2	1.3	7.9	7.1	4.5	9.5	8.2

As an aside, PowerClerk now tracks PV module mounting by standoff which will facilitate more accurate PV temperature modeling due to variation of convection losses on the back side of the module. In particular, standoffs are classified into, flush, 0 to 1, 1 to 3, and 3 to 6 inches.

Conclusions

A PV performance model was developed to calculate the expected power output for each PV system from SolarAnywhere (SAW) irradiation data (hereinafter referred to as SAW+P). Modeled power output was validated against measured power output from 192 PV systems in California (in SDG&E, SCE, and PGE territories) in 2010. An average PV efficiency derate (η_{Tot}) of 79% (loss of 21%) describes losses due to panel temperature, AC conversion, MPP tracking, and annual calibration. The average calibration factor of 0.91 confirms that the SAW+P model generally overestimates, likely due to line losses and soiling of the PV systems that are not considered in Eq. 6-5. The bias error between modeled and measured power output was found to be less than 3% during middays while SAW+P underestimates the measured data in winter and spring times.

8 Hours-Ahead Solar Forecasting using Satellite Imagery and Numerical Weather Prediction

Introduction

Hours-ahead solar forecasting is important for grid operators to manage intra-day solar variability through procurement of reserves and operation of energy markets. As Geostationary Operational Environmental Satellite (GOES-15) was deployed in 2011 and the more advanced GOES-R will be launched in 2017, the ability of solar forecasting using satellite imagery is expected to improve and to be widely used by forecast providers.

In this section a technique for solar forecasting using satellite imagery with the ability to characterize and predict forecast error is presented. Many forecast techniques have been developed for solar irradiance using satellite imagery and Numerical Weather Prediction (NWP) (e.g. Hammer et al, 2000, Perez et al., 2010, 2011, Marquez et al., 2012). Perez et al (2010) showed that the satellite cloud motion-based forecasts improve upon NWP forecasts up to forecast horizons of 4-5 hours on average over the United States. For operational forecasting, however, the relative skill of different forecasts depends on meteorological conditions. The motivation of this study is to determine this "cross-over time" under different conditions and determine parameters (cloud cover fraction, cloud speed, etc.) that are good predictors for this cross-over time.

Knowledge of the cross-over time and determining parameters will inform the composition of a forecast across all time horizons, which is based on optimal combinations of inputs from different models. For example, a satellite model would be chosen as the primary forecast if it had been determined to be most accurate for frontal passages, while a persistence model may be preferred for coastal marine layer clouds. More accurate and certain solar forecasts would facilitate more economical high solar penetration on the electric grid.

GOES-15 Satellite and Validation Data

The GOES-15 satellite dataset from April 24th to May24th, 2012 is used. The domain of interest covers an open ocean region with an approximate size of 2000 km x 2000 km from $32°22'$ to $34°31'$ N and from $122°62'$ to $127°62'$ W (Fig. 8-1). The open ocean area is selected to avoid complex terrain effects that introduce stationary (mountain) clouds and high spatial variability in satellite measured reflectivity. By selecting an ocean area, more generalizable results are obtained to illustrate the potential of the approach. Over land the approach is expected to apply as well especially in coastal areas that exhibit similar stratocumulus cloud types, but forecast errors may be larger due to the aforementioned terrain effects and more dynamic cloud fields. The next step should be to apply the developed techniques over different land areas. GOES-15 images starting at 1645 and 1700 UTC (09 PST) are used to determine cloud speed and the 1700 UTC cloud transmission image is advected with hourly time steps out to a forecast horizon of 6 hours. NOAA North American Model (NAM) forecast data are used for comparison.

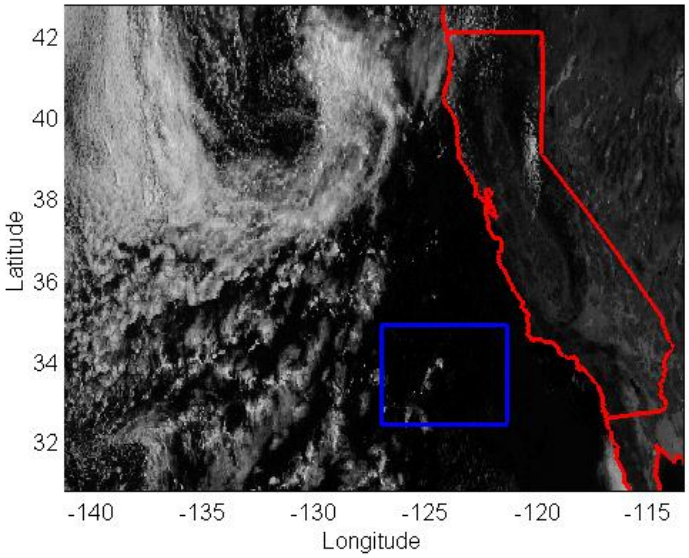


Fig. 8-1: GOES-15 visible channel image that shows the region of interest (blue) with the coast and California state lines (red).

Image Features and Selection

Seven different regional features are used to characterize the cloud conditions within the image (Fig. 8-1):

- average cloud speed,
- average kt (AVGKT) indicating the cloud fraction and cloud optical depth in the image,
- cloud fraction (Cf) indicating the cloudiness in the image,
- standard deviation of kt (std) indicating the variability of cloudiness in the image,
- correlation coefficient (CC) between the CMV advected and actual image after the first 15 min,
- entropy, a measure of the randomness of the kt field, and
- uniformity.

To establish the important determinants of forecast error, a multivariate linear regression analysis of forecast error versus different combinations of features is performed as in

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_i x_i + \epsilon, \qquad \text{Eq. (7-1),}$$

where y represents the RMSE of the forecast and and x_i are the features at a specific forecast hour respectively, β_i is the fit coefficient solved for in the regression analysis, and ϵ is the error between this linear model and the observations.

To address the limitations of the multivariable linear regression model, an artificial intelligence (AI) method called analogue method is also used to estimate the forecast error. The analogue of a forecast for a given location and time is defined as a past prediction that matches selected features of the current forecast. Analyzing historical forecasts errors for the analogues, forecast error can be inferred (Delle Monache et al. 2011). The search for analogue forecasts is controlled by a K-nearest

neighborhood (KNN) algorithm that ranks past predictions by how similar they are to the current forecast.

Results

Forecast performance of different methods

The performance of CMV forecast is compared to the NAM forecast and basic persistence forecast that equates future with current weather condition, i.e. the current kt satellite image is assumed to persist. Forecasts are validated against the true satellite image.

Fig. 8-2 shows the RMSE at different forecast horizons for each forecast model. The NAM forecast error varies only weakly with forecast horizon. The CMV forecast error increases with increasing forecast horizon as expected. The persistence forecast errors are slightly larger than CMV forecast errors up to a 6 hours forecast horizon. The cross-over time between satellite CMV and NAM forecast is found to be in between forecast horizons of 6 hours and 1 day, which is longer than in Perez et al. (2010) probably because Perez used ground measurements for validation.

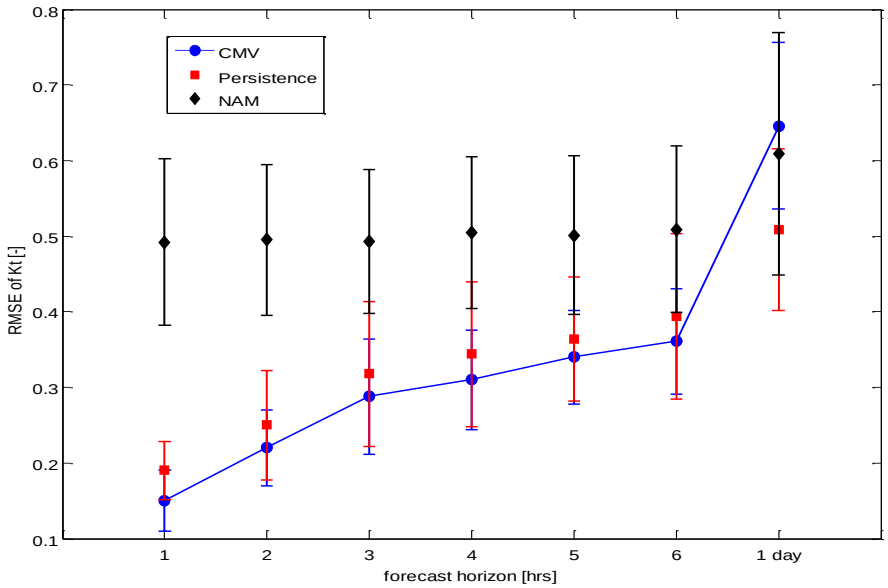


Fig. 8-2: RMSE of kt as a function of forecast horizon for each forecast model. Error bars show the standard deviation of the RMSE. Note the jump from 6 hours to 1 day on the *x* axis.

Multivariate linear regression for satellite forecast RMSE

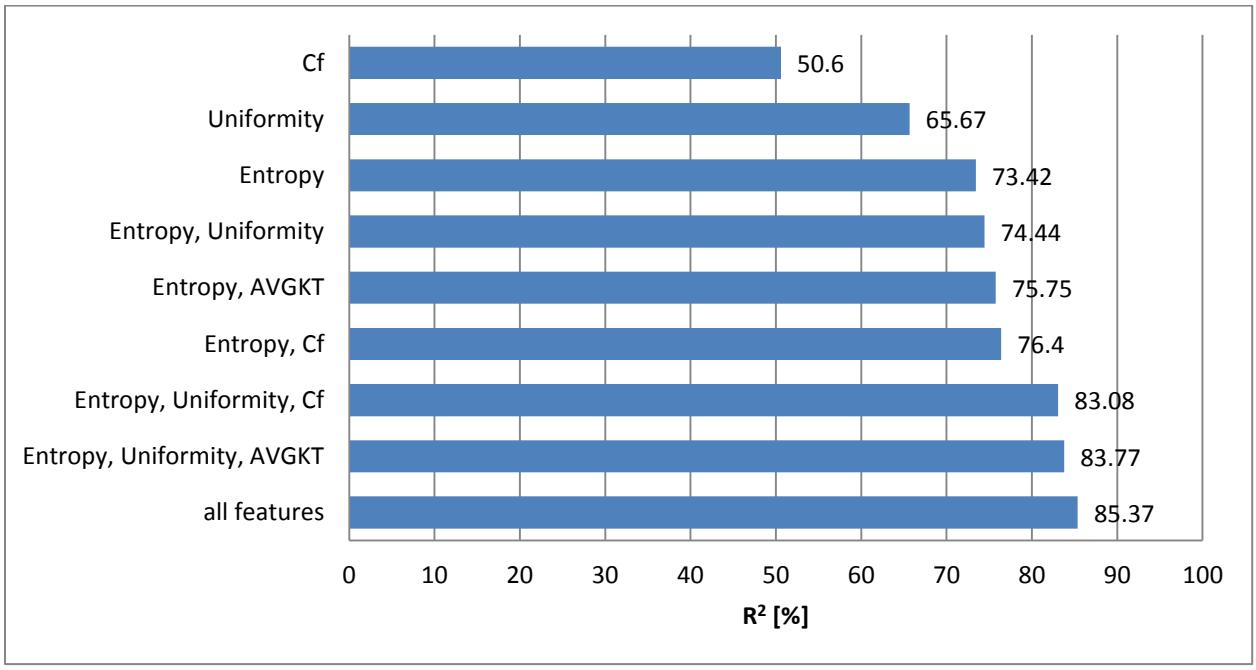


Fig. 8-3: Coefficient of determination (R^2) for different combination of features for the 1st hour of CMV forecast. The larger R^2 the better the predictive capability of the feature(s).

Fig. 8-3 shows the coefficient of determination (R^2) for several multivariate linear models (Eq. 7-1), each representing a different combination of features. R^2 describes the goodness of multivariate linear fit of RMSE. For the first forecast hour the satellite forecast RMSE is most correlated to Entropy (R^2=73.4%), followed by Uniformity (R^2=65.7%). If an additional feature is considered, Entropy and cloud fraction perform best with a 3% increase (R^2=76.4%) compared to the single feature using Entropy. Entropy and Uniformity are the most important features for 1-3 hour forecasts (not shown). At the fourth to sixth hour of forecast, entropy is still the dominant variable, but CC has a larger impact on the forecast RMSE for longer (> 3 hours) forecast horizon.

The multivariate regression analysis suggests that Entropy (i.e. the randomness of the kt field) is the best predictor of CMV forecast RMSE. A high randomness typically indicates that the cloud pattern is made of many small scale clouds or clouds with high spatial variability of optical thickness. Since smaller scale clouds tend to have a smaller 'half-life' than large clouds, the assumption of a 'frozen' cloud field in the CMV method is violated, resulting in larger forecast error.

