Integrating PV into Utility Planning and Operation Tools

Final Report
April 28, 2014

Project Partners:	Clean Power Research (lead)
	California Independent System Operator
	Sacramento Municipal Utility District
	Pacific Gas and Electric Company
	State University of New York
	Solar Electric Power Association
	UC San Diego
	Electric Power Research Institute
Project Manager:	Smita Gupta (Itron)
Investigators:	Thomas E. Hoff
	Benjamin Norris
	Phil Gruenhagen
	Richard Perez
	Andrew Parkins
	Kristian Lieberg

Acknowledgements

This project was a success as a result of the contributions of people at a variety of organizations. The California Public Utilities Commission (CPUC), Sacramento Municipal Utility District (SMUD), and Pacific Gas and Electric Company (PG&E) provided support. Smita Gupta (Itron) managed the grant in such a way so as to result in a usable set of products, and Ann Peterson (Itron) provided support. Jim Blatchford and his team (California ISO) and Obadiah Barthomy and his team (SMUD) provided valuable direction and support from the independent system operator and municipal utility perspectives. Mike Taylor and his team (SEPA) was effective at disseminating information. Dr. Richard Perez and his team (SUNY) made valuable SolarAnywhere advancements, including the production of High Resolution (1 km, 1 minute) data. Dr. Jan Kleissl and his team (UC San Diego) assessed the performance of SolarAnywhere. Thanks to all of these individuals and organizations, as well as many others, for their support and assistance.

Contents

Acknowledgements 2

Abstract 5

Executive Summary 6

1. Introduction 11

2. Task 1: SolarAnywhere Data 12

   2.1. Introduction 12

   2.2. SolarAnywhere Enhanced Resolution 12

   2.3. SolarAnywhere High Resolution 13

   2.4. SolarAnywhere Validation 14

   2.5. Conclusions 24

3. Task 2: Validate PV Fleet Simulation 25

   3.1. Introduction 25

   3.2. Results for Large PV Systems Connected to CAISO 25

   3.3. Results for Small Distributed PV Connected to SMUD 26

   3.4. Conclusions 27

4. Task 3: Integrate PV Fleet Simulation Into Utility Software Tools 29

   4.1. Introduction 29

   4.2. Original Plan 29

   4.3. Revised Plan 29

   4.4. PV Fleet Simulation 30

   4.5. Solar Resource Data 30

   4.6. PV System Specifications 30

   4.7. PV Simulation Model 31

   4.8. Challenges 31

   4.9. Obtain Data 31

   4.10. Behind-the-meter PV System Specifications 31

   4.11. Perform Simulation Quickly 32

   4.12. Keep Data Up-to-Date 32

   4.13. PV Fleet Simulation Validation (CAISO Data) 33

   4.14. PV Fleet Simulation Validation (SMUD Data) 45

5. Technology Transfer 55

6. Conclusions 57

6.1. Key Findings 57

6.2. Benefits to California Ratepayers 57

6.3. Potential Next Steps 58

Abstract

The California Solar Initiative (CSI) has a goal of installing 3,000 MW of new solar electricity by 2016. CSI has identified that one potential barrier to accomplishing this goal is planning and modeling for high-penetration PV grid integration issues. A team led by Clean Power Research (CPR) received approval from the California Public Utilities (CPUC) for a grant titled, "Integrating PV into Utility Planning and Operation Tools." CPR led the team in the development, validation, and integration of PV fleet simulation tools that enable utilities and ISOs to cost-effectively integrate distributed PV resources into their planning, scheduling and operating strategies.

This project builds upon Clean Power Research's CSI Grant Solicitation #1 award, which was awarded in 2010 and titled, "Advanced Modeling and Verification for High Penetration PV." Two key accomplishments under that award were: production of a publicly-available enhanced resolution solar resource database for every location in California (SolarAnywhere® Enhanced Resolution, 1-km, 30-minute resolution, available at www.SolarAnywhere.com); and development of an advanced methodology to simulate PV fleet power production for any PV fleet configuration (SolarAnywhere FleetView).

This current project accomplishes the following grid-integration tasks:

1. Extend the SolarAnywhere Enhanced Resolution solar resource database, create high resolution (1-km, 1-minute resolution) solar resource data, and benchmark data accuracy.
2. Validate previously developed PV fleet simulation methodologies using measured ground data from fleets of PV systems connected to California's grid.
3. Integrate PV fleet simulation methodologies into utility software tools for use in activities ranging from distribution planning to balancing area operation using CAISO as a test case.

The tools and data streams developed as part of this work will be made available using CPR's existing software services (e.g., www.SolarAnywhere.com) to California utilities, ISOs and others to help cost-effectively and reliably integrate distributed PV into the grid.

Executive Summary

Introduction

The California Solar Initiative (CSI) has a goal of installing 3,000 MW of new solar electricity by 2016. CSI has identified that one potential barrier to accomplishing this goal is planning and modeling for high-penetration PV grid integration issues. A team led by Clean Power Research (CPR) received funding from the California Public Utilities Commission (CPUC) titled, "Integrating PV into Utility Planning and Operation Tools." The team included the California Independent System Operator (CAISO), Sacramento Municipal Utility District (SMUD), Pacific Gas and Electric Company (PG&E), State University of New York (SUNY), Solar Electric Power Association (SEPA), UC San Diego (UCSD), and Electric Power Research Institute (EPRI). CPR led the team in the development, validation, and integration of PV fleet simulation tools that enable utilities and ISOs to cost-effectively integrate distributed PV resources into their planning, scheduling and operating strategies.

Project Objectives

This project builds upon CPR's CSI Grant Solicitation #1 award, which was granted in 2010 and titled, "Advanced Modeling and Verification for High Penetration PV." Two key accomplishments under that award were: production of a publicly-available enhanced resolution solar resource database for every location in California (SolarAnywhere® Enhanced Resolution, 1-km, 30-minute resolution, available at www.SolarAnywhere.com); and development of an advanced methodology to simulate PV fleet power production for any PV fleet configuration (SolarAnywhere FleetView).

This current project accomplishes the following grid-integration tasks:[1]

1. Extend the SolarAnywhere Enhanced Resolution solar resource database, create high resolution (1-km, 1-minute resolution) solar resource data, and benchmark data accuracy.
2. Validate previously developed PV fleet simulation methodologies using measured ground data from fleets of PV systems connected to California's grid.
3. Integrate PV fleet simulation methodologies into utility software tools for use in activities ranging from distribution planning to balancing area operation using CAISO as a test case.

Results

Task 1: SolarAnywhere Data

The first task was to extend the SolarAnywhere Enhanced Resolution solar resource database (1-km, 30-minute resolution), to create high resolution (1-km, 1-minute resolution) data, and to benchmark data accuracy.

[1] The Grant Agreement lists four tasks with Task 1 being "Project Management." This report does not include the Project Management task. As such, all tasks are shifted back by one for this report. For example, Task 2 in the Grant Agreement corresponds to Task 1 in this report.

CPR continued to update the SolarAnywhere Enhanced Resolution data throughout the project. The data is publicly available at www.solaranywhere.com.

The updated SolarAnywhere Enhanced Resolution data was then used as an input to create the SolarAnywhere High Resolution (1-km, 1-minute) solar resource data. This was accomplished by applying the Cloud Motion Vector (CMV) approach developed by Dr. Richard Perez. The CMV approach projects cloud movement by comparing two consecutive enhanced resolution images. The result produces high temporal resolution (1-minute) data that is then used to create the SolarAnywhere High Resolution solar resource data.

It was challenging to produce such state-of-the-art data. It was even more challenging to produce the data at a speed fast enough to make it available for PV fleet forecasting for several hundred thousand individual PV systems every 30 minutes. CPR had to move its code base from running on servers in a local datacenter to a “massively parallel” architecture using Internet “cloud” computing. CPR performed this transition over a multi-month period that included designing, porting, testing, and running the system. The cloud computing approach allows CPR to flexibly and efficiently add additional compute power as the number of PV systems grows or as simulation time horizons are adjusted. By the end of the process, CPR was able to produce solar forecasts that could be used to forecast production for 170,000 PV systems (current number of systems as of the writing of this report) every 30 minutes. The system has been operating well for almost one year.

The accuracy of the SolarAnywhere data was validated in conjunction with a CEC project, titled, “Demonstration and Validation of PV Output Variability Modeling,” Project number CEC 500-10-059 (see Appendix 1). Results indicated that the SolarAnywhere High Resolution data is more accurate than SolarAnywhere Enhanced Resolution data which, in turn, is more accurate than SolarAnywhere Standard Resolution data.

Task 2: Validate PV Fleet Simulation

The second task was to validate PV fleet simulation methodologies using measured ground data from fleets of PV systems connected to California’s grid. This task was accomplished using measured PV production data for fleets of systems from two separate sources. CAISO provided data for large PV plants connected to its system. SMUD provided data for small distributed PV systems connected to its system.

CAISO provided measured fifteen-minute production data for 46 metered PV plants from March 10, 2013 to April 19, 2013. Results suggest that the relative mean absolute error (rMAE) ranged from 3 to 7 percent depending upon the level of model tuning and data filtering. SMUD provided measured hourly production data for 2,206 distributed PV systems from April 16, 2012 to October 10, 2012. SMUD also provided specifications for all of the PV systems. Model tuning was not applied to this data set. Results indicate an accuracy of 6 percent rMAE. The overall conclusion was that SolarAnywhere’s PV fleet simulation capabilities result in fairly accurate results.

Task 3: Integrate PV Fleet Simulation Into Utility Software Tools

The third task was to integrate PV fleet simulation methodologies into utility software tools using the results from Tasks 1 and 2.

A fleet of PV systems can be defined very broadly. At one extreme, a fleet can refer to single PV system on the roof of one person's house. At the other extreme, a fleet can refer to all PV systems located within a balancing area across a state or even all PV systems in the U.S. or the world. User-defined collections of systems ("virtual" fleets) based on location, system attributes or other criteria are useful for planning and modeling purposes.

PV fleet simulations can be based on historical, real-time, or forecasted solar resource data. Historical data is useful for system planning. Real-time data is useful for assessing PV fleet operation. Forecasted data is useful for determining how to operate the rest of the utility system.

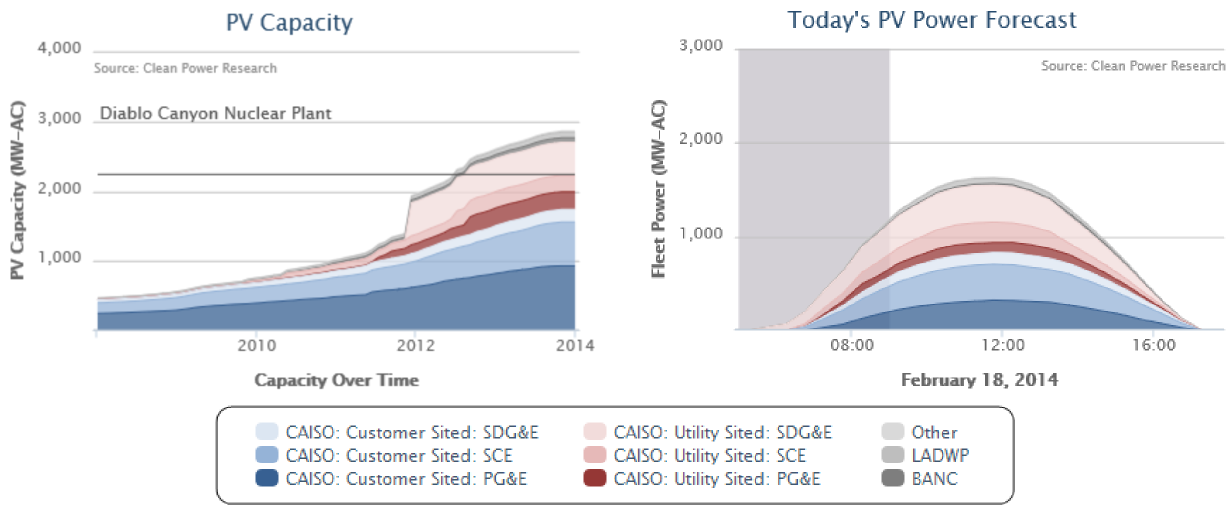
This broad definition of a PV fleet makes the simulation capabilities useful across a wide spectrum of utility applications. These applications range from planning and smart grid operation in the distribution system, utility load scheduling in the utility system, and balancing area planning and operation functions covering multiple utilities.