Predicting Cloud Motion Vector (CMV) versus NAM forecast performance

The RMSE predicted by the linear regression method in Eq. 7-1 and the weighted analogue are validated against the true RMSE. To compare the predictability of relative errors of CMV and NAM forecasts, the relative performance (RP) is used:

$$RP_i = \frac{RMSE_{NAM,i}}{RMSE_{CMV,i}}, \quad \text{Eq. (7-2),}$$

where *i* indicates forecast horizon. RP < 1 indicates that CMV performs better than the NAM forecast, and vice versa. Satellite image Uniformity is observed to be the most important feature for predicting the RP. The impact of other features varies by forecast horizon.

Table 8-1 shows a comparison of the performance of the multivariate regression and analogue method in predicting CMV forecast error and RP. The analogue method is shown to outperform the multivariate linear regression. The average rRMSE of predicting CMV forecast error is 14.4% and 23.9% for analog and regression method respectively, while the average rRMSE of predicting RP_{NAM} is 27.6% and 31.7%. The errors of both methods are found to be independent of forecast horizon.

Table 8-1: Error in predicting CMV forecast error and RP with analogue and multivariate linear regression model. Training data is from April 24 - May 24 and test data is May 25 – June 25, 2012.

	Predicted CMV forecast error		Predicted RP_{NAM}	
Forecast hour	$rRMSE_{analog}$	$rRMSE_{regression}$	$rRMSE_{analog}$	$rRMSE_{regression}$
1	16.5%	21.4%	31.2%	37.3%
2	13.0%	28.6%	27.2%	30.6%
3	13.5%	19.7%	20.3%	27.2%
4	15.6%	23.5%	29.6%	31.0%
5	14.0%	23.2%	30.3%	35.1%
6	13.9%	27.1%	27.1%	29.0%

Conclusions

A technique for solar forecasting using satellite imagery with the ability to characterize and predict the forecast error using multivariate linear regression and the analogue model is presented. CMV solar forecasts are computed using the cross-correlation algorithm applied pixel-by-pixel to satellite cloud transmissivity. On average, the RMSE is found to increase with increasing forecast horizon for both persistence and CMV forecast and becoming large than numerical weather prediction model (NAM) forecast error (i.e. with a cross-over time) at forecast horizons between 6 hours and 1 day.

The dependence of forecast RMSE on different features in the initial satellite image is analyzed. Entropy is shown to be the most influential determinant to the CMV forecast RMSE, since large entropy is associated with broken cloud fields where there is greater likelihood of cloud deformation near cloud field boundaries and greater difficulty of a CMV forecast to match the future cloud field.

Then the ability of two statistical models to characterize forecast error is analyzed to provide an indication on whether confidence in satellite versus NWP forecasts could be characterized ab-initio. The analogue model was more successful than a linear regression model in predicting forecast error. This indicates that the CMV model errors show a non-linear dependence on the selected features that can be better captured by the analogue method versus a linear regression.

The relative performance of NAM and satellite CMV models is also analyzed. Based on Fig. 8-2, RP decreases from about 3.5 at the first forecast hour to 1.5 at the 6th forecast hour, on average. The cross-

over time (when CMV becomes less accurate than NAM forecasts) is found to be more than 6 hours, on average, but as early as 2 hours under conditions of cloud fields with large entropy or strongly rotational flow. The rRMSE of predicting RP is shown to be 27.6% and 31.7% for analog and regression method respectively. Consequently, both models have similar skill in informing forecasters at what forecast horizon a numerical weather prediction forecast should start to be used instead of a satellite forecast.

With only two months of available satellite data, the model performance is limited by the data size being used to train the models. Especially for the analogue method a larger training data set increases the chance to find similar analogues improving model accuracy.

Table 8-2 shows the pipeline of operational satellite CMV forecast and the processing time for each step using a desktop computer. Using this strategy an operational forecast can be provided within less than 25 minutes after satellite image capture. This limits CMV forecast value for short (2 hour) cross-over times as the remaining high accuracy portion of the forecast is short, especially after other time lags in the electric power market systems are considered. However, for the large majority of times when crossover times are several hours, the satellite forecast is valuable for improving the accuracy of solar forecasting and reducing solar integration costs.

Table 8-2: Pipeline of CMV forecast and the processing time for each step.

Procedures	Processing time
Download and extract satellite data	20 min
Compute cloud motion vectors	40 sec
Advect kt pixels and register forecast images out to 6 hours	20 sec
Predict errors and RP of CMV and NAM	90 sec
Total	23.5 min

9 Intra-hour Forecasting using Cloud Speed Sensors

Introduction and Data

The key barrier against high PV penetration scenario is power output variability. Clouds cause spatio-temporal variability of solar radiation that is the main cause of fluctuating photovoltaic power feed-in to the grid. Cloud velocity is a principal input to most short-term forecast and variability models. For example, a simple estimate for the maximum possible ramp rate of a solar power plant is $dP/dt = P_{csk}\, k_t\, v / l$, where P_{csk} is the clear sky power output, kt is the cloud clear sky index (related to its optical depth), v is the cloud velocity, and l is the length of the power plant in the direction of cloud motion. In words, the ramp magnitude depends on the kt; in addition the ramp rate is a function of how long it takes for a cloud to cover the plant (assuming that the cloud is larger than the plant). Short term forecasts often rely on advection of a frozen cloud field using cloud motion vectors (CMVs). CMVs are usually derived from satellites, but not only are satellite data complex to acquire and process, but infrequent data update (every 15 to 30 min) and data transfer delays also may not allow detection of mesoscale convective clouds in a timely manner. Consequently, local ground measurements of cloud speed are advantageous for short-term solar variability and solar forecasting.

Short-term solar forecast using sky imagers shows great promise, but the deployment and operation of sky imagers is capital and labor intensive [10]. The main goal in this study is to design a sensor array and automated algorithm to estimate CMV using a simpler approach. Time delays in cloud arrival times estimated using ground-based pyranometers can be used to estimate cloud speed [8]. For this purpose Global Horizontal Irradiance (GHI) was measured using 8 photodiode pyranometers (Li-200SZ, Licor, Inc.) and logged to a single CR1000 (Campbell Scientific Inc.) data logger with an acquisition frequency of 20 Hz. The sensors were deployed in a semicircular shape with a radius r of 6 m, at the UC San Diego Solar Energy Test Bed (UCSD-SETB, Fig. 9-1). The sensors were placed onto an exposed rooftop without obstacles or shading and were cross-calibrated on a clear day using linear regression fits.

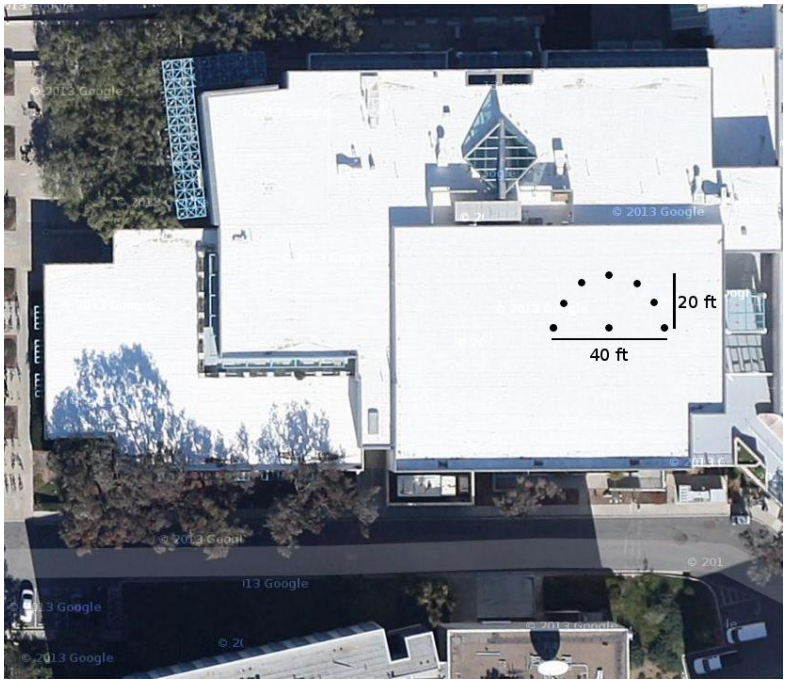


Figure 9-1**:** Sketch of the layout of the cloud speed sensor. Sensors were spread out over an area of 20 x 40 ft.

Cloud Speed Algorithms

Two different methods to determine CMVs were developed based on the correlation between the GHI timeseries. The basic premise is that the GHI timeseries for a pair of sensors aligned with the CMV will be highly correlated, but the correlation is largest if the upwind sensor timeseries is lagged by the travel time of the cloud between the two sensors. Once the lag has been determined, cloud speed can be calculated from the sensor spacing. Since cloud direction is variable several pairs of sensors covering many directions are needed. The direction of cloud motion is then given by the alignment of the most correlated pair. An alternative method to calculate CMV is proposed using a reduced set of three sensors. However, in this case the angle of the cloud shadow edge relative to the array also has to be determined.

Most Correlated Pair Method (MCP): The basic premise is that for a pair of sensors Sa and Sb aligned with the cloud motion, separated by a distance D, the irradiances Ia and Ib are highly correlated, but with a time lag tab. Once the lag has been determined (e.g. by analyzing the signal cross correlation), cloud speed can be calculated as
$v = D / t_{ab}$

For the experiment, 8 sensors are grouped in 7 pairs, where all pairs share the central sensor. The cross correlation was performed for all 7 pairs with a time step of 5s, and using a time window of 30s (600 data points). The pair with the largest cross correlation coefficient is assumed to be the most aligned with the CMV. Cloud speed is then calculated using the most correlated pair.

Linear Cloud Edge Method (LCE): For reference data from an alternative method using a triplet of sensors is shown; CMVs are calculated assuming a linear cloud shadow edge (LCE) passing through the array. We assume (i) a linear cloud edge shape across the sensor array; (ii) constant CMV while passing over the three sensors; (iii) the cloud is large enough to cover all three sensors. Quality control is applied to both the raw MCP and LCE results.

Cloud Speed Results on 3 Days

The methods were tested the three remaining days with partial cloud cover, October 20th, 21st, and 25th, 2011. Review of total sky images indicated that cumulus was the main type of cloud observed for October 20th and stratocumulus for October 21st, while overcast skies were observed for most of October 25th. METAR reported low cloud heights for October 20th and 21st, in the range [274-640]~m, and higher clouds in [1200-1800]~m for October 25th. For October 20th clouds were reported to be broken from 8 to 11PST, scattered/few from 11 to 17 PST and then overcast for the rest of the day. For October 21st clouds were reported to be few for most of the day with periods of clear sky and scattered clouds. For October 25th the sky was reported as overcast for most of the day.

To illustrate the performance of the MCP method, Fig. 9-2 shows 20~min of data on October 21st. During the first 10 min in clear skies MCP outputs a random direction alpha in the range [-180,180]. In this interval both the maximum cross correlation coefficient and the standard deviation of GHI are small.

After 12.7 h, alpha becomes consistent at -60^{o} (i.e., 300^{o}). Solar irradiance in this second interval shows larger variability and maximum cross correlation coefficient due to cumulus clouds.

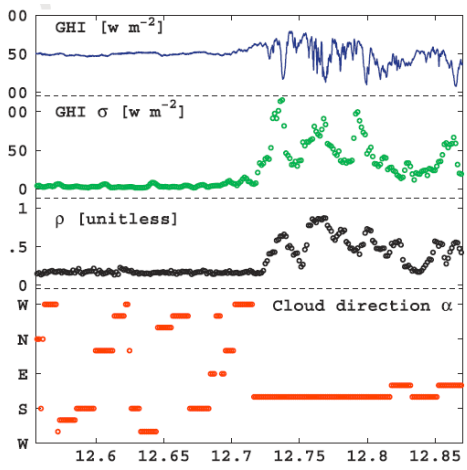


Fig. 9-2: GHI and its standard deviation every 5s over 30s intervals and MCP results for maximum correlation coefficient ρ and cloud direction α on October 21, 2011.

Fig. 9-3 shows the results obtained for the studied days. For October 20th, overcast conditions prevail in the morning and both methods fail to extract usable CMVs. In the interval 1300-1600 PST, both methods result in similar directions, also comparable to the 1623 PST radiosonde data. The cloud direction was different for the three studied days; both methods reflect the intra and inter-day variability and agree with the radiosonde measurement. The only daytime radiosonde measurement occurs at 1623 PST, which does not overlap but is only 30 min after the last results for cloud direction were obtained from the MCP and LCE methods. The cloud direction results also agree with cloud directions determined visually from total sky imagery.