CAISO is responsible to maintain reliability and accessibility for California's utility grid. As such, they are concerned with the effect of power production from customer-owned PV systems on the balancing area. CAISO clearly understood the performance of the large PV systems through detailed production monitoring. They had no visibility, however, into the performance of the behind-the-meter PV systems. This was a concern to them. They needed to forecast behind-the-meter PV fleet performance.

It was initially planned to demonstrated limited PV fleet simulations across the variety of possible applications. Instead, it was determined that the greatest market need was the most complex application originally anticipated: behind-the-meter PV fleet forecasting for an entire balancing area. Furthermore, it was clear that what was needed was a full-scale application, not a limited scope test. As a result, CPR designed, tested, and implemented a PV fleet forecasting system that included all distributed PV systems in the state of California.

The three critical elements in performing a PV fleet simulation include: high resolution solar resource data, PV system specifications, and a simulation model to convert this information into production. The solar resource data was developed in Task 1. Detailed specifications for all PV systems in California were collected as part of a partner CEC project. The fleet simulation model was validated in Task 2. The result is that CPR began generating high resolution forecasts every 30 minutes for the entire state based on detailed PV system specifications for all behind-the-meter and metered PV systems. At the time of this report, there were 170,000 PV systems. The result is the capability of forecasting PV fleet output for the entire state of California as illustrated in the following figure.

Figure 1. California solar resource portfolio (Feb. 18, 2014).



Key Findings

This project built on the results produced under CSI RD&D Program Solicitation #1 for the CPR proposal entitled "Advanced Modeling and Verification for High Penetration PV." Key conclusions from this work are:

- High resolution solar resource data can be accurately produced.
- This solar resource data can be combined with PV system specifications to accurately simulate PV fleet production.
- The simulation process can be performed quickly enough to support even the challenging application of forecasting production for hundreds of thousands of systems while meeting forecasting time horizon requirements using the appropriate computing resources and underlying system architecture.

Benefits to California Ratepayers

This project has provided a number of benefits to the state of California.

Solar Resource Data

The first task was to extend SolarAnywhere. SolarAnywhere Enhanced Resolution provides 1 km spatial resolution with half-hour temporal resolution irradiance data. It is beneficial in that it is comprehensive for all of California and is freely available at www.SolarAnywhere.com. California project developers are also leveraging the increased Enhanced Resolution data accuracy to obtain lower financing rates because of reduced project risk; this lowers the cost of solar and increases the penetration of PV in the state. SolarAnywhere High Resolution extends the Enhanced Resolution to one-minute temporal resolution. The High Resolution data is used in PV penetration and variability studies.

PV Fleet Simulation Validation

The second task was to validate PV fleet simulation methodologies using measured ground data from fleets of PV systems connected to California's grid. It is critical to the utilities and balancing area authorities responsible to run the grid that they validate models using real-world data. The validation provides public benefits because grid operators need to gain confidence in the models intended to inform grid operation prior to their use.

PV Fleet Simulation Integration Into Utility Software Tools

The third task was to integrate PV fleet simulation methodologies into utility software tools. CAISO has the responsibility of maintaining reliability and accessibility for California's utility grid. As such, they are concerned with the effect of power production from customer-owned PV systems on the balancing area. Prior to this contract, CAISO did not have visibility into the performance of behind-the-meter PV systems. CPR has been providing behind-the-meter PV fleet forecasts every 30 minutes to CAISO for one year. This is beneficial to California in that CAISO has visibility into behind-the-meter PV performance when none existed prior to this grant. It has the additional benefit of being a valuable case study for California's IOUs as they consider using the same approach for their needs.

1. Introduction

The California Solar Initiative (CSI) has a goal of installing 3,000 MW of new solar electricity by 2016. CSI has identified that one potential barrier to accomplishing this goal is planning and modeling for high-penetration PV grid integration issues. A team led by Clean Power Research (CPR) received approval from the California Public Utilities (CPUC) for a grant titled, "Integrating PV into Utility Planning and Operation Tools."

CPR assembled a team that included the California Independent System Operator (CAISO), Sacramento Municipal Utility District (SMUD), Pacific Gas and Electric Company (PG&E), State University of New York (SUNY), Solar Electric Power Association (SEPA), UC San Diego (UCSD), and Electric Power Research Institute (EPRI). CPR led the team in the development, validation, and integration of PV fleet simulation tools that enable utilities and ISOs to cost-effectively integrate distributed PV resources into their planning, scheduling and operating strategies.

This project builds upon CPR's CSI Grant Solicitation #1 award, which was granted in 2010 and titled, "Advanced Modeling and Verification for High Penetration PV." Two key accomplishments under that award were: production of a publicly-available enhanced resolution solar resource database for every location in California (SolarAnywhere® Enhanced Resolution, 1 km, half-hour resolution, available at www.SolarAnywhere.com); and development of an advanced methodology to simulate PV fleet power production for any PV fleet configuration (SolarAnywhere FleetView).

This current project will accomplish the following grid-integration tasks:

1. Extend the SolarAnywhere Enhanced Resolution solar resource database to create high resolution (1-km, 1-minute resolution) data and benchmark data accuracy.
2. Validate previously developed PV fleet simulation methodologies using measured ground data from fleets of PV systems connected to California's grid.
3. Integrate PV fleet simulation methodologies into utility software tools for use in activities ranging from distribution planning to balancing area operation using CAISO as a test case.

The tools and data streams developed as part of this work will be made available to California utilities, ISOs and others to help cost-effectively and reliably integrate distributed PV into the grid.

2. Task 1: SolarAnywhere Data

The first task was to extend the SolarAnywhere Enhanced Resolution solar resource database, to create high resolution (1-km, 1-minute resolution) data, and to benchmark data accuracy.

2.1. Introduction

SolarAnywhere is a subscription-based, online satellite-based irradiance dataset (www.SolarAnywhere.com) currently available from Mexico to Canada. Prior to CPR's CSI Phase 1 grant, it contained hourly irradiance data at a 10-km by 10-km spatial resolution and 1-hour temporal resolution dating back from 1998. The functionality of SolarAnywhere was extended in three ways under CPR's CSI Phase 1 grant: (1) finer spatial resolution (1-km by 1-km grid); (2) finer temporal resolution (30-minute interval); and (3) freely available to users throughout California for the term of the project. The resulting product was referred to as SolarAnywhere Enhanced Resolution.

The objective of this task was to continue to provide SolarAnywhere Enhanced Resolution data for the duration of this contract, to extend the data to include SolarAnywhere High Resolution data (1-km by 1-km grid, 1-minute interval), and to benchmark data accuracy.

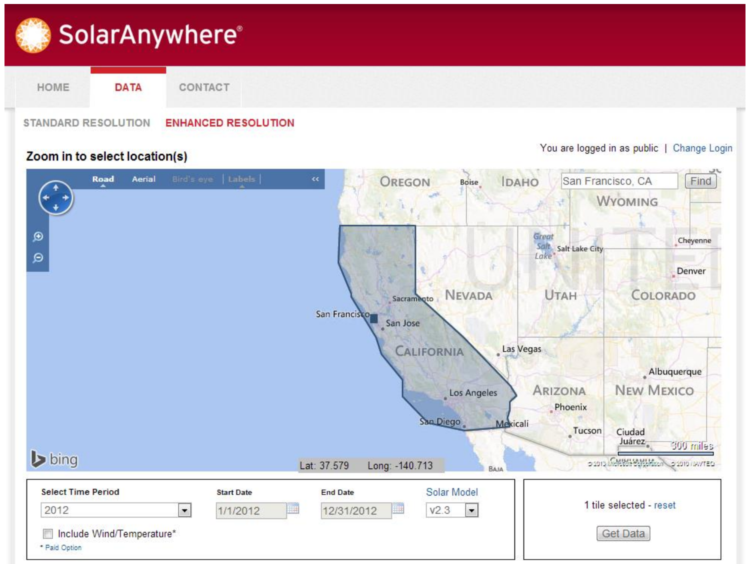
2.2. SolarAnywhere Enhanced Resolution

Production of the SolarAnywhere Enhanced Resolution data (1-km spatial resolution, 30-minute temporal resolution) began under CPR's CSI Phase 1 contract. Production was continued in order to keep it up-to-date and accessible.

The data has been used in multiple ways:

- The data is publicly available through the www.solaranywhere.com website (see Figure 2).
- High Resolution data uses Enhanced Resolution data as a core input.
- The data was used in some of the other tasks for this current project as they relate to tools and reports.
- Others have used the data for CPUC studies. For example, CPUC's subcontractor E3 used the data to perform a detailed net metering analysis for the California IOUs. CPR provided E3 with half-hour solar irradiance data for every PV system in the state of California for their analysis.

Figure 2. SolarAnywhere public data access.



Several activities were associated with continued production of the data. First, there was the daily activity of ensuring that the system operated correctly; operational issues were addressed immediately when they occurred. Second, there was the monthly activity of ensuring that end-of-month processing operated correctly. Third, there was the activity of speeding up data processing. In particular, end-of-month processing required a two-day process at the beginning of this project. This caused significant server performance problems and made it slow for any users to access the data. End-of-month server performance issues were eliminated as part of this project.

2.3. SolarAnywhere High Resolution

SolarAnywhere Enhanced Resolution data has a 1-km spatial resolution and 30-minute temporal resolution. Utility applications, however, require higher temporal resolution data. This project launched the first application of SolarAnywhere High Resolution data. This data set has 1-km spatial resolution and 1-minute temporal resolution. This data needed to be available on demand to generate the PV time series data for use in the validation and PV fleet forecasting described in the subsequent sections.

Two issues were addressed as part of this subtask. The first issue was to produce the high resolution data. The second issue was to produce the data at a speed fast enough to make it available for PV fleet forecasting for several hundred thousand individual PV systems every 30 minutes.

The first issue was to produce the data. The data is created using the enhanced resolution data as input and then applying a specially-designed Cloud Motion Vector (CMV) approach. Two consecutive enhanced resolution images are compared and the CMV approach is used to project the movement of the clouds based on a comparison of the two images. This produces a series of new enhanced resolution images. The 1-minute temporal resolution data was obtained from these images. This approach has been under development by Dr. Richard Perez for several years. It was further refined as part of this project.

The second issue was to produce the data quickly. While it was challenging to produce the 1-km, 1-minute data for the entire state of California, it was an even greater challenge to produce the data fast enough to work for forecasting behind-the-meter PV production for the entire state. In particular, new data needed to be produced for every location in California every half hour in order to update the forecast. This was an issue that had not been addressed in any previous projects.

The challenge of rapidly producing such a large volume of data required CPR to reevaluate how the solar resource data was produced. CPR ultimately decided to move its entire code base from running on servers in a local datacenter to a massively parallel architecture using Internet “cloud” computing. This enabled CPR to match the need for computing resources to the available supply without continually purchasing (and thus maintaining) new servers.

CPR performed this transition over a multi-month period. This included designing, porting, testing, and running the system. By the end of the process, CPR was able to produce solar forecasts that could be used to simulate the 10-day forecast of the output from 170,000 PV systems with 30-minute observations in less than 30 minutes. The system has been operating well for almost a year.

2.4. SolarAnywhere Validation

The final subtask was to validate the accuracy of the SolarAnywhere data. This subtask was performed in conjunction with a CEC project, titled, “Demonstration and Validation of PV Output Variability Modeling,” Project number CEC 500-10-059. This CEC report is attached as Appendix 1.

2.4.1. *Definitions*

It is important to clearly define what is meant by accuracy before discussing solar resource data accuracy. Accuracy validation often means different things to different people. As such, it is useful to begin with a definition of how accuracy quantification can be performed.

Three fundamental questions need to be answered to provide a clear definition of how accuracy quantification is performed.

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

2.4.2. *Data Source*

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

2.4.3. *Time Attributes*

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

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

2.4.4. *Evaluation Metric*

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

MAE relative to available energy (rMAE) is a good method to measure relative dispersion error. This is the method used in the present analysis. The MAE relative to the average energy available is calculated by summing the absolute error for each time interval over the time period, and then dividing by the total available energy.

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

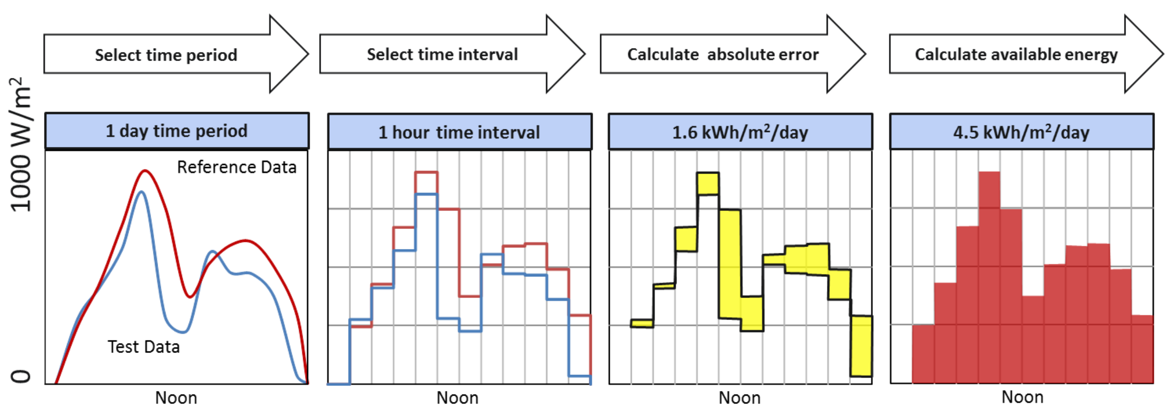
where I_t^{test} is the test irradiance at time t, I_t^{ref} is the reference irradiance at time t, and N is the number of time intervals.

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

As presented in Figure 3, the process is follows:

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

Figure 3. Mean absolute error relative to available energy calculation example.



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

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

2.4.5. Locations Selected for Validation

This metric can be used to quantify irradiance data accuracy for a one-year time period (2011) with time intervals ranging from one-minute to one-year using a historical time perspective. The analysis was performed for both individual locations and the ensemble of those locations.

Ten of the 46 metered locations were randomly selected for validation purposes. In order to perform the detailed analysis, each location had to have two global horizontal insolation (GHI) monitoring devices available on site and have one year's (2011) worth of data available. There were six locations that passed this initial screening.

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

The validation approach involved the following steps:

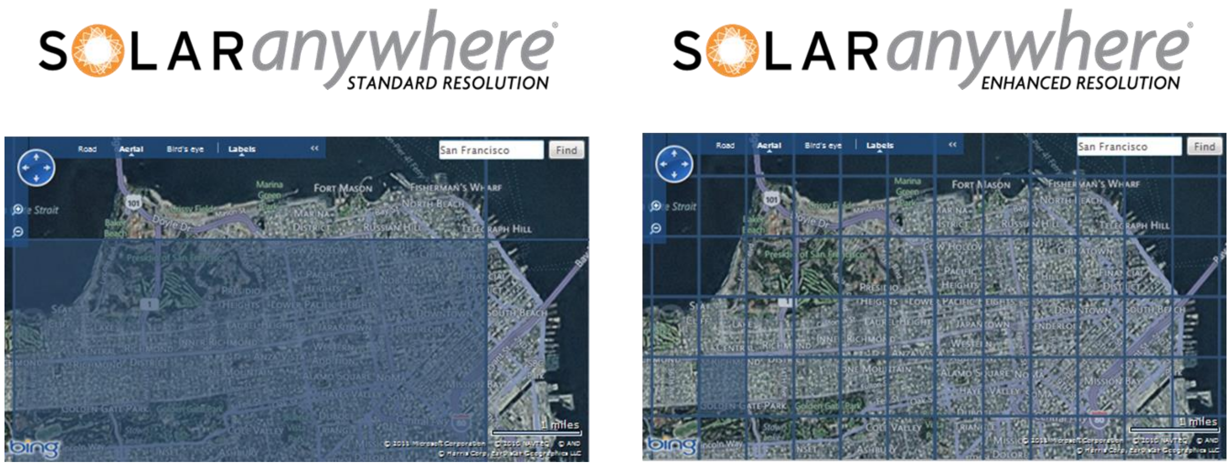
- Obtain time-series GHI data for 2011 for six locations:
    - 4-second data averaged into 1-minute time intervals from two separate sensors at each location.
    - Satellite based data at the following SolarAnywhere resolutions:
        - 1 minute, 1 km grid (High Resolution).
        - ½ hour, 1 km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

2.4.6. Obtain Time Series Data

CPR extended SolarAnywhere Standard Resolution (10 km spatial/1 hour temporal resolution) to SolarAnywhere Enhanced Resolution (1 km spatial/ 30 minute temporal resolution) under a previous contract.[2] Figure 4 illustrates the increase in resolution for San Francisco, CA.

[2] California Solar Initiative Solicitation #1 Grant Agreement, "Advanced Modeling and Verification for High Penetration PV".

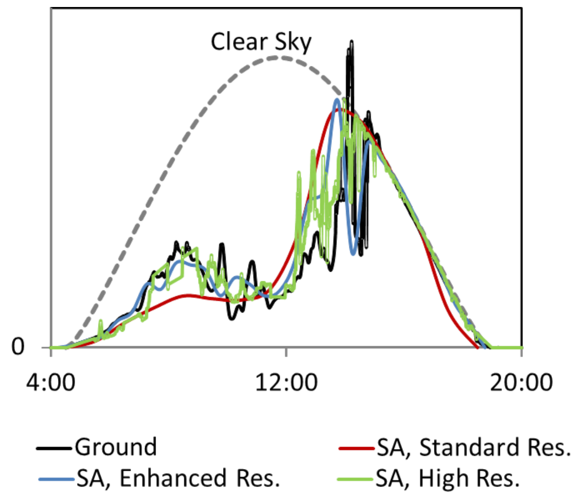
Figure 4. SolarAnywhere Standard and Enhanced Resolution



(San Francisco, CA)

A critical part of project was to extend SolarAnywhere Enhanced Resolution to SolarAnywhere High Resolution (1 km spatial/ 1 minute temporal resolution). The data was generated for all selected locations. Figure 5 presents a sample of the data for one day (July 4, 2011) at one location (CAISO Site A).

Figure 5. Time series data for all data sources on July 4, 2011 at CAISO Site A.



Note: only one ground source is shown for clarity purposes

2.4.7. Evaluate All Observations for Data Quality

As mentioned above, one of the steps in the analysis was to evaluate all observations for data quality. When evaluating accuracy, it is often simply assumed that reference data is correct. This assumption is made due to the difficulty in determining whether or not the reference data is correct: to what can the reference data be compared?

A unique aspect of the data provided by the CAISO is that all the locations have two ground sensors. As a result, since either sensor could be the reference, the data quality of the ground sensors was assessed by comparing the two ground data sets.

This was the process used to assess data quality:

(1) Compare the two sets of ground sensor data to each other to determine when one value is substantially different than the other value.
(2) Compare the enhanced resolution satellite and ground sensor data to search for 0 values occurring at incorrect times (e.g., mid-day on otherwise clear day) to determine when the satellite data is invalid.
(3) Compare ground sensor data to the SolarAnywhere Enhanced Res. data to determine if both ground observations are the same but are obviously incorrect (e.g., the irradiance value remains at a constant level for many hours).

The complete data set was evaluated and then potential outliers were manually evaluated and screened for each of these steps. Figure 6 illustrates the screening result when comparing the two ground sensors at one location. All of the data points would lie on the 45 degree red line if they were identical. The blue symbols correspond to valid data and the black symbols correspond to invalid data. Figure 7 illustrates the issue for one of the invalid observations when one of the sensor's recorded values remained constant after solar noon. Figure 8 illustrates the case when both ground sensors produced a similar value but were obviously incorrect, reading a constant low value on an otherwise clear day as assessed from the satellite data. Figure 9 illustrates the case when there was a night-time calibration error across the year. Site E was missing more than a month of data during the first part of the year as well as a five percent difference between the two ground sensors.

Sites E and F were eliminated from the analysis as a result of the data filtering process. The remaining sites had about one percent of the ground data marked as invalid.

Figure 6. Half-hour energy production in 2011 from meter 2 vs. meter 1 (Site A).

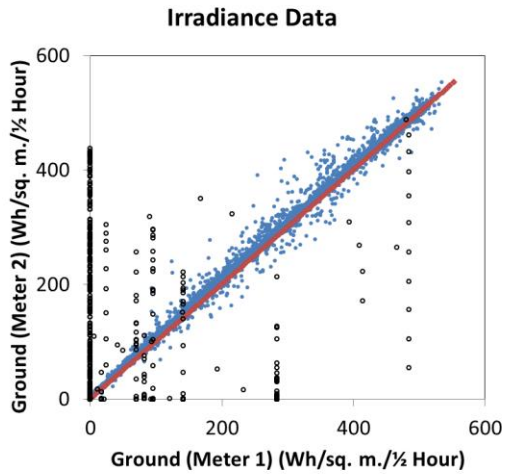
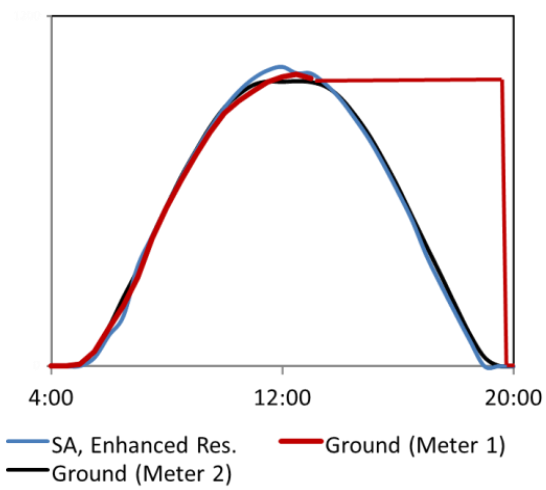
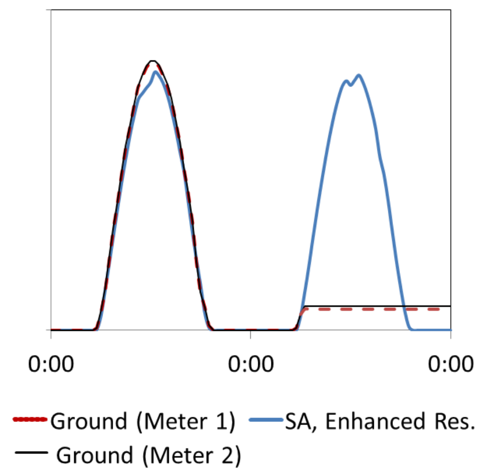


Figure 7. Example of when only one of the ground sensors has invalid data.