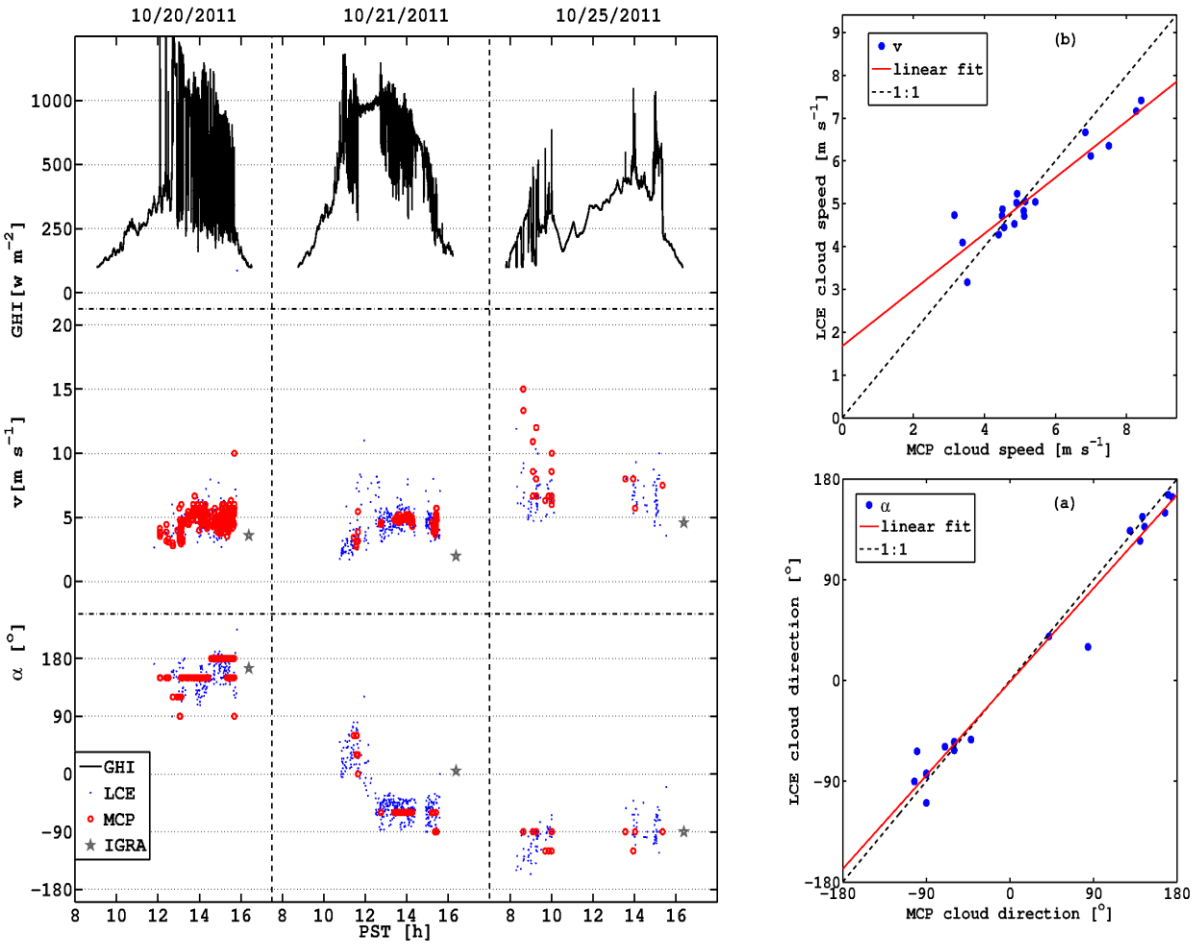


Fig. 9-3, Left: GHI and cloud motion vector comparison for October 20th, 21st and 25th, 2011. Radiosonde (IGRA) results obtained at 1623 PST are shown as stars. Right: Validation of Linear Cloud Edge (LCE) versus Maximum Correlated Pair (MCP) for half-hourly cloud speed (*v*) and direction (a).

While cloud direction serves as a point of comparison of the different methods, the primary output relevant to solar power variability is cloud speed. The results confirm those obtained for the cloud direction, with a similar output from both MCP and LCE methods of about 5 ms^{-1} on October 20th and 21st and 5 to 10 ms^{-1} on the 25th, in agreement with radiosonde observations.

MCP and LCE results are not temporarily aligned. To objectively analyze the consistency of both methods, 30~min vector averages of speed and direction were compared on the three study days. Fig. 9-3 a and b show the results for direction and speed, respectively. Linear regression between the LCE alpha and MCP directions show a regression factor R^2 of 0.977 and a root mean squared error of RMSE=16.5^{o}. The same regression performed for the speeds resulted in a regression factor R^2 of 0.875 and RMSE=0.40ms^{-1}.

Overall we conclude that encouraging cloud motion vector results were obtained from the LCE and MCP methodologies. Both methods require high variability in the radiation from clear-cloudy or cloudy-clear transitions. Cloud speed of an overcast cloud layer or the atmospheric velocity in clear skies cannot be obtained, but there is also little practical relevance at these times. The Most Correlated Pair Method (MCP) is the more robust approach and should be considered the best of the two presented options.

Based on the successful demonstration, a provisional patent has been filed and a cloud speed sensor was designed that could provide self-contained cloud motion vector measurements at utility-scale solar power plants.

Design of a Cloud Speed Sensor

Based on the encouraging preliminary results, the research team examined the design of a cloud speed sensor. The quest for compactness, faster acquisition rates, durability, and on-board processing are the primary motivations behind the CSS design. Thirteen illuminance sensors are used to determine cloud motion. There are twelve outer sensors covering 0-110° in 10° direction steps arranged in a circle formation with radius of 0.152 m relative to a central sensor. Without incorporating more sensors, the remaining angles to complete a full semicircle can be obtained using equilateral triangles constructed from the sensors in the circle (Fig. 9-4).

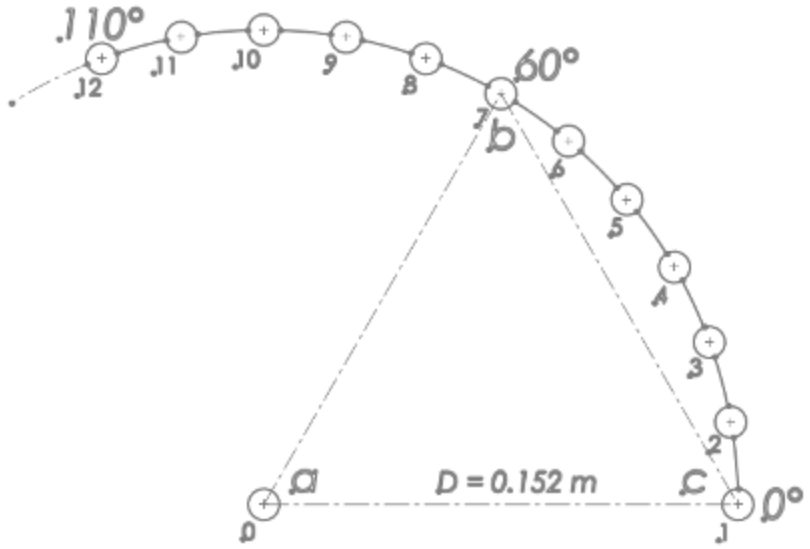


Fig. 9-4: Sensor arrangement. Each circle represents a sensor arranged in 10° spacings. Angles from 120 to 170 degrees are obtained through equilateral triangles. For example, for triangle *abc* the line from b to c results in an angle of 120°.

The cross correlation coefficient is calculated for sensor pairs in 10° increments and used to determine cloud direction and speed.

The CSS uses Vishay TEPT 4400 silicon NPN epitaxial planar phototransistors to observe light intensity. Similar to the human eye, the phototransistor is sensitive to visible light, (range of spectral bandwidth: 440 to 800 nm) with peak sensitivity at 570 nm. TEPT 4400s are designed to operate in harsh environments such as high temperatures without performance decrease. A white opaque acrylic mounting plate that supports the 13 sensors helps to reduce the internal temperature of CSS and shields sensitive components such as the voltage regulators and microcontroller.

10 Aggregate Ramp Rates Analysis of Distributed PV Systems in CAISO Territory

Introduction

Integration of large amounts of photovoltaic (PV) into the electricity grid poses technical challenges due to the variable solar resource. Solar distributed generation (DG) is often behind the meter and consequently invisible to grid operators. The ability to understand actual variability of solar DG will allow grid operators to better accommodate the variable electricity generation for resource adequacy considerations that inform scheduling and dispatching of power. From a system operator standpoint, it is especially important to understand when aggregate power output is subject to large ramp rates. If in a future with high PV penetration all PV power systems were to strongly increase or decrease power production simultaneously, it may lead to additional cost or challenges for the system operator to ensure that sufficient flexibility and reserves are available for reliable operations.

In this study, aggregate ramp rates of distributed PV systems installed in CAISO territory are analyzed [9] (Table 10-1). The objective of the ramp rate analysis is two-fold. Firstly, knowledge about the largest possible aggregate ramp rates and underlying meteorological conditions is useful for system operators to plan for worst-cases. Secondly, under extreme ramp rate conditions, knowing the PV output in real time would be most valuable since regulation up or regulation down capacity may have to be quickly procured. The ability of CIMIS and SAW modeled PV performance to match actual output is therefore of interest.

Table 10-1: Statistics of PV systems in SDG&E, SCE, and PGE territories

IOU	No. of PV systems	Total PTC rated capacity(MW)	Mean PTC rated (kW)	Median PTC rated (kW)
SDG&E	45	4.73	105.1	46.43
SCE	81	17.48	215.8	192.9
PGE	66	16.29	229.4	165.8

We will show that the largest absolute ramp rates are solely a result of diurnal cycles. To verify that, 1-min GHI in clear sky conditions (GHICS) at each PV site is calculated based on the Ineichen model with Linke Turbidity from the SoDa database. Then, GHICS is averaged over the CSI time interval (15-min). The aggregate GHICS is calculated at each time step and differences in the aggregate GHICS are also calculated for different ramp duration intervals; 15-min through 5-hour in 15-min increments.

Absolute Ramps

We present normalized absolute ramp rates to facilitate scaling the results to future PV penetration scenarios (assuming a similar geographic diversity). Therefore, the aggregate power outputs are normalized by the aggregate (PTC) kWAC capacity of the PV systems for each time period. Also to facilitate the comparison between CIMIS and SAW irradiances, and CSI power output, each timeseries is normalized by the respective maximum daily values (normalized to a maximum of 1 for each day in Fig. 10-5).

Absolute ramp rates in aggregate output and clear sky irradiance by time horizon

The largest step sizes in the absolute aggregate PV power output (normalized by kW_{AC}) and the aggregate GHI_{CS} (normalized by 1000 W m^{-2}) are detected over the year for different intervals (Fig. 10-1). As expected, the maximum ramp magnitude increases with the ramp interval. For the absolute ramp rates, the increase is near linear up to about 120 minutes at about 0.46% of PTC per minute. The maximum ramp magnitude approach 90% for 5 hour ramps reflective of the diurnal cycle (e.g. from zero output at 0700 to near maximum output at 1200) on a clear day.