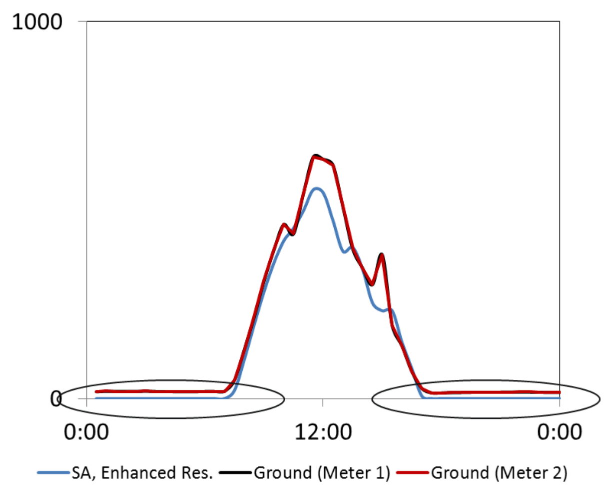
(Site A, June 22, 2011)

Figure 8. Example of when both ground sensors have invalid data.



(Site C, May 1-2, 2011)

Figure 9. Site F has a night-time calibration error across the year.



2.4.8. Results

rMAE was calculated for three scenarios:

- Each location individually.
- Average of individual locations.
- Fleet of locations.

Figure 10 presents the rMAE for each of the four locations using time intervals ranging from 1 minute to 1 year.

Figure 10. Relative MAE for each location individually.

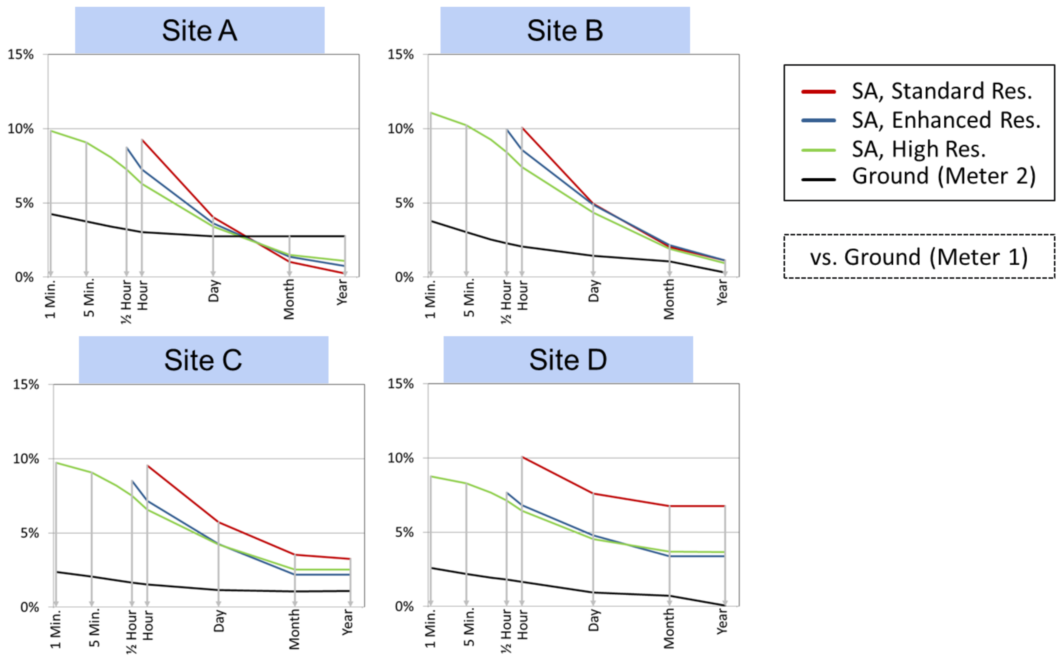
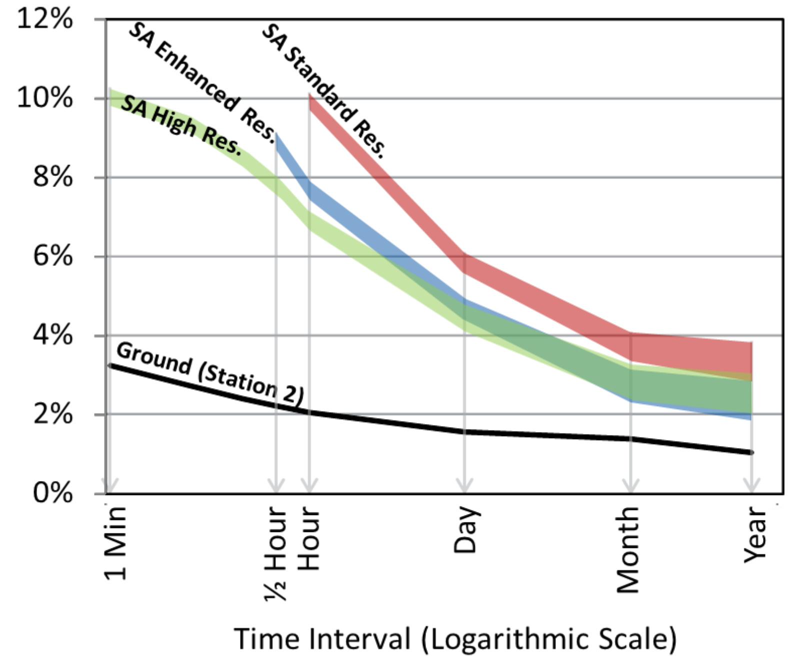


Figure 11 presents the average rMAE of four individual locations. The black line summarizes the error when two ground stations were used (one was the reference and the other was the test). The green, blue, and red regions summarize the error when SolarAnywhere High, Standard, and Enhanced Resolution were compared to the ground sensor. The green, blue, and red areas are regions rather than lines because they compare satellite data to ground data using the two different ground sensors: the top of the region is the comparison using the ground sensor that maximizes error; the bottom of the region is the comparison using the ground sensor that minimizes error.

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

Figure 11. Average MAE of 4 individual locations.

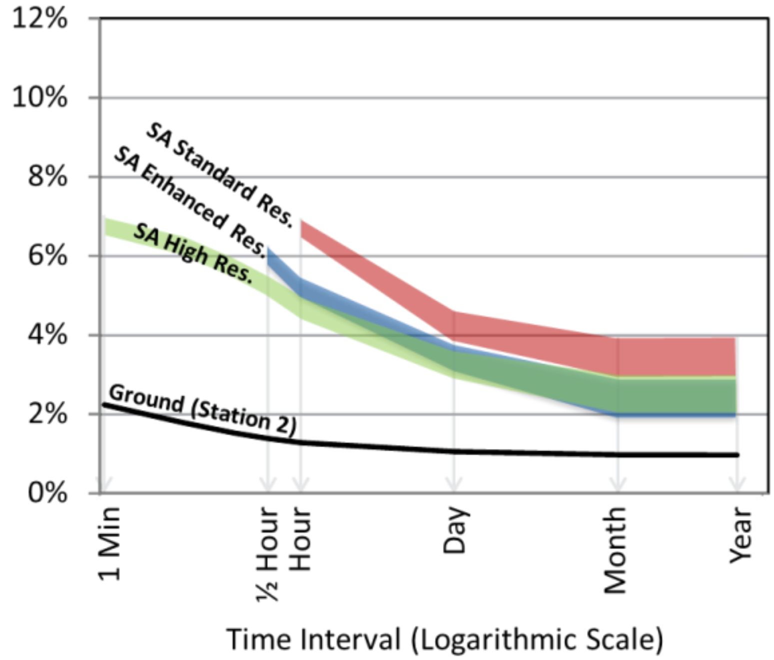


A consistent finding of many PV variability studies is that variability is reduced when PV systems are geographically dispersed. That is, variability is reduced as the number of systems increases across a sufficiently large geographic region.

So far, this section has focused on the error associated with individual locations. While individual locations are of interest in some cases, there are certainly many other cases in the utility industry when users are most interested in the error associated with a set of locations.

The rMAE analysis was repeated with the input data being the combined irradiance across four locations. The results are presented in Figure 12. A clear reduction in error due to combining locations can be seen by comparing Figure 12 to Figure 11. That is, the effect of geographic dispersion on reducing output variability reduction that has been observed by others is now also observed with regard to prediction accuracy: accuracy improves as a geographically diverse set of independent locations are combined.

Figure 12. MAE of 4 locations combined.



2.5. Conclusions

This task provided SolarAnywhere data and benchmarked data accuracy. The SolarAnywhere data included both Enhanced Resolution and High Resolution data. Enhanced Resolution is 1 km spatial resolution and half-hour temporal resolution while High Resolution is 1 km spatial resolution and 1 minute temporal resolution. Both versions were successfully developed and delivered over the term of the grant. Benchmarking results indicated that High Resolution data had about 7 percent rMAE on an hourly basis for a single location. Similar results were obtained for data provided for SMUD's extensive solar resource monitoring network.[3]

[3] "Solar Monitoring, Forecasting, and Variability Assessment at SMUD." Presented at WREF 2012 (SOLAR 2012). Denver, CO, May 2012. Paper available at. http://www.cleanpower.com/wp-content/uploads/SMUD-Solar-Assessment_2012.pdf.

3. Task 2: Validate PV Fleet Simulation

3.1. Introduction

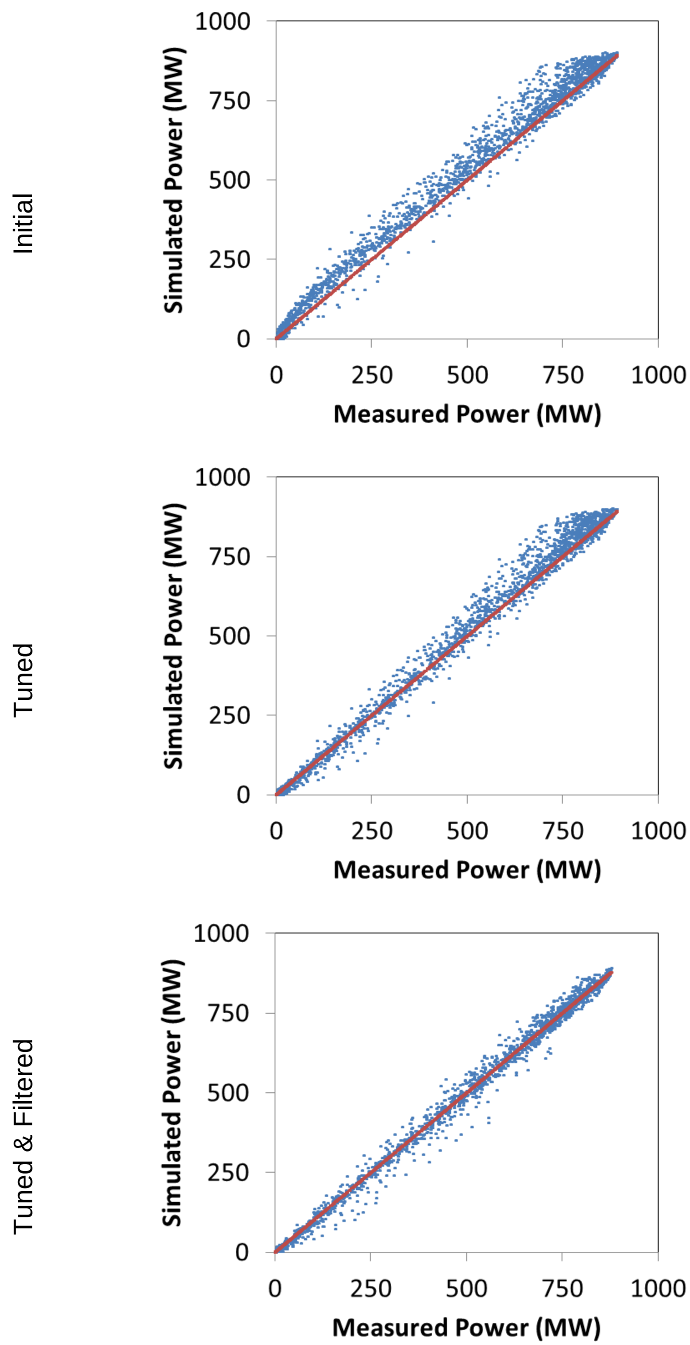
The second task was to validate PV fleet simulation methodologies using measured ground data from fleets of PV systems connected to California's grid. This task was accomplished using measured PV production data from two separate sources.

One set of data was provided by CAISO for large PV plants connected to its system. The other set of data was provided by SMUD for small distributed PV systems connected to its system. Analysis of the CAISO data was performed in conjunction with a CEC project, titled, "Demonstration and Validation of PV Output Variability Modeling," Project number CEC 500-10-059. This report is attached as Appendix 1. Analysis of the SMUD data was documented in a separate report. It is attached as Appendix 2. This section highlights results from those two reports. Additional details are presented in the appendices.

3.2. Results for Large PV Systems Connected to CAISO

CAISO provide fifteen-minute measured production data for 46 metered PV plants from March 10, 2013 to April 19, 2013. The measured data were used to infer system specifications including rating, azimuth angle, and tilt. These system specifications were then used to simulate production. Figure 13 presents simulated vs. measured 15-minute average PV fleet power for each interval. All of the blue markers would be on the red line if simulated and measured results matched perfectly. The top of the figure corresponds to the "Initial" case of PV fleet production without PV performance filtering for plant performance problems (it corresponds to Figure 20 in the CEC report). A consistent power-related bias can be observed. This bias can be reduced by applying an inverter model tuning curve. The "Tuned" case is presented in the center of Figure 13. Significant scatter, however, can still be observed. This can be reduced by filtering the data for PV performance using the filtering from the previous section. The "Tuned & Filtered" case is presented in the bottom of Figure 13. There is a good alignment between simulated and measured data after making the tuning and filtering adjustments. Results show that the Initial, Tuned, and Tuned & Filtered cases have 7.2, 5.2, and 3.1 percent rMAE, respectively.

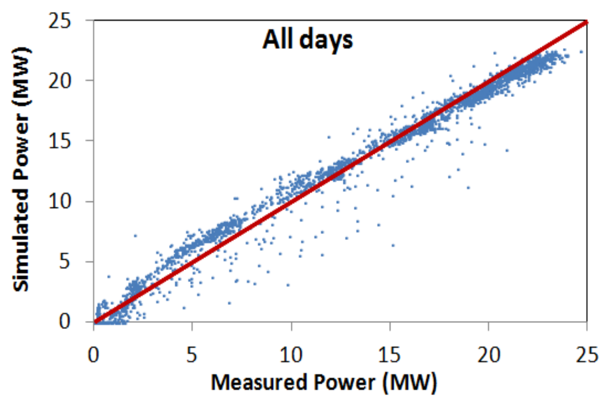
Figure 13. Simulated vs. measured average 15-minute power for CAISO metered PV fleet.



3.3. Results for Small Distributed PV Connected to SMUD

SMUD provided measured hourly production data for 2,206 distributed PV systems from April 16, 2012 to October 10, 2012. SMUD also provided specifications for all of the PV systems. PV production was then simulated using the system specification data combined with SolarAnywhere data. Figure 14 presents simulated versus measured hourly energy production for the fleet of 2,206 distributed PV systems from April 16, 2012 to October 10, 2012. Results indicate an accuracy of 6.2 percent Mean Absolute Error relative to energy (rMAE). These results demonstrate that accurate simulations of a large fleet of PV systems are obtainable. As was the case at CAISO, it is also likely that additional accuracy will be realized by improving the PV inverter model (e.g., see the CAISO results after tuning presented above). Note that this tuning is not applied for SMUD.

Figure 14. Simulated vs. measured hourly energy production for 2,206 distributed PV systems from April 16, 2012 to October 10, 2012.



3.4. Conclusions

Understanding the accuracy at which one can simulate fleet wide PV system energy production is a critical step towards facilitating increased PV penetration into California's electricity system. Factors such as irradiance, shading, soiling, and system configuration greatly influence the performance of an installed PV system. Proper characterization of these factors is important to the simulation of PV system energy.

Several conclusions can be drawn from these results. First, SolarAnywhere's PV fleet simulation capabilities result in fairly accurate results, especially for fleets of PV systems. Results for the large metered PV plants connected to CAISO suggest that total rMAE was 7 percent for 15-minute time interval data. This result can be reduced to about 3 percent by tuning the model and incorporating plant operating status in the simulation. Results for the small distributed systems at SMUD demonstrated an accuracy of 6 percent rMAE when all systems and all days were included. The error was reduced to 5 percent rMAE for a subset of well-behaved PV systems. Results further improved to 4 percent, when partly cloudy day conditions were removed

Second, there is room for improvement in the underlying PV simulation methodologies by further inspection of simulated and measured data at the hourly and sub-hourly levels. Additional work can be done to develop a more accurate inverter model and to understand better application of PV modeling derate factors. Better data tuning and measured data cleaning methods would help identify and rectify faulty PV system specifications and help improve simulations.

4. Task 3: Integrate PV Fleet Simulation Into Utility Software Tools

4.1. Introduction

Task 1 was to create a high resolution (1-km, 1-minute resolution) solar resource database. Task 2 was to validate PV fleet simulation methodologies. Task 3 was to use the results from Tasks 1 and 2 to integrate PV fleet simulation methodologies into utility software tools.

A fleet of PV systems can be defined very broadly. At one extreme, a fleet can refer to single PV system on the roof of one person's house. That is, it is a fleet of one. At the other extreme, a fleet can refer to all PV systems located within the balancing area of a major authority such as the California ISO (CAISO). A fleet could even be defined more broadly and refer to all PV systems in the U.S. or even the world. User-defined collections of systems ("virtual" fleets) based on location, system attributes or other criteria are useful for planning and modeling purposes.

The PV fleet simulation can be based on historical, real-time, or forecasted solar resource data. Historical data is useful for system planning. Real-time data is useful for assessing PV fleet operation. Forecasted data is useful for determining how to operate the rest of the utility system.

4.2. Original Plan

This broad definition of a PV fleet makes the simulation capabilities useful across a wide spectrum of utility applications. These applications range from planning and smart grid operation in the distribution system, utility load scheduling in the utility system, and balancing area planning and operation functions covering multiple utilities.

The original project plan was to demonstrate PV fleet simulation capabilities across the range of applications. The approach was to have short-duration, limited scope demonstrations of how fleet simulation could be used. The simplest planned application was to simulate the historical production of a small PV fleet on a single distribution feeder. The most complex planned application was to forecast the output of all PV systems in California and supply this information to the CAISO.

4.3. Revised Plan

Simulating the historical PV fleet production for CAISO was one of the first applications that CPR began to work on. CAISO has the responsibility of maintaining reliability and accessibility for California's utility grid. As such, they are concerned with the effect of power production from customer-owned PV systems on the balancing area.

CPR's interaction with CAISO revealed that, while CAISO clearly understood the performance of the approximately 50 large PV systems through detailed production monitoring, they had no visibility into the performance of the hundreds of thousands behind-the-meter PV systems. This was concerning to them. They needed to begin to develop a forecast for the fleet of behind-the-meter PV systems. That is, the greatest market need was the most complex application that CPR had originally anticipated. Furthermore, this was needed on a full-scale, not a limited scope.

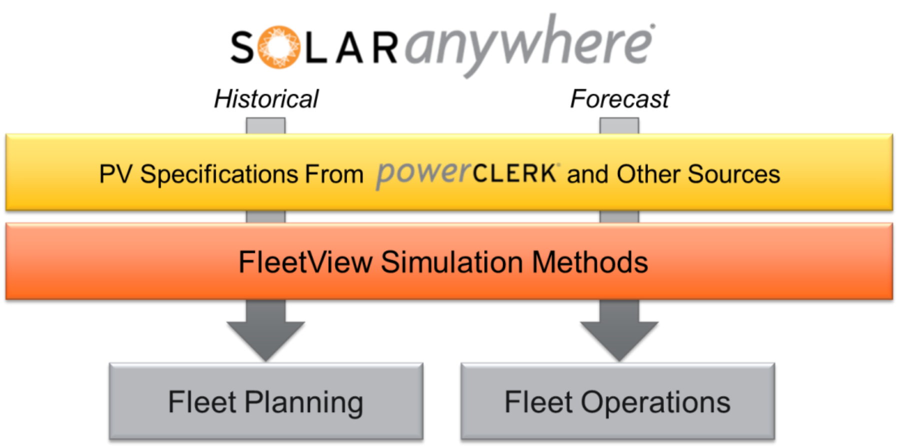
It became clear that even beginning to address CAISO's issue would require significant resources. It was determined that, rather than demonstrating PV fleet simulation for a wide range of applications with a limited scope, it would be more beneficial to begin the implementation of PV fleet forecasting for the most complex application: balancing area-wide, behind-the-meter PV fleet forecasting. PV fleet forecasting could then be applied to all other applications if it could be demonstrated for the CAISO balancing area.

CPR embarked on the task of designing, testing, implementing, and operating the PV fleet forecasting system for the entire state of California.

4.4. PV Fleet Simulation

Figure 15 illustrates that there are three critical elements in performing a PV fleet simulation: solar resource data (listed as SolarAnywhere in the figure), PV system specifications, and a simulation model to convert this information into production.

Figure 15. PV fleet simulation procedure.



4.5. Solar Resource Data

The first component required is the solar resource data. This data set was discussed above. Task 1 produced a SolarAnywhere High Resolution data for the entire state of California. This 1-minute data set is produced every half hour and covers the subsequent 60 minutes at which point the time interval shifts to 30 or 60 minutes, depending upon time frame. As discussed above, this step required significant effort to move CPR's processes to the Internet "cloud" in order to generate the data in an acceptable timeframe.

4.6. PV System Specifications

The second component required is the list of specifications for every PV system. At a minimum, this includes PV system rating, azimuth, tilt orientation, and fixed vs. tracking mode. This task was completed in conjunction with a CEC project. Together, these two projects obtained specifications for all PV systems in California that had received an incentive to get installed. Many of the PV systems for California are included in CPR's PowerClerk® database. As of the writing of this report, there are a total of more than 170,000 PV systems, with the number of systems continuing to grow each month.

4.7. PV Simulation Model

The third component required is a model that combines solar resource data with PV system specifications to simulate PV production. The simulation model needs to be capable of performing the simulation for each PV system individually and then combining the results for all systems into the PV fleet output. It needs to be able to perform the calculations rapidly, ideally in a parallel (concurrent) fashion.

4.8. Challenges

Three challenges were encountered in accomplishing this task. The first challenge was to obtain the data. This included both the solar resource data and PV system specification data for all systems in California. The second challenge was to perform the simulation quickly. The third challenge was to keep the data updated.

4.9. Obtain Data

The first challenge was to obtain the data. This included both the solar resource data and PV system specification data for all systems in California.

The production of the solar resource data was described above. The result was that data was available in a 1-km grid for the entire state of California.

The initial scope of the project was to simulate behind-the-meter PV production for a sample of PV systems in California. The scope was to simulate historical half-hour PV fleet output from 2008 to 2011.

It became clear early on in the project that CAISO needed visibility into all PV systems in California. Thus, the decision was made to provide comprehensive simulation for all behind-the-meter PV systems in California.