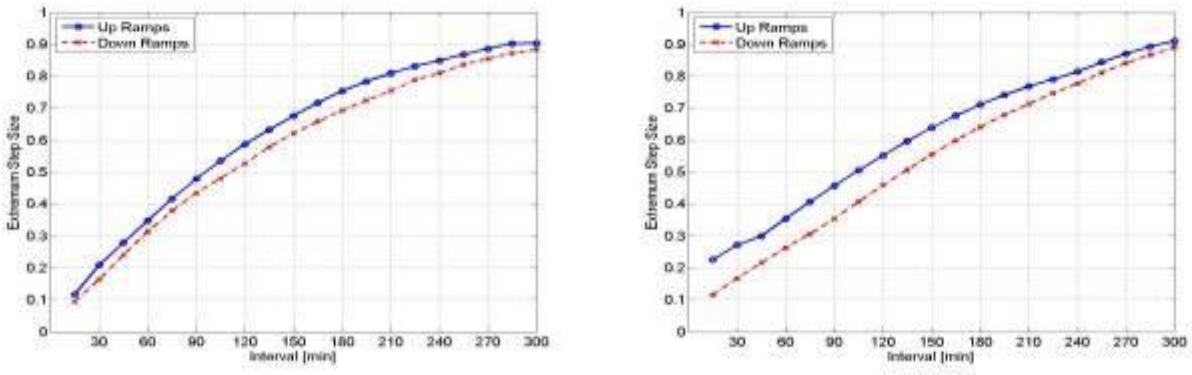
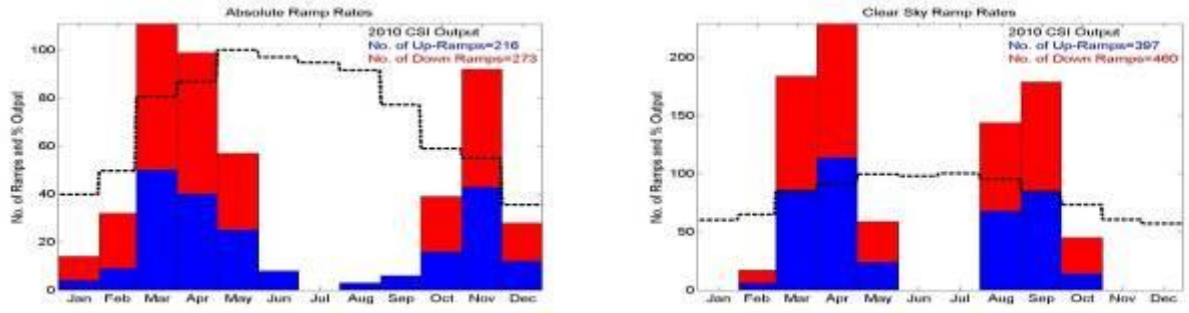


Fig. 10-1, Largest absolute ramps: Largest ramp magnitude versus ramp time interval (from 15-min upto 5-hours) for aggregate (a) normalized output (P_{CSI}/kW_{AC}) and (b) clear sky GHI (GHI_{CS}/1000 W m^{-2}) from 192 PV sites in CAISO territory.

1-hour ramps have a special significance as most energy exchange between electric balancing areas is currently scheduled over hourly intervals. The clear sky rate is of interest because it simulates the output ramps that would be experienced if there were no clouds or fog and the weather was always clear. This is the precisely predictable rate that is not governed by weather.

Fig. 10-2 shows a histogram (by month) of the 1-hour absolute ramp rates of aggregate normalized power output (normalized by kW_{AC}) and 1-hour ramp rates in the aggregate clear sky GHI, which are larger than 23% of PV capacity and 22% of maximum GHI_{CS} respectively. The largest absolute ramps in March and April can be explained by clear sky ramps.



(a) (b)

Fig. 10-2, Month of occurrence and direction of large absolute ramps: Histogram of the largest 1-hour ramp rates of aggregate (a) absolute 15-min output (P_{CSI}/kW_{AC}) and (b) clear sky GHI (GHI_{CS}/1000 W m^{-2}) from all 192 PV sites. The black lines show percentage of the measured aggregate 15-minute PV output (in Fig. a) and the aggregate 15-min clear sky GHI (in Fig. b) from all 192 PV sites normalized to the maximum month (May).

Weather-induced Ramps

Overall statistics

Since the largest absolute ramp rates are solely a result of diurnal cycles, and are therefore predictable, weather-induced ramp rates with reference to a 30-day average of power output are helpful to detect unexpected variations. These unexpected variations are more likely caused by weather than the sun's movement through the sky. First the average aggregate PV power output of the previous 30 days at a given time of day (ToD), corrected for differences in aggregate PV capacity, is subtracted from the aggregate PV power output at that ToD. Then, the differences in the resulting timeseries constitute weather-induced ramp rates, which are calculated for different ramp duration intervals; 15-min through 5-hour in 15-min increments.

Similar to Fig. 10-1, the largest step sizes in the weather-induced aggregate ramp rates (normalized by kW_{AC}) are presented in Fig. 10-3. The maximum ramp magnitude increases with the ramp interval and approaches 50% for 5 hour ramps.

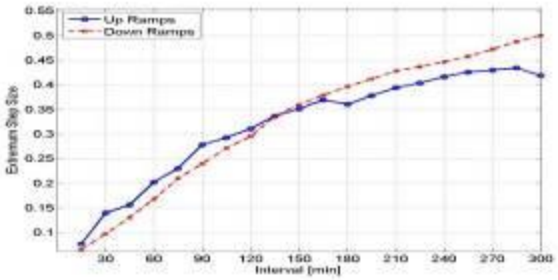


Fig. 10-3, Largest weather-induced ramps: same as Fig. 10-1 but for weather-induced normalized ramps (normalized by kW_{AC}).

Fig. 10-4 shows a histogram (by month) of the 1-hour weather-induced ramp rates of aggregate normalized power output (normalized by kW_{AC}) which is larger than 9% of PV capacity. The large weather-induced ramps were most likely in the winter months (primarily December and January) when they occurred about once per day. Presumably, this is related to overcast conditions (large morning down or evening up-ramp compared to the 30 day – mostly clear - average) or when storm systems moving into (large down ramp) or out of the area (large up-ramp). Large ramps are anti-correlated with the average output over a month.

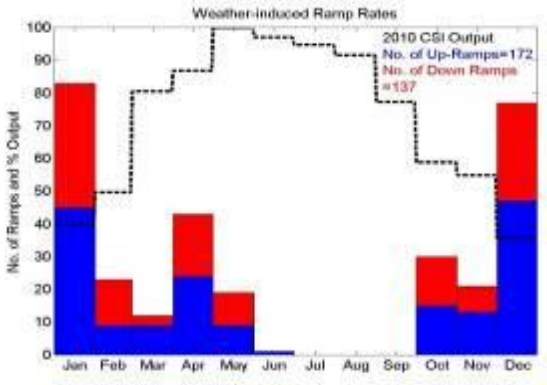


Fig. 10-4, Month of occurrence and direction of large weather-induced ramps: Same as Fig. 10-2a but for weather-induced ramps.

Days with the largest hourly weather-induced ramps

Fig. 10-5 shows daily profiles for the days when the four largest weather-induced ramps were observed. The largest ramp caused a change of 20% expressed in PTC capacity within one hour and was caused by a transition from overcast to clear during the afternoon when a large downramp was expected based on the 30 day average. The other large ramps are 16% or less and all result from overcast days with very small PV output. SAW estimates tracked the CSI power output typically within 9-15% rMAE at 30-min resolution.

The 15 minute consecutive GOES images for the time period with the largest weather-induced ramp (Fig. 10-5a) are illustrated in Fig. 10-6. A large cloud band that is parallel to the cost covers most PV sites south of the San Francisco bay area around noon, but rapidly clears the Southern California coastal area starting at 1300h. This day also caused the largest weather-induced ramp in SCE territory with a 29.8% magnitude. The greater geographic diversity of sites serves to reduce the ramp magnitude for the whole state by almost one-third.

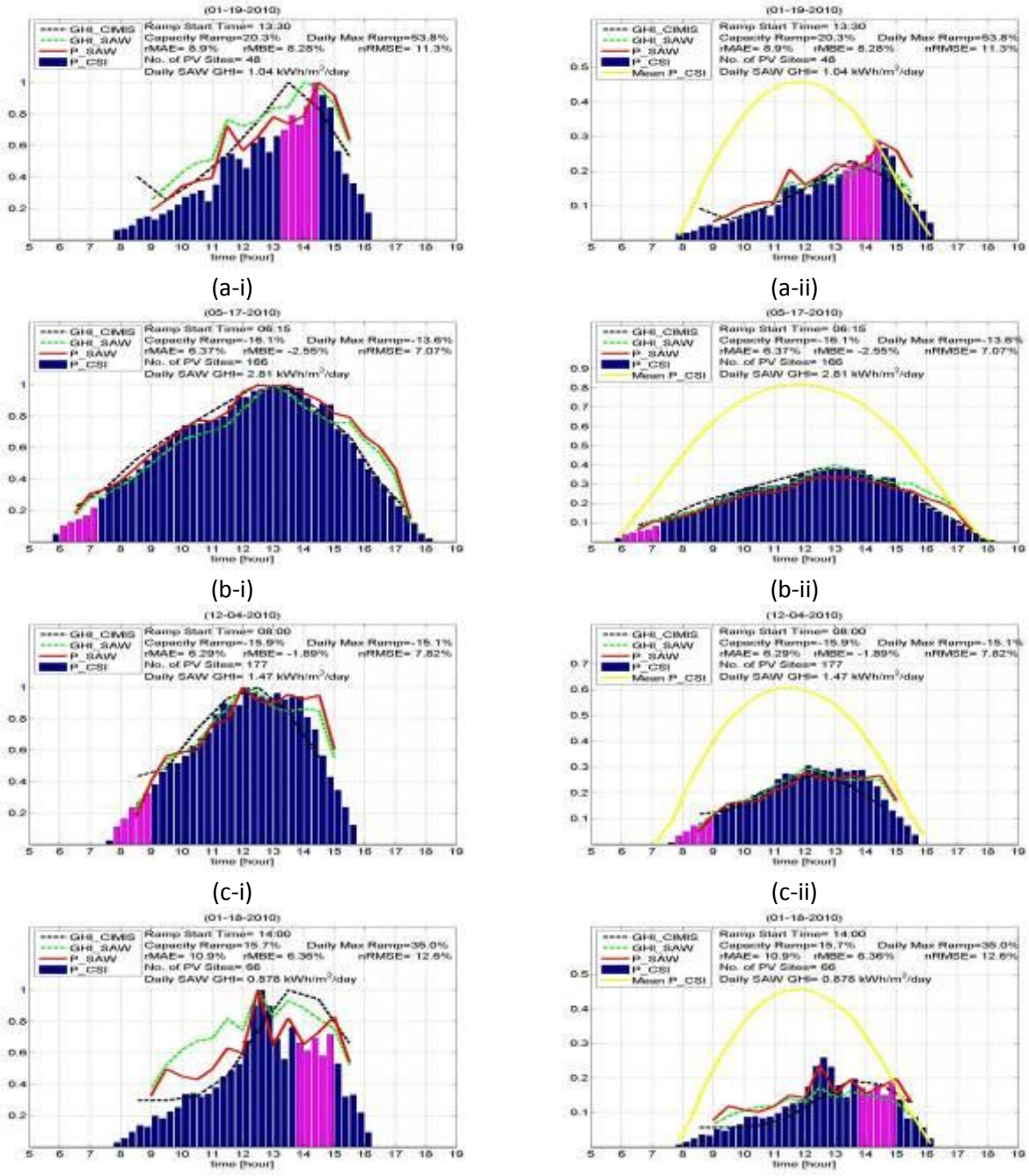


Fig. 10-5, Four days with largest weather-induced ramps: Normalized aggregate 15-minute PV output from all 192 PV sites (bars) for the days with the largest weather-induced ramp rates (aggregate power output minus 30 day average diurnal power output). Normalized (to a maximum of 1) aggregate 30-minute output (red) and GHI (green) obtained from SolarAnywhere at each pixel and normalized aggregate hourly measured GHI of 39 weather stations (black) are also shown. (a) Jan. 19, 2010, (b) May 17, (c) Dec. 4, 2010, and (d) Jan. 18, 2010. Magenta bars show the timing of the large ramp. The (ii) graphs also show the 30 day average diurnal power output (yellow lines) and are not normalized to 1.

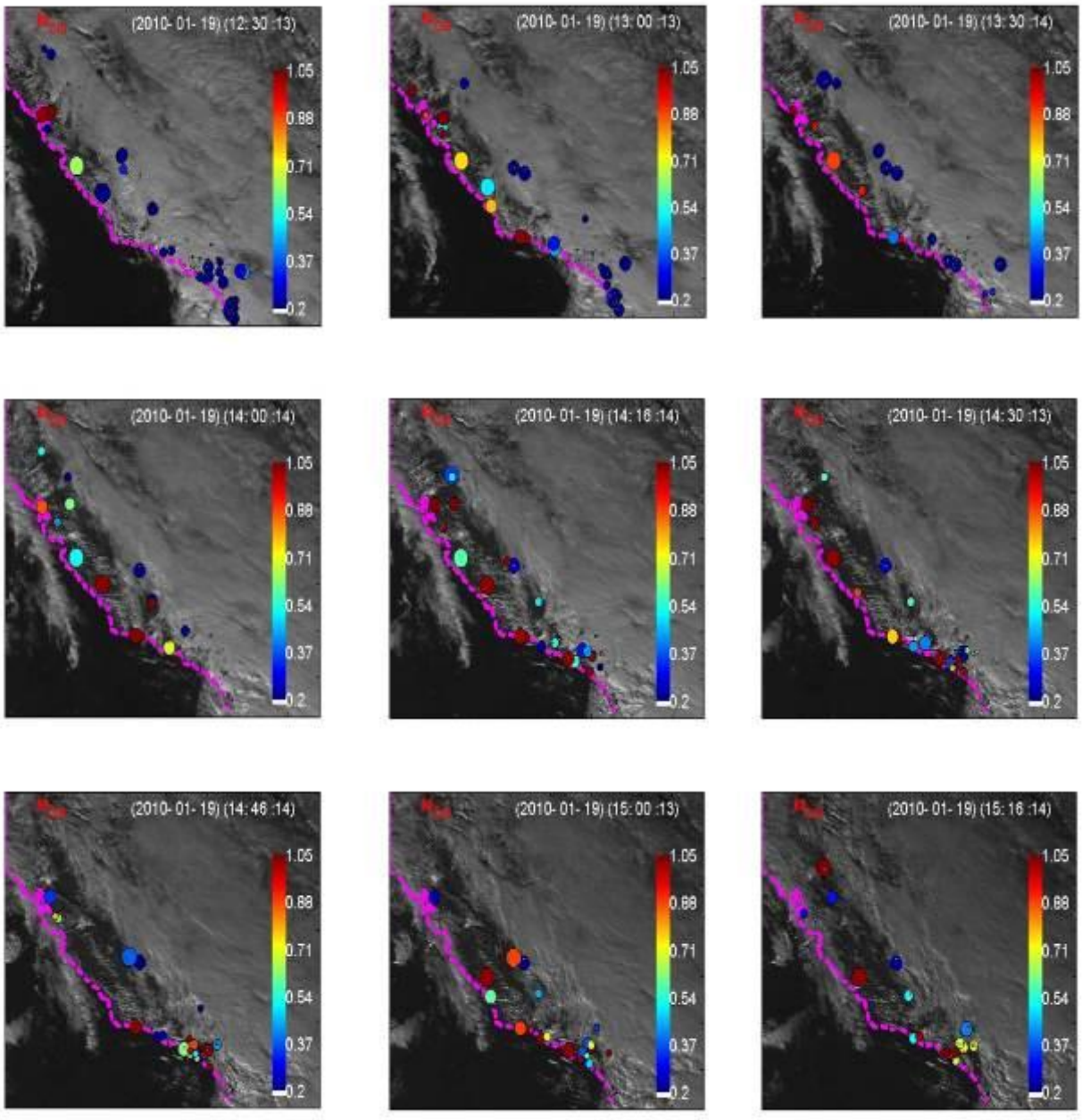


Fig. 10-6, GOES images for the day with the largest weather-induced ramp: GOES satellite images at 15 minute resolution on Jan. 19, 2010 (Fig. 10-5a). The circles represent 192 PV systems shown in Fig. 2-1b. The area of the circles is proportional to the power rating of the PV system and the largest system is 1000 kW. The color bar shows the ratio of 15-min averaged output to annual maximum output at that time of day (ToD). The largest aggregated 1 hour ramp for this period was 20% of PV capacity and occurred from 1330 to 1430 PST.

Conclusions

Aggregate power output of 192 PV systems in SDG&E, SCE, and PGE territories was presented and compared to satellite-derived Solar Anywhere irradiation and measured GHI at 39 weather stations (CIMIS). Results for the IOUs individually are presented in more detail separately on the CSI website.

The PV performance model applied to the satellite solar resource data was able to follow the power output measured over 192 systems typically within 2-10% during the four largest ramps. The largest hourly absolute and weather-induced ramps were 30% and 20% of PTC capacity respectively. In a very high PV penetration scenario, if such ramps hit the operator unprepared, they may indeed cause reliability challenges and additional costs for the system operator.

By investigating ramp rates of aggregate irradiance at clear sky conditions, the largest absolute ramp rates were found to be solely a result of diurnal cycles, and are therefore predictable. Weather-induced ramp rates with reference to a 30-day average of power output, on the other hand, were helpful to detect unexpected variations.

Since maximum absolute ramp rates during intervals shorter than 1 hour decrease about linearly with time interval, short term aggregate ramp rates are relatively benign at about 0.46% per minute. However, the analysis does not apply to single PV systems, which will show much larger short-term ramp rates that – in extreme cases - can result in exceedance of voltage tolerance bands on distribution feeders. This analysis was focused on distributed generation systems that are relatively well distributed across the state. Groups of larger but less geographically diverse systems may experience larger weather induced ramps. Table 10-2 compares the largest absolute and weather-induced ramp rates for different utility territories within California.

Table 10-2: Comparison of the largest ramp rates for different utility territories.

	SDG&E	SCE	PG&E	All
Largest absolute ramp	60.4%	31.4%	30.2%	30.9%
Largest weather induced ramp	55.5%	29.8%	28.1%	20.3%

References

[1]	"Web-based Clean Power Research service database, SolarAnywhere," [Online]. Available: https://www.solaranywhere.com/Public/About.aspx.
[2]	"CIMIS Sensor Specifications," [Online]. Available: http://wwwcimis.water.ca.gov/cimis/infoStnSensorSpec.jsp. [Accessed 15 Mar. 2011].
[3]	"Integrated Surface Irradiance Study (ISIS) Network," [Online]. Available: http://www.srrb.noaa.gov/isis. [Accessed 15 Mar. 2012].
[4]	"California Solar Initiative, "California Public Utilities Commission California Solar Initiative Program Handbook"," [Online]. Available: http://www.gosolarcalifornia.org/documents/CSI_HANDBOOK.PDF. [Accessed 15 Sep. 2011].
[5]	Lave, M, J Kleissl "Testing a wavelet-based variability model (WVM) for solar PV powerplants," in IEEE Power & Energy Society General Meeting, San Diego, CA, 2012.
[6]	Kimber, A, L Mitchell, S Nogradi, H Wenger "The Effect of Soiling on Large Grid-Connected Photovoltaic Systems in California and the Southwest Region of the United States," in IEEE 4th World Conference, 2006.
[7]	Mejia F, J Kleissl, Soiling Losses for Solar Photovoltaic Systems in California, Solar Energy, 95:357-363, 2013
[8]	Bosch JL, Y Zheng, J Kleissl, Deriving cloud velocity from an array of solar radiation measurements, Solar Energy, 87: 196-203, 2013, 10.1016/j.solener.2012.10.020.
[9]	M Jamaly, JL Bosch, J Kleissl, Aggregate Ramp Rates of Distributed Photovoltaic Systems in San Diego County, IEEE Transactions on Sustainable Energy, 99, 2012, 10.1109/TSTE.2012.2201966
[10]	Bryan Urquhart, Mohamed Ghonima, Dung Nguyen, Ben Kurtz, Chi Wai Chow and Jan Kleissl, Sky Imaging Systems for Short-term Forecasting, in: Solar Resource Assessment and Forecasting (Editor Jan Kleissl), Elsevier, 2013

Appendix

Appendix A: Summary of Quality Control for 2010 CIMIS Data

Data from146 CIMIS sites were processed for the year 2010. Of the 146 sites, 19 sites had 100 days or more of data missing. Data for the remaining 127 sites was checked for missing data (time steps greater than 1 hour). The missing data was replaced with 'NaN' values. For each site, the daily maximum power, the total daily energy production, the monthly maximum power, and the total monthly energy production were calculated. Data was also replaced with 'NaN' values when the clear sky GHI is zero.

Of the 127 sites, 70 sites were determined to be good sites, having reliable output for at least 265 consecutive days over the year. Sites were excluded if:

a) There was not 265 *continuous* days with data (frequently data for a site would be missing at the beginning of the year and also for a large timespan during or at the end of the year). (8 sites, Fig. A1)
b) There was a *small* shift in the magnitude of the output that seemed unusual. (21 sites, Fig. A2)
c) There was a *significant* shift in the magnitude of the output that seemed unusual. (15 sites, Fig. A3)
d) There was a *significant* shift in the magnitude of the output that seemed unusual, and the shift occurred directly after an outage in output (6 sites, Fig. A4)
e) There was a time shift between the CIMIS GHI data and the clear sky GHI. (1 site)
f) There was a spike in the output due to apparent recording issues or there appeared to be recording issues in general (4 sites, Fig. A5)
g) There was a spike in the output but no obvious recording issues (2 sites, Fig. A6)

Fig. A7 shows the location of all 127 CIMIS sites and the location of the 70 good sites.

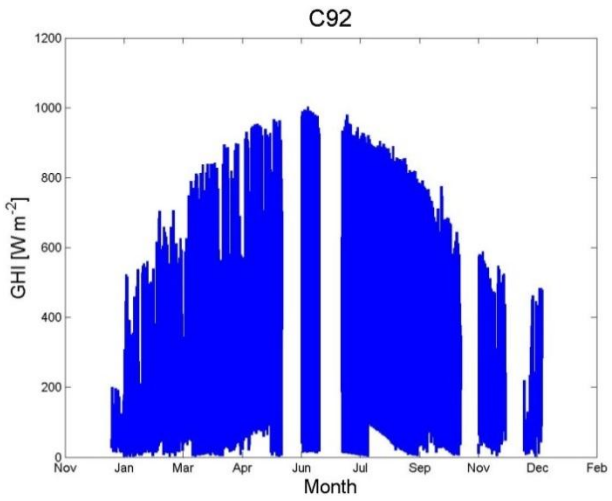


Fig. A1: Site does not have continuous 265 days of data.

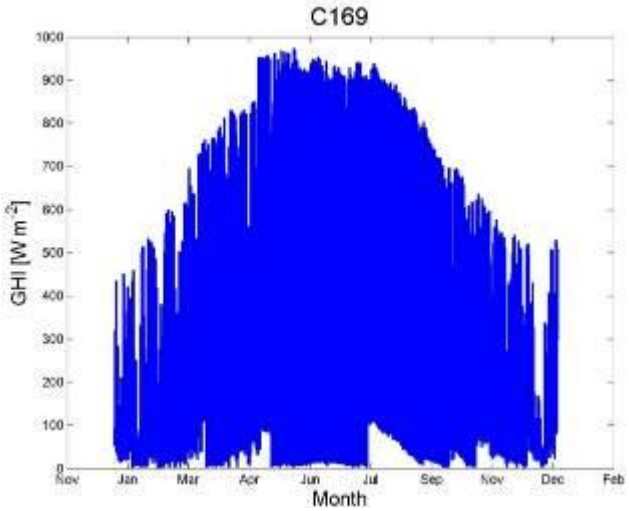


Fig. A2: Output increases unusually in April.

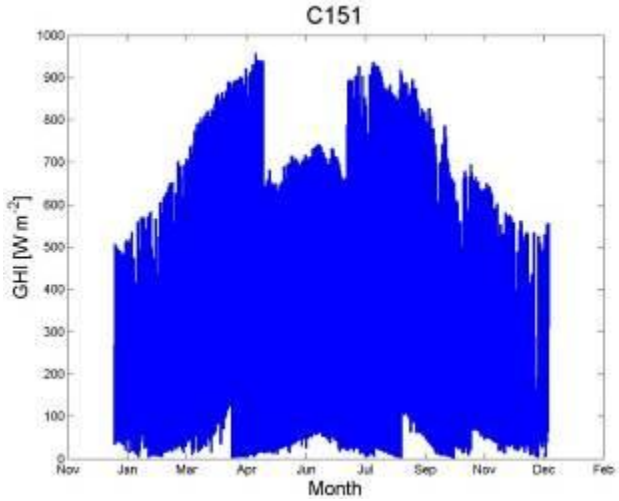


Fig. A3: Large descrease in output in April.

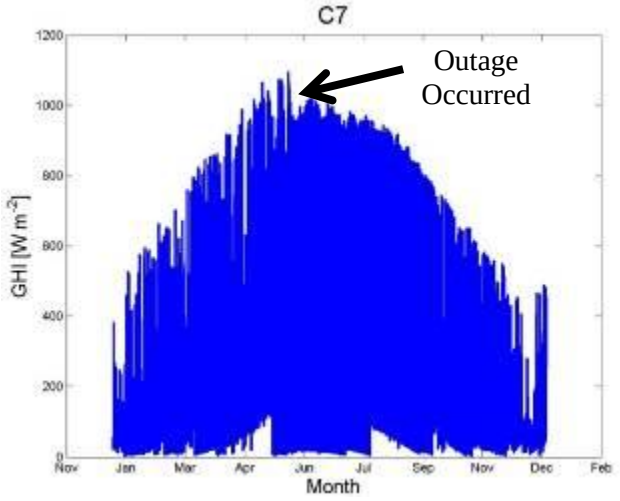


Fig. A4: Large descrease in output in May.