4.10. Behind-the-meter PV System Specifications

The task of collecting the specifications for all PV systems was performed in conjunction with a CEC-sponsored project. The details of the data collection are described in the Appendix. As described in the Appendix, covered programs included:

- Renewable Portfolio Standard systems
- Publicly owned utility Senate Bill-1 programs
- CEC's Emerging Renewables Program
- New Solar Homes Partnership
- Single Family Affordable Solar Homes
- Self-Generation Incentive Program
- California Solar Initiative

In addition to collecting data for all behind-the-meter systems, CPR needed to obtain specifications for PV systems connected to CAISO. CAISO provided high speed historical data for all of these systems. The specifications, however, were not available. CPR developed an approach to infer specifications from PV system performance data.[4]

4.11. Perform Simulation Quickly

The second challenge was to perform the simulation quickly. CPR originally planned to provide CAISO with simulated PV fleet production to use for planning purposes. As mentioned above, the decision was made to focus efforts on providing forecasts for all behind-the-meter PV system to CAISO. As a result, CPR needed to simulate production for hundreds of thousands of individual PV systems that could then be combined and delivered as fleets to CAISO. This needed to be done every half-hour.

CPR's forecasting system was not designed to accommodate high volume simulation requirements with short delivery schedules. As a result, this requirement presented CPR with a significant challenge. CPR's system was re-architected and redeployed to an Internet "cloud"-based platform. This allowed CPR to satisfy CAISO's technical requirements of high volume simulations delivered in short time-frames.

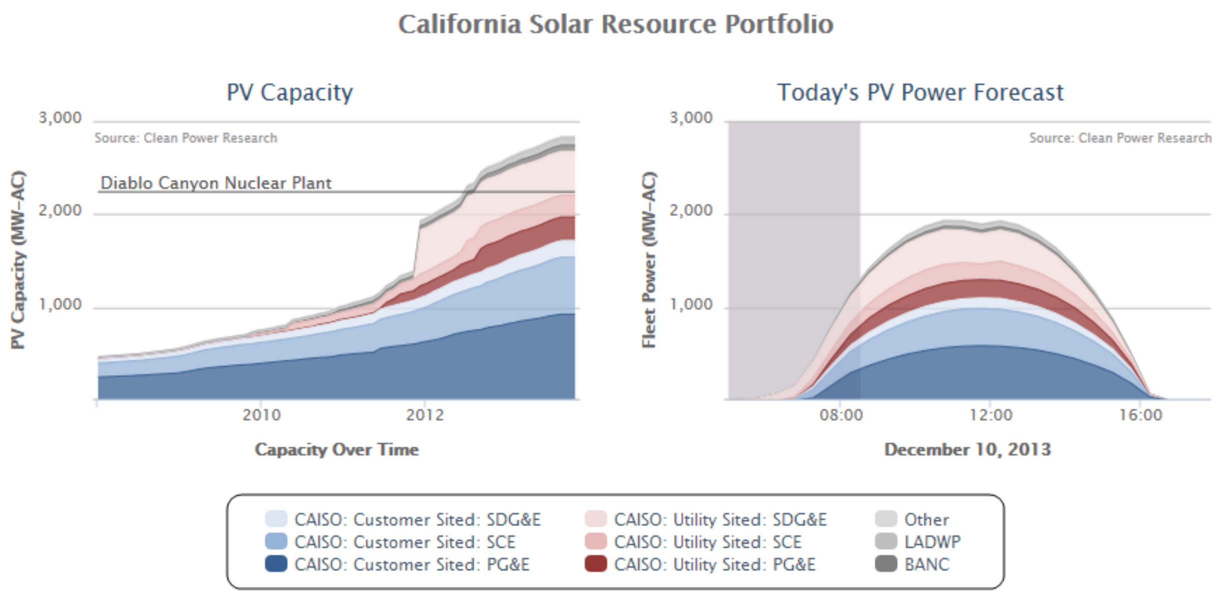
4.12. Keep Data Up-to-Date

The third challenge was to keep the data updated. The number of PV systems is growing rapidly in California. There were 70,000 PV systems in the state as of the end of 2011 (near the start of this project). There were more than 170,000 PV systems near the end of 2013 (as of the writing of this report). This means that the number of systems grew at a rate of faster than 50 percent per year. The number of systems has more than doubled in two years. There will be more than a million behind-the-meter PV systems before 2020 if the market continues to grow at this rate. The capacity of these systems and the associated PV fleet forecast is presented in Figure 7. The data is presented according to the five regions defined by CAISO with the blue regions corresponding to behind-the-meter data and the red to the directly connected systems.

To date, the maintenance of this information has been sustainable because most PV systems have received an incentive. More specifically, the PV system specification information has been collected in the process of receiving an incentive. This will change in places where incentives no longer drive the behind-the-meter market. This is one of the reasons CPR is extending PowerClerk to manage interconnection processes; in an incentive-less world, the system will provide a reliable, user- and utility-friendly online system to collect accurate, thorough system specification data in a format relevant not only to interconnection optimization but also for downstream purposes including planning and operations.

[4] Patent approved for "Computer-implemented system and method for inferring operational specifications of a photovoltaic power generation system."

Figure 16. California solar resource portfolio (Dec. 10, 2013).



4.13. PV Fleet Simulation Validation (CAISO Data)

PV fleet simulation results were validated using two sets of measured PV production data. The first set of data was provided by CAISO. The second set of data was provided by SMUD. Consider, first, results based on data provided by CAISO.

The CAISO measures power production every four seconds for 46 PV plants. A 15-minute time interval is critical to the CAISO's forecasting efforts above. Thus, the four-second measured PV power production was averaged to 15-minute data. CPR interacted with CAISO to determine data availability, resolve time synchronization issues, and take steps necessary to ensure data integrity.

4.13.1. Sources of Error

Inaccuracies degrade the ability of the simulation to reflect measured performance. These inaccuracies can be grouped into three categories.

1. Solar resource.
2. PV modeling.
3. PV plant performance issues.

Solar resource inaccuracies include errors in historical or forecasted solar resource data. PV modeling inaccuracies refer to limitations in the PV fleet modeling algorithms. PV plant performance issues reflect errors that occur because the plant is not operating as expected.

The effects of solar resource and PV modeling inaccuracies are fairly obvious. Inaccurate solar resource data (historical or forecasted) and/or PV fleet modeling algorithms clearly limit the simulation's ability to reflect measured performance.

PV plant performance issues are more subtle. Differences between simulated and measured PV production can still occur even if the simulation method perfectly predicts measured PV fleet power production for a fleet that is operating perfectly. Differences can occur if the actual PV fleet does not operate as expected due to system performance issues. That is, inaccuracies can occur that are unrelated to the fundamental simulation methodology. They are related to lack of incorporation of poor performance into the simulation.

4.13.2. PV Plant Performance Issues

The first step of the evaluation, therefore, is to determine how to address PV performance issues. One option is to incorporate plant status into the simulation methodology. The simulation, for example, would reflect a capacity reduction if a plant was only operating at 50 percent capacity. This option requires obtaining PV plant status information. This information, unfortunately, was unavailable for the CAISO fleet of PV systems.

An alternative approach is to identify days when the individual plants had sub-par performance. These days and plants are then eliminated from the fleet simulation. This is the approach that was taken for this project.

Fifteen-minute measured and simulated data were obtained for 46 CAISO metered PV plants from March 10, 2013 to April 19, 2013. The time series data were compared for each of the plants individually. The data was visually examined to assess days when the PV plant was either not operating or was clearly underperforming. Figure 17 and Figure 18 present the results of the analysis for two of the 46 plants. The red and blue lines correspond to simulated and measured data. The shaded areas represent days with plant performance issues. The dashed line corresponds to the daily rMAE. Figure 17 corresponds to a plant that operated well during the whole time period. Figure 18 corresponds to a plant that had significant operational issues.

Figure 17. Example of PV Plant that operated as expected.

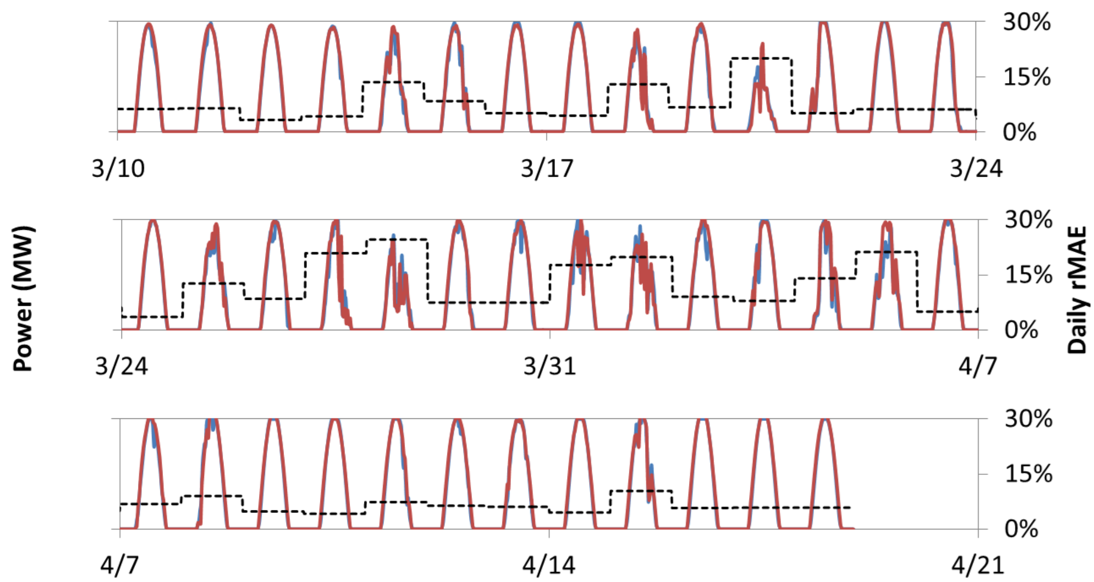
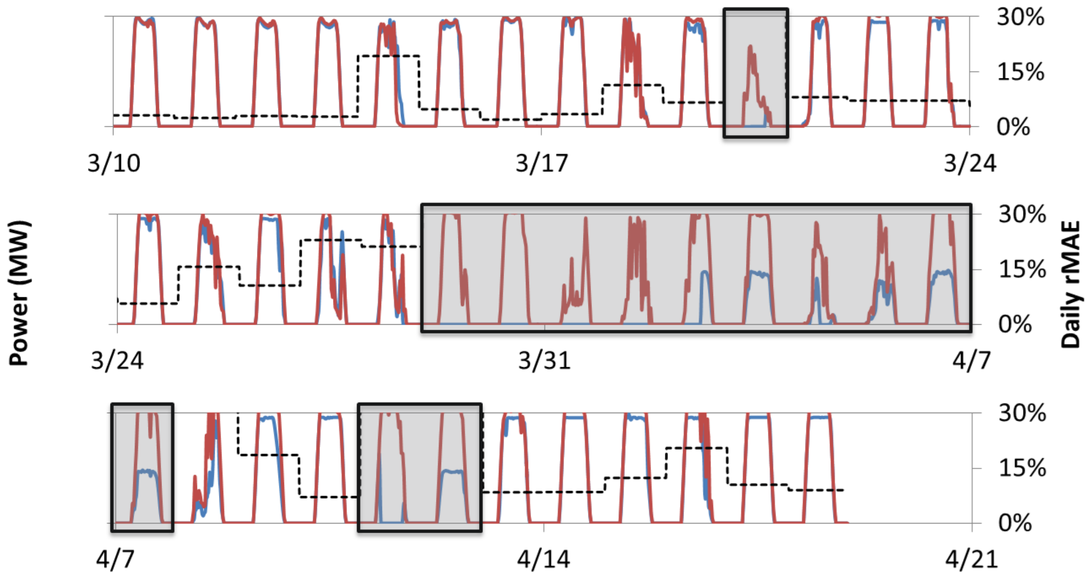
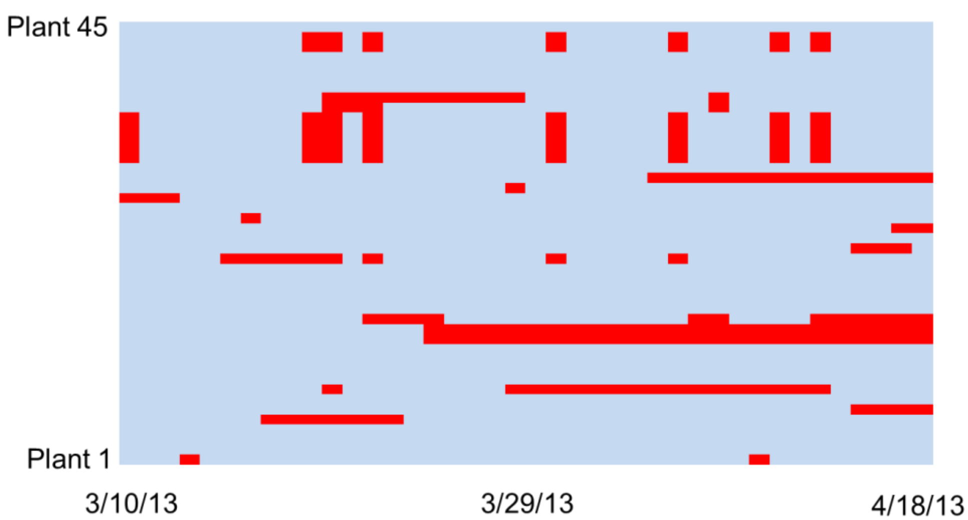