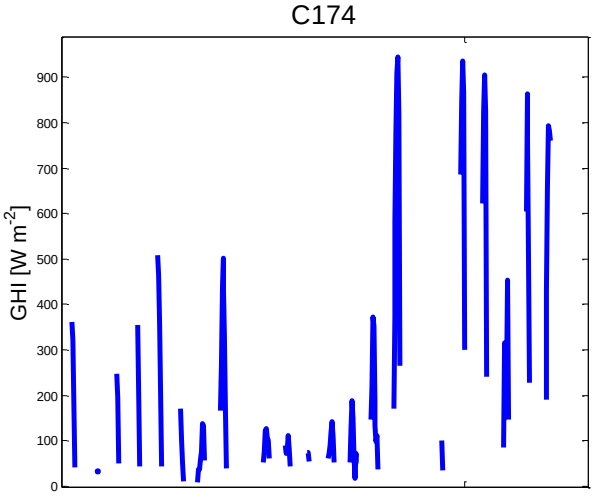


Fig. A5: Data exhibits recording issues.

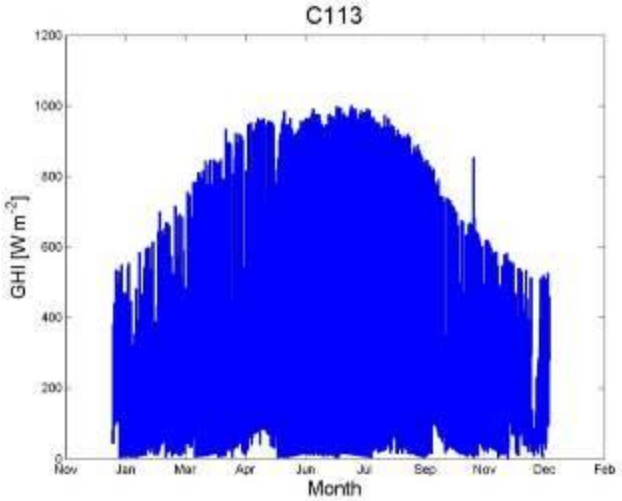


Fig. A6: Data has a spike in output.

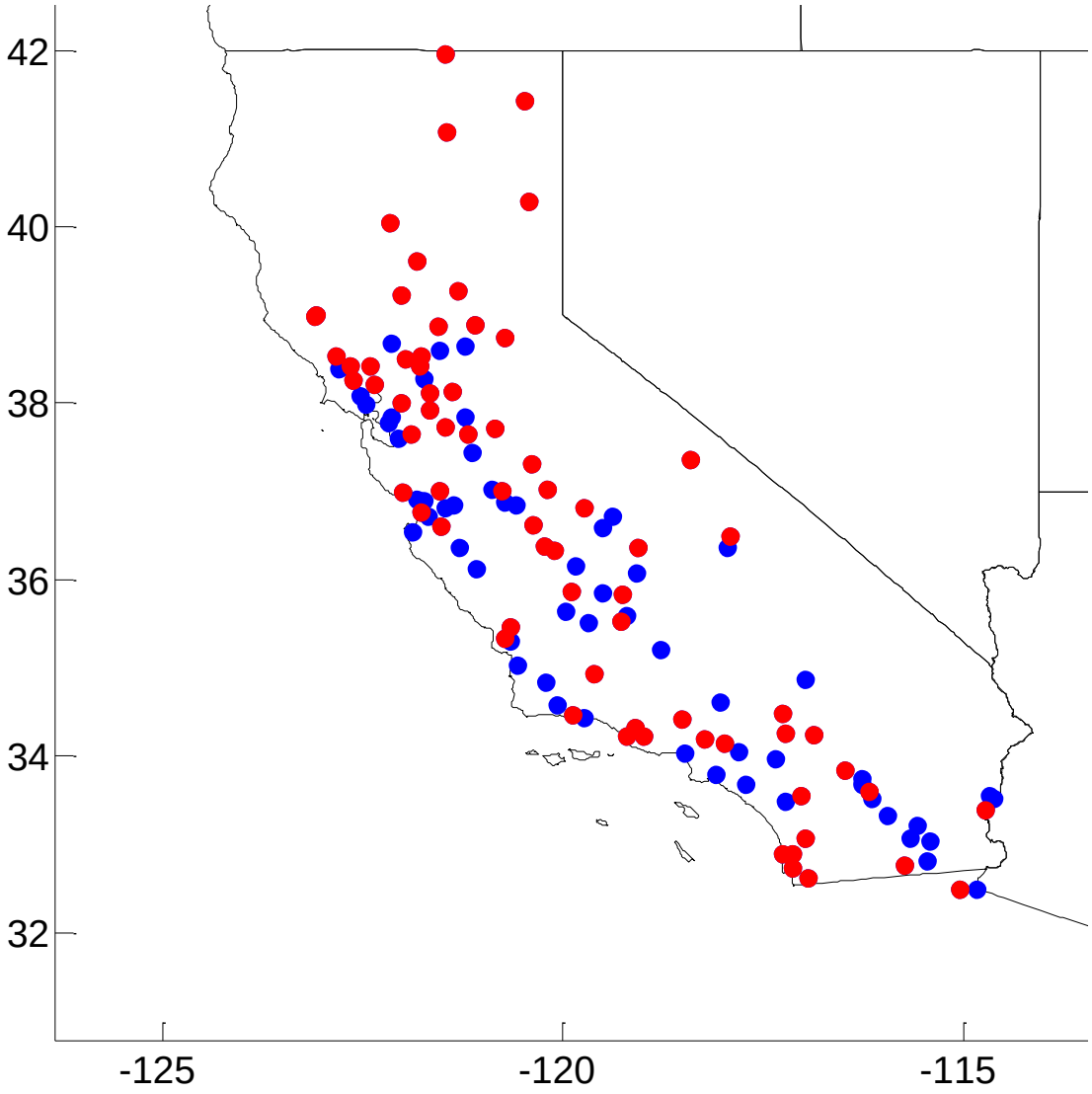


Fig. A7: All 127 sites that had data for 2010 are represented by a marker. Red markers indicate a good site; blue markers represent a site that was excluded.

Appendix B: Summary of Quality Control for CSI Data

Data from 981 sites was processed (194 from SDG&E, 384 from SCE, 403 from PG&E). Of the 981 sites, the location of the site was unknown for 5 sites and they were excluded. Fig. B6a shows the location for the 976 sites.

For the year 2010, the data was checked for missing data (time steps greater than 15 minutes). The missing data was replaced with 'NaN' values. Negative values were replaced with zero. For each site, the daily maximum power, the total daily energy production, the monthly maximum power, and the total monthly energy production were calculated.

Sites were excluded if:

a) There were less than 32 discrete power values over the entire year. This criterion resulted in the exclusion of most sites with 1 kWh data resolution (Fig B1), except for large systems. For example for a 128 kW system 1 kWh data resolution is sufficient to resolve output variability over 15 minute time steps (up to 32 kWh per 15 min). (Approximately 200 sites)
b) There was significant noise in the data (Fig. B2). (Approximately 15 sites)
c) It appeared that some strings were offline for a portion of the year (Fig B3). (Approximately 25 sites)
d) There were large spikes in power due to recording issues (Fig B4). For sites in which only a few spikes in power occurred for the entire year, the spike in power was replaced by the average of the power for the timestep just before through the timestep just after the power spike. (Approximately 25 sites)
e) There appeared to be decreases in power due to soiling, coupled with an increase in peak daily power on the order of 20% after a rainy day (Fig B5). (Approximately 25 sites)
f) There was not 265 days with data of 265 continuous days with data (frequently data for a site would be missing at the beginning of the year and also for a large timespan during or at the end of the year). (Approximately 200 sites)
g) Significant clipping of power due to undersized inverters occurred. (Approximately 15 sites)
h) The sites had single or dual axis tracking panels. (Approximately 25 sites)

Fig. B7 shows the frequency of good sites as a function of number of missing days. Fig. B6b shows the locations of the 286 good sites that have 100 missing days or fewer.

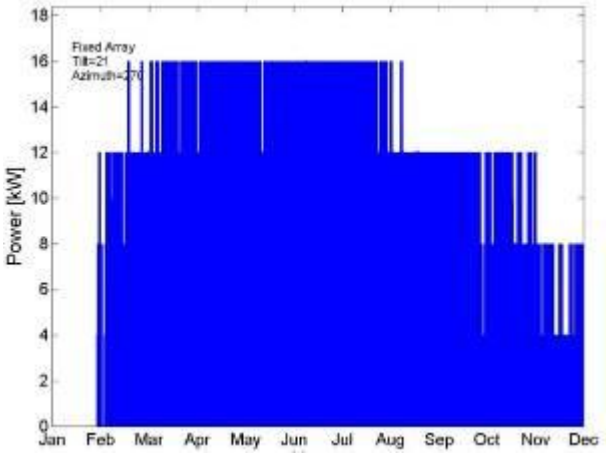


Fig. B1: Site that has blocky, low resolution data.

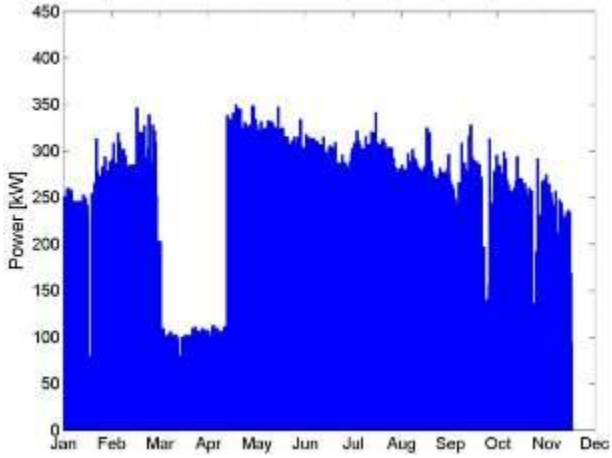


Fig. B3: Site that appeared to have some strings offline around March to April.

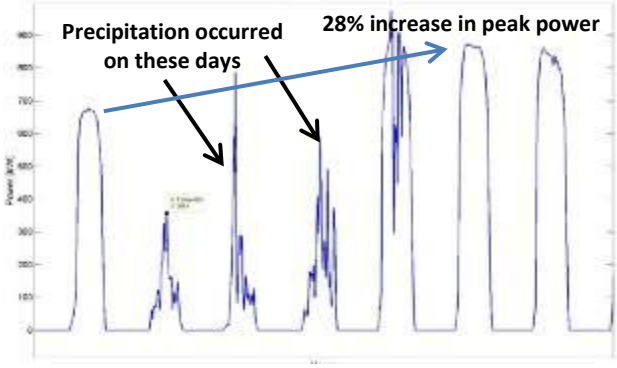


Fig. B5: Site that shows an increase in peak power after rainfall. These sites were likely not well maintained or improperly sited such that significant soiling occurred.

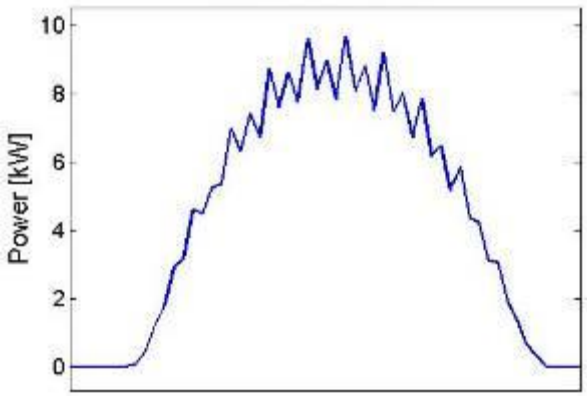


Fig. B2: Site that has noisy data presumably because recording intervals were not constant at 15 min, resulting in 'redistribution' of energy to neighboring periods.

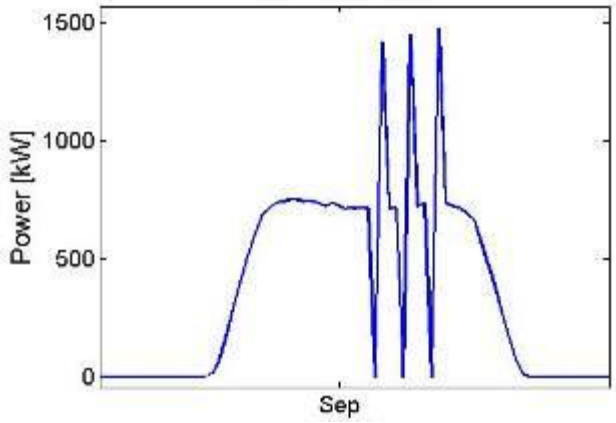


Fig. B4: Site with large, unrealistic spikes in power likely for the same reason as b). Sites with only a few such events were manually corrected.