Figure 18. Example of PV Plant with possible performance issues.



This process was repeated for all of the plants. Figure 19 summarizes plant performance for all 46 plants. The y-axis corresponds to the plant number and the x-axis corresponds to the date. Blue corresponds to normal operation and red corresponds to performance issues. The figure suggests that the PV fleet experienced a significant number of performance issues over the six-week analysis period.

Figure 19. Summary of performance issues for all metered plants.



4.13.3. PV Fleet Simulations: Time Series Data

Simulations were performed using FleetView with and without plant filtering results from the previous section. Figure 20 presents PV fleet output without filtering. Figure 21 presents PV fleet output with filtering. A comparison of the two figures illustrates the improvement in accuracy by taking PV plant performance issues into consideration.

Figure 20. PV fleet production before PV performance filtering.

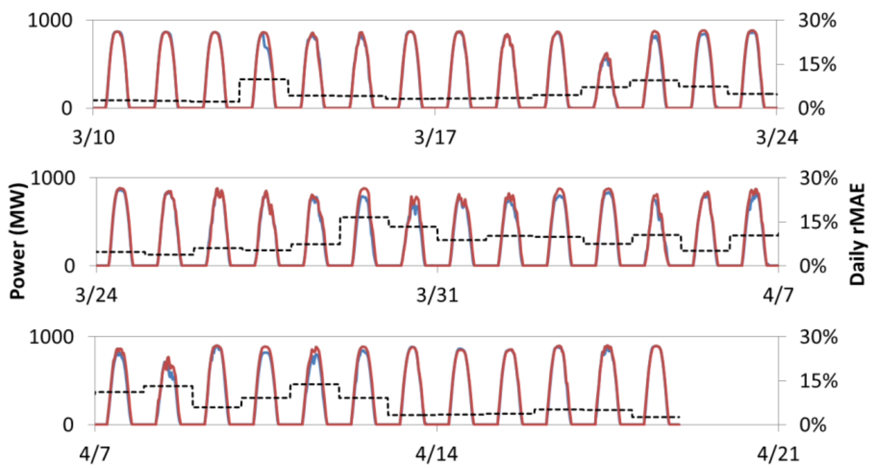
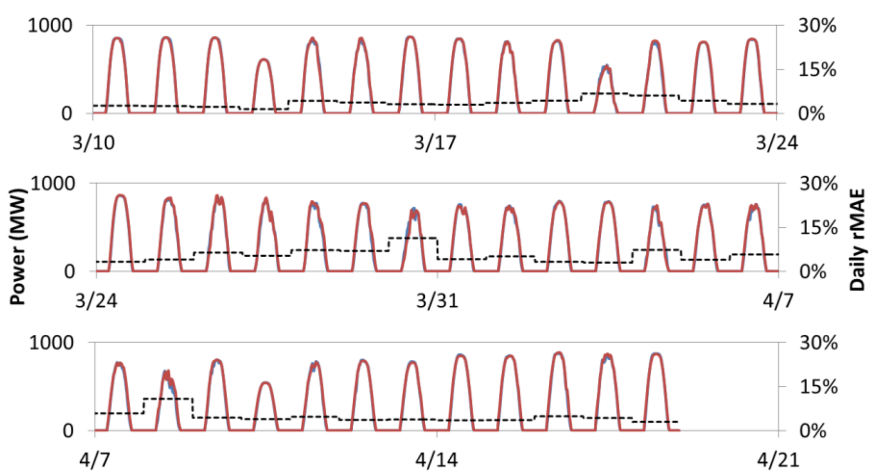


Figure 21. PV fleet production after PV performance filtering.



4.13.4. PV Fleet Simulations: Simulated vs. Measured Data

An alternative way to present the data in Figure 20 is to plot simulated vs. measured average power for each 15-minute interval. Figure 22 presents the data in this manner. All of the blue markers would be on the red line if simulated and measured results matched perfectly. The top of the figure corresponds to the "Initial" case of PV fleet output without PV performance filtering (it corresponds to Figure 20). A consistent power-related bias can be observed.

This bias can be reduced by applying the tuning curve presented in Figure 23. The "Tuned" case is presented in the center of Figure 22. Significant scatter, however, can still be observed. This can be reduced by filtering the data for PV performance using the filtering from the previous section.

The "Tuned & Filtered" case is presented in the bottom of Figure 22. There is a good alignment between simulated and measured data after making the tuning and filtering adjustments.

Figure 22. Simulated vs. measured average 15-Minute power for CAISO metered PV fleet.

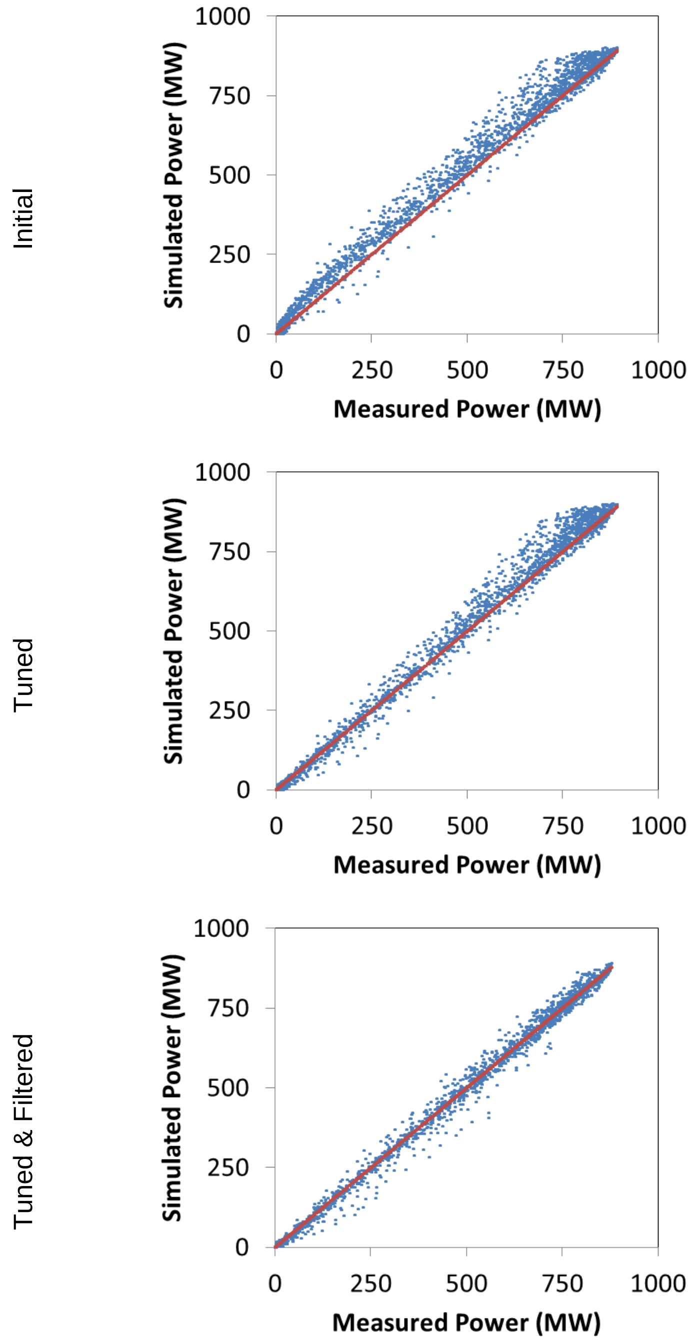
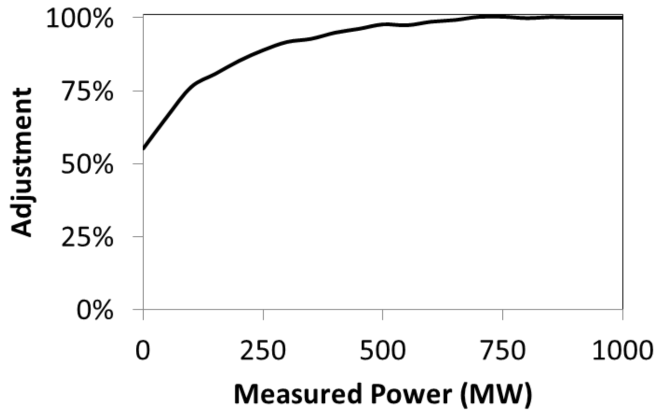


Figure 23. Power-based simulation tuning.



4.13.5. Relative Mean Absolute Error

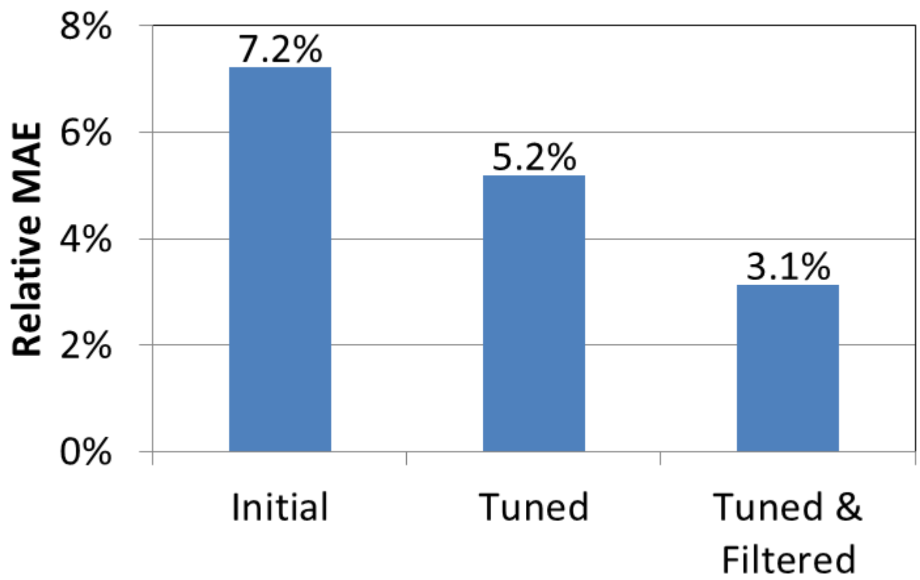
The final step of the analysis is to calculate the rMAE. The time series data were evaluated over the approximately six-week time period for the 15-minute time interval data. Figure 24 presents results for three cases: Initial, Tuned, and Tuned & Filtered. These cases correspond to the results presented in Figure 22. Results show that the Initial, Tuned and, Tuned & Filtered cases have 7.2, 5.2, and 3.1 percent rMAE.

Several observations can be made based on these results. First, overall, FleetView PV power modeling is pretty accurate. There is, however, room for improvement. In particular, improving the inverter power curve model for individual PV systems will substantially improve simulation results (i.e., the improvement identified by applying the tuning).

Second, there is a substantially negative effect due to poorly performing plants even after the PV fleet model has been tuned. Accurately representing plant status reduces error by more than 40 percent.

Third, three percent rMAE can be achieved for 15-minute time interval data using a well-tuned model that accounts for poor PV plant performance. This requires that: (1) accurate location-specific solar resource data is supplied; (2) correct PV specifications are used; (3) the inverter power curve is properly represented (i.e., the simulation is tuned); and (4) actual PV plant status is incorporated into the simulation.

Figure 24. Total rMAE.



It is useful investigate the error on a daily basis in addition to an analysis over the entire time period. Figure 25 and Figure 26 present the daily rMAE for the 15-minute time interval before and after tuning the model. The blue and red colors correspond to simulation error and PV plant performance error respectively. PV plant performance error is estimated by subtracting simulation error with and without filtering. The figure shows that rMAE varies from day to day. While absolute error increases on some of days, rMAE tends to be higher on low energy days. This is because the rMAE calculation is defined as absolute error divided by measured energy.

Figure 25. Daily relative MAE using 15-minute time interval before tuning.

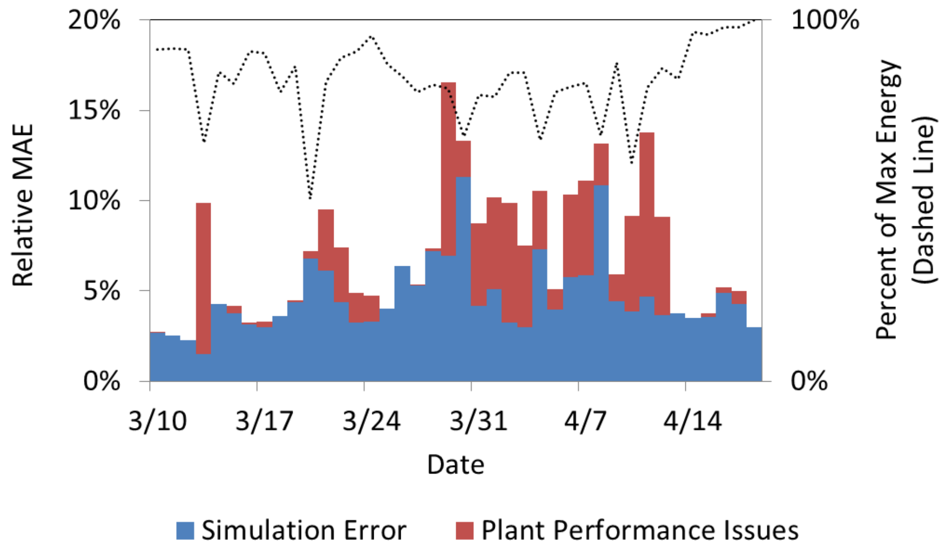
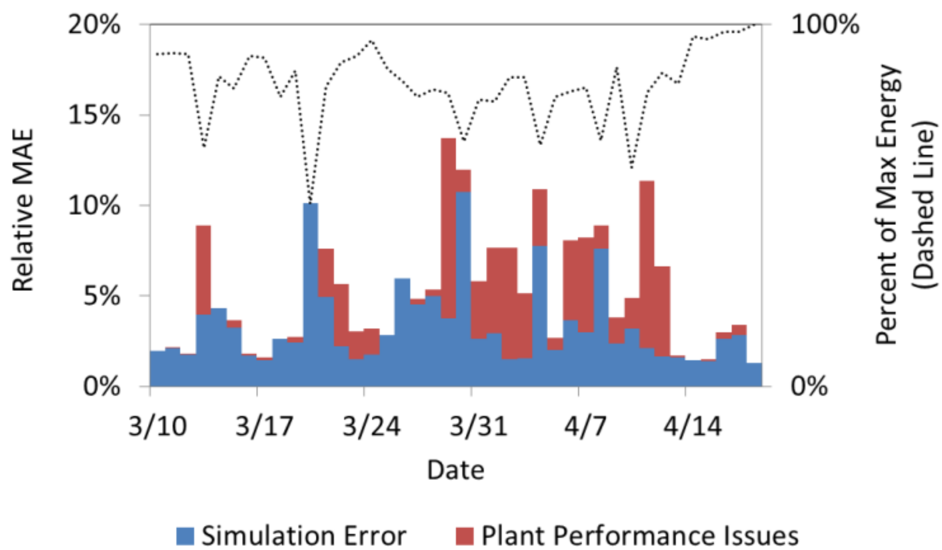


Figure 26. Daily relative MAE using 15-minute time interval after tuning.



4.13.6. Sample Days After Tuning and Filtering

It is useful to compare simulated and measured data for a range of days after tuning and filtering. Figure 27, Figure 28, and Figure 29 present measured and simulated PV fleet production. Figure 27 corresponds to a clear day. Figure 28 corresponds to a day with PV performance issues. Figure 29 corresponds to a day with variable weather and PV performance issues.

Several observations can be made. First, tuning the simulation model increases accuracy for all days. Second, modeling on a clear day is very good with a rMAE of less than 2 percent. Third, filtering for PV plant performance issues can be very important; rMAE was reduced from 20 percent to 4 percent on one particular day. Fourth, simulated data tracks measured data fairly well even for the worst performing day.

Figure 27. PV fleet production on clear day.

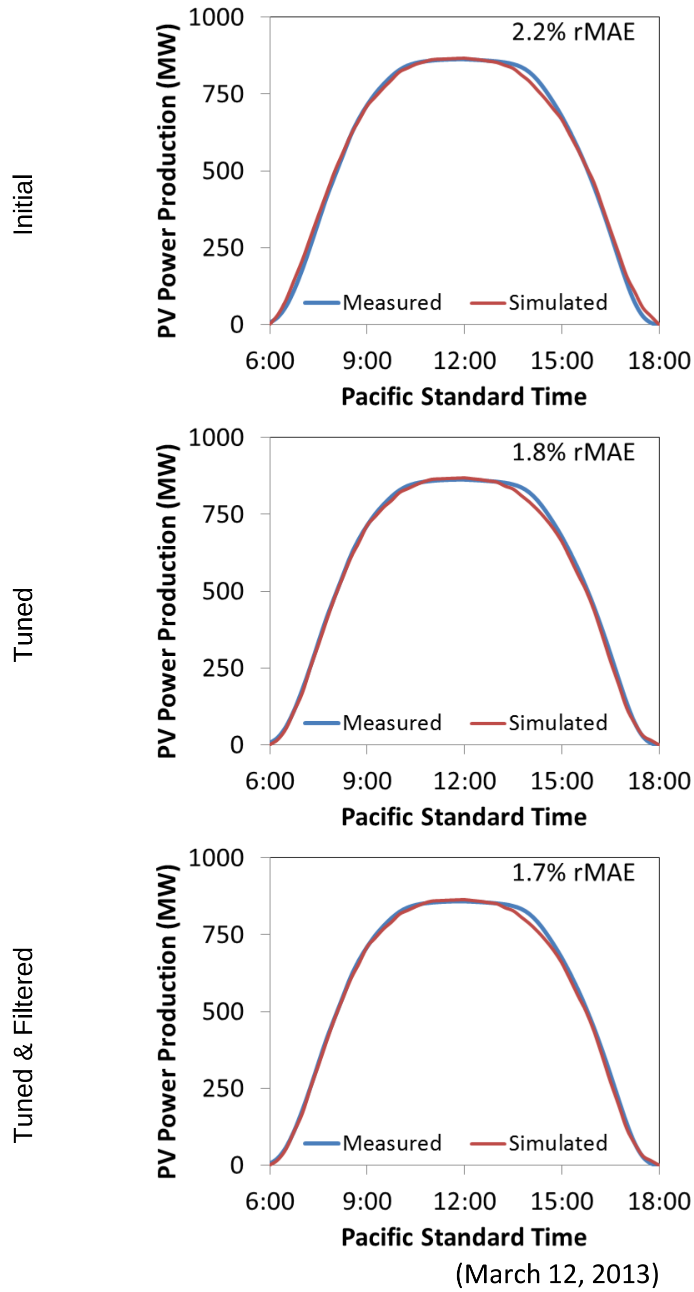


Figure 28. PV fleet production on clear day with production issues.

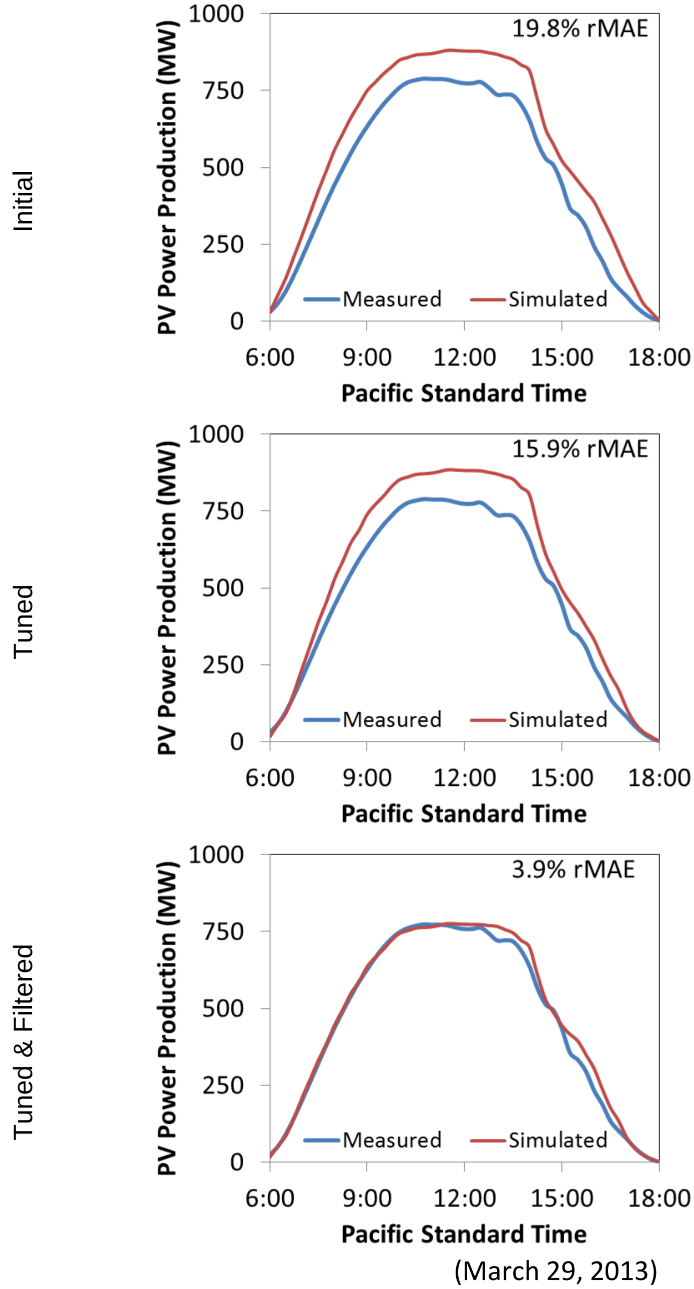