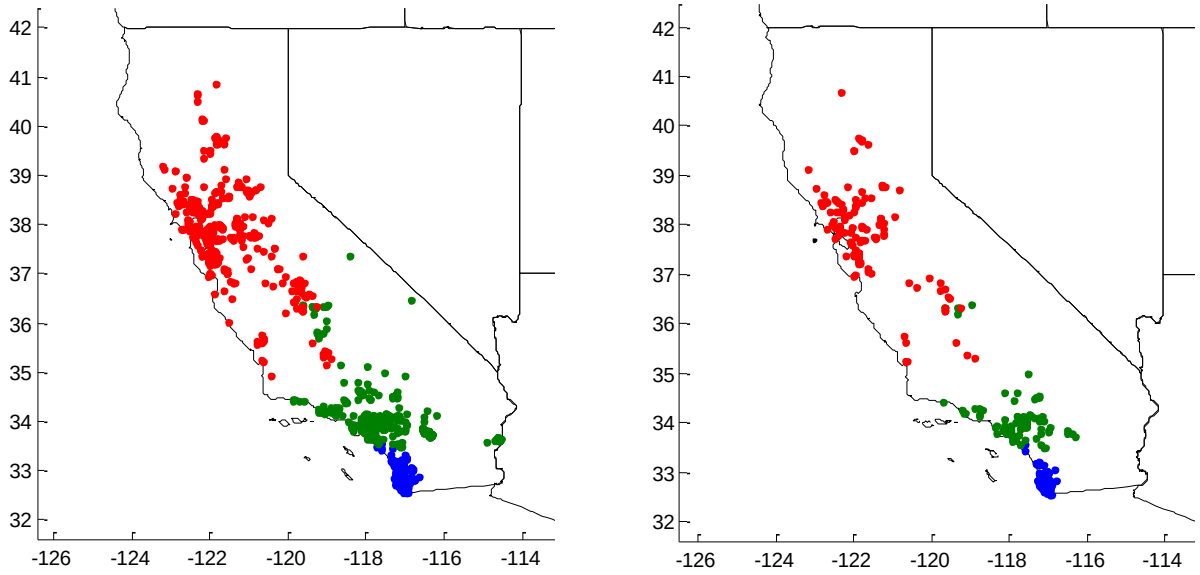


Fig. B6: (a) Site locations for all sites and (b) sites with no more than 100 days of missing data, excluding sites as described above. SDG&E sites are blue, SCE sites are green, and PG&E sites are red.

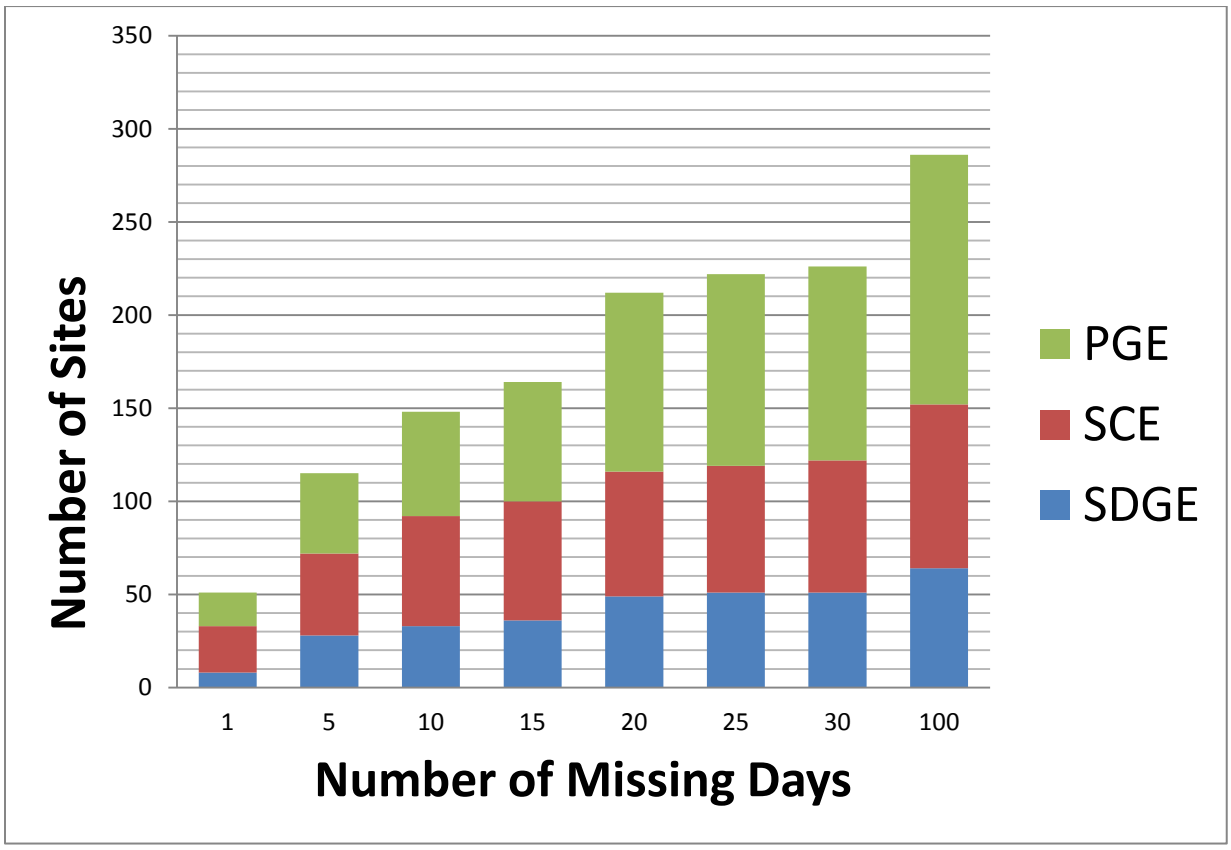


Fig. B7: Number of missing days versus the number of sites that have that many missing days or fewer, excluding sites as described above.

Appendix C Surveying Stakeholders for Solar Resource and Forecasting

C.1 Background

The California Solar Initiative (CSI) RD&D program is making a substantial investment in improved solar resource and solar forecasting data. The market for such products is relatively immature. To understand the needs of stakeholders a survey was sent to list serves of the California Energy Commission (CEC) and California Public Utilities Commission (CPUC) between October 2010 and January 2011. Participants self-identified as shown in Fig. C-1. Generally, survey questions asked to rate whether something was 5: very important to 1: least important. The scores given below are the average scores.

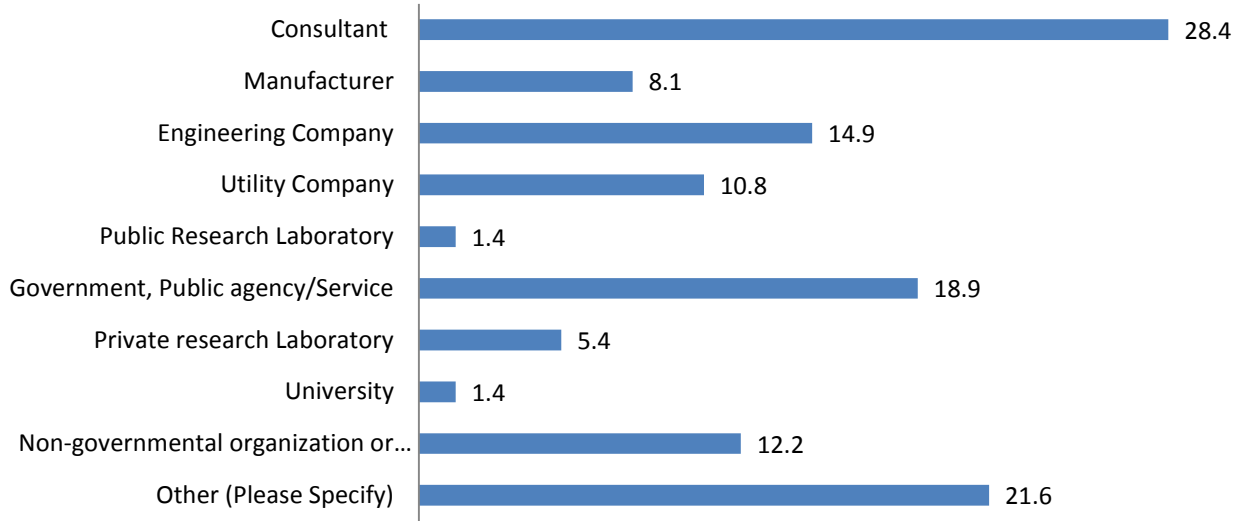


Fig. C-1: Self-identified profession of respondents [%].

The survey was divided into a solar resource survey and solar forecasting survey. 58 responses were received. Conditional results for utility responses are reported in bold after the overall results. For example 'PV was the most relevant technology' (4.3, **5**) means that the average of all respondents' scores was 4.3 (important to very important) and the average of the utility respondents was 5 (all voted very important).

However, except for seasonal forecasts, the majority of respondents indicated that either forecasts were not important (score of 1) or that the question was N/A. In addition, answers to the question 'Where do you currently get the solar forecasting information you need?' (28 respondents) revealed that most did not define solar forecasting in the same way as system operators, as 10 responded that they acquired such information from NREL, DOE, CSI, or CEC (these agencies do not provide solar forecast data). 9 responded that they did not use any forecasts. Consequently, the survey did not have the right respondents to address the solar forecasting questions and the results to the solar forecasting section of the survey will not be presented in detail.

C.2 General Results

Respondents were first asked which solar energy technology was most relevant to them (Fig. C-2). Photovoltaics (PV) was the most relevant technology (4.3, **5**) followed by CSP (3.65, **3**), Solar Heating (3.05, **2.8**), Building Engineering / Architecture (2.89, **1.6**), Solar Cooling (2.86, **2.6**), Water Desalination (2.6, **1.6**), and Chemical Systems (1.98, **1.25**). PV is by far the most relevant technology and – considering the continuing construction and financing trends in favor of PV and against CSP after the survey ended – PV would likely be even further ahead if the survey was repeated today.

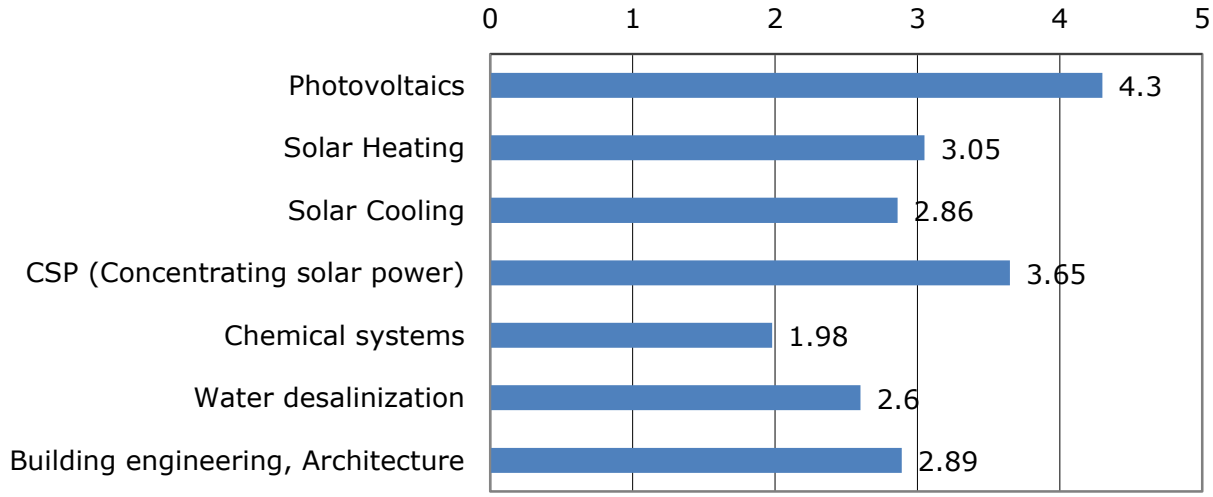


Fig. C-2: Responses to "Which technologies are most relevant to you? (Rate each item by 1: not important, to 5: very important)".

Depending on the professional field of the respondent, different aspects of these solar technologies would be more important (Fig. C-3). The most important aspects of solar technologies were Cost Assessment (4.29, **4.8**), Site Selection and Feasibility Study (4.09, **3.4**), System Design (4.09, **3.8**). Decommissioning and fault detection received the lowest scores. For utilities only, Monitoring (**4.4**), Investment Decision (4.2), and Grid Operation (4) also received high scores. The results reflect the current growth trajectory of solar energy where issues related to the build-out of solar power plants dominate over the operation of existing plants.

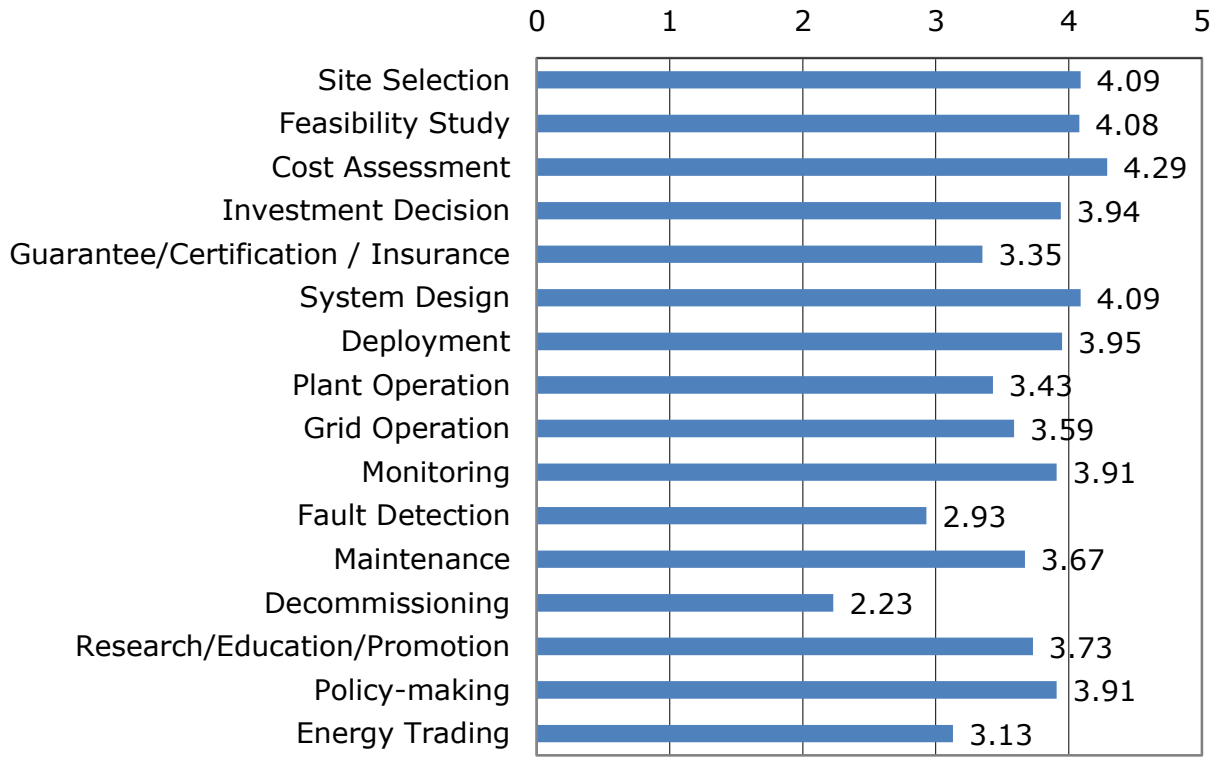


Fig. C-3: Responses to "What aspects of these solar technologies are you most interested in? (Rate each item by 1: not important, to 5: very important)".

C.3 Survey Results for Solar Resource Assessment

Depending on the application (Fig. C-2) different types of solar resource or ancillary meteorological data is required (Fig. C-4). The most important types of data were direct radiation (4.0, **3.6**), radiation on tilted and tracking surfaces (4.0, **3.8**), and ambient temperature (3.89). The high ratings of direct radiation data are surprising given the preference for PV over CSP (Fig. C-2); for PV only global radiation on titled surfaces is required. Global Horizontal Radiation was deemed relatively unimportant and atmospheric pressure was the least important (2.93), while utilities scored Global Horizontal Radiation equal to direct (**3.6**).

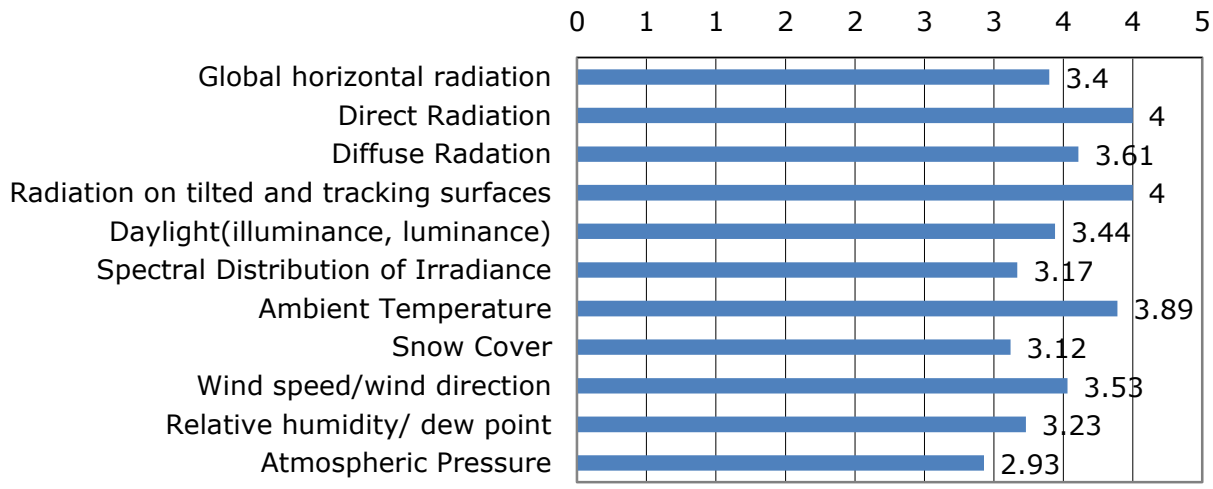


Fig. C-4: Responses to "Which type of data will be most useful/applicable to you (Rate each item by 1: not important, to 5: very important)".

Solar resource data are usually averaged over time periods from hours to years. While day-to-day variability in solar resource is large, the year-to-year variability is usually less than a few percent. The temporal resolution of solar resource data preferred by the respondents was annual (4.27, **3.75**), monthly (4.05, **4.25**), and hourly (4.01, **4.25**) averages. Subhourly values are considered much less important (2.4/**1.7** for instantaneous to 2.9/**2.8** for 15 minutes). Recent (last month / last year) and data from older archives was considered more important (4.0, **4.25**) than data from the last few hours or days. Consistent with Fig. C-3 the preference for planning solar power plants means that primarily long-term averages of the solar resource are needed for siting and cash flow analysis.

Geographically, averaged data gridded over a large region was considered the most important (3.7, **4.0**), but many respondents also considered single site data to be valuable (3.2), whereas utility respondents preferred at least several sites (**4.3**) compared to a single site (**2.7**). The spatial resolution (Fig. C-5) should be 6 x 6 miles or better (3.9, **4.7**) as is currently provided in the National Renewable Energy Laboratory (NREL) National Solar Radiation Database (NSRDB). Resolutions used in the NREL PVWATTS solar power production tool of 60 x 60 miles were deemed suboptimal (3.18, **3.33**). Most respondents used solar resource observations as input into simulators and software (4.1, 3.8) or used the observations themselves (3.62).

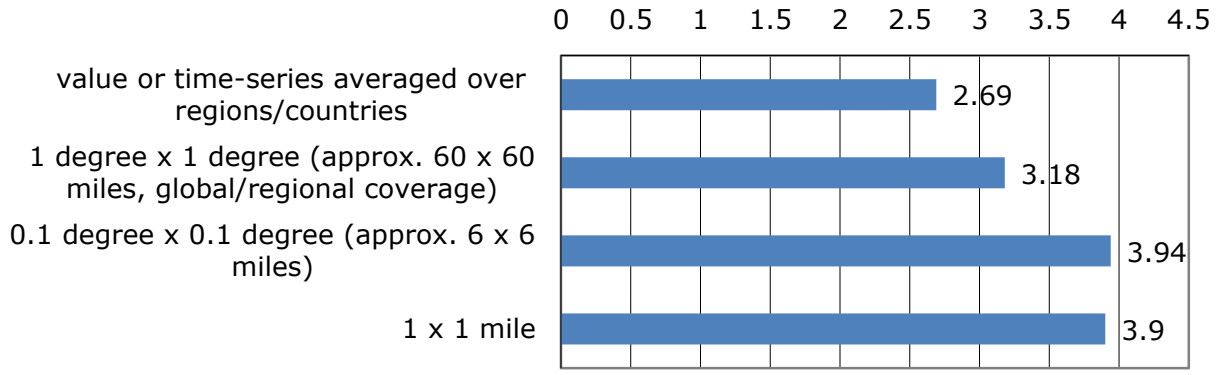


Fig. C-5: Responses to "What spatial resolution of maps (what pixel size) do you think is needed for your main purpose? (Rate each item by 1: not important, to 5: very important)".

Respondents were also asked "Is the solar resource and meteorological information currently available satisfactory when considering various factors. There is considerable room for improvement in the currently available data (overall rating of the experience with the data is 2.9, **2.5**), especially in up-to-date-ness and clarity in description of products for all respondents (both 2.8) and access and geographic coverage for utility respondents (**2.5**).

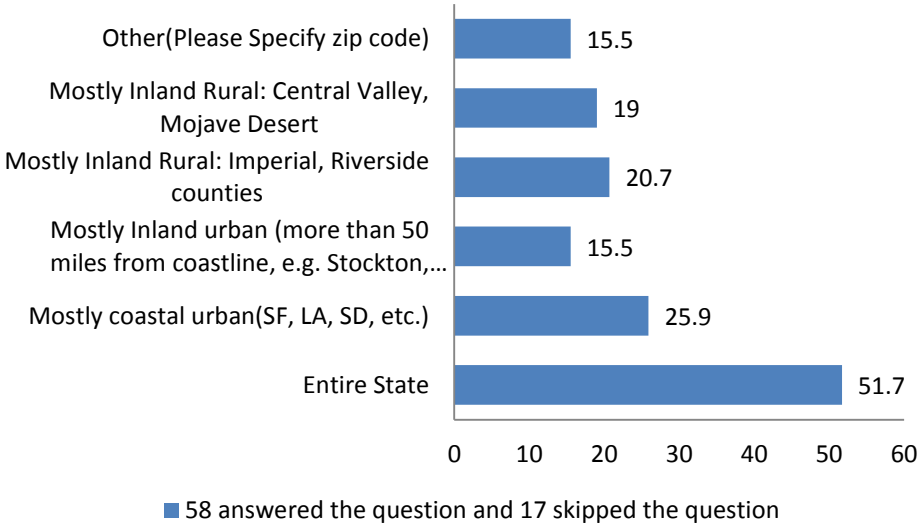


Fig. C-6: Regions of interest for solar resource data [%].

Most respondents were interested in solar resource data for the entire state, while utility respondents were obviously only interested in their specific territory (Fig. C-6). Coastal urban (26%) received higher ratings than inland rural (20%), and inland urban (16%). The majority of respondents (60%) get their current data from NREL through the National Solar Radiation Database (NSRDB), PVWATTS, or the Solar Prospector. A few respondents use their own meteorological information, NASA, or 3Tier. Utilities used NREL, SolarAnywhere, and local stations. Some specific comments raised by (one user each) were request for estimates of year-to-year variability, more historical data, and specific data for urban areas to help communities and municipalities select buildings for solar installations, and solar power prospector-like tools for other countries.

The results indicate that temporal and spatial resolution of existing gridded solar resource data including the NREL NSRDB is sufficient for most applications. However, improvements are needed in keeping data up to date (achieved through the Clean Power Research CSI RD&D grant) and clarity of the product descriptions. Most respondents were primary users of solar resource data and were mostly using free data sources from established sources such as NREL.

C.4 Implications for the California Solar Initiative RD&D Program

- For solar resource data the entire state should be covered, but most effort to improve accuracy should go into coastal urban areas and inland rural areas in southern California.
- Improvements in accuracy, clarity, and format of solar resource data are considered most important, while improvements in spatial resolution beyond 6 x 6 miles were not deemed critical.
- Collaboration with NREL should be sought to serve improved solar resource data through NREL tools, which are predominantly used in the industry. However, the industry is developing improved solar resource models that may be better able to address metadata issues noted above.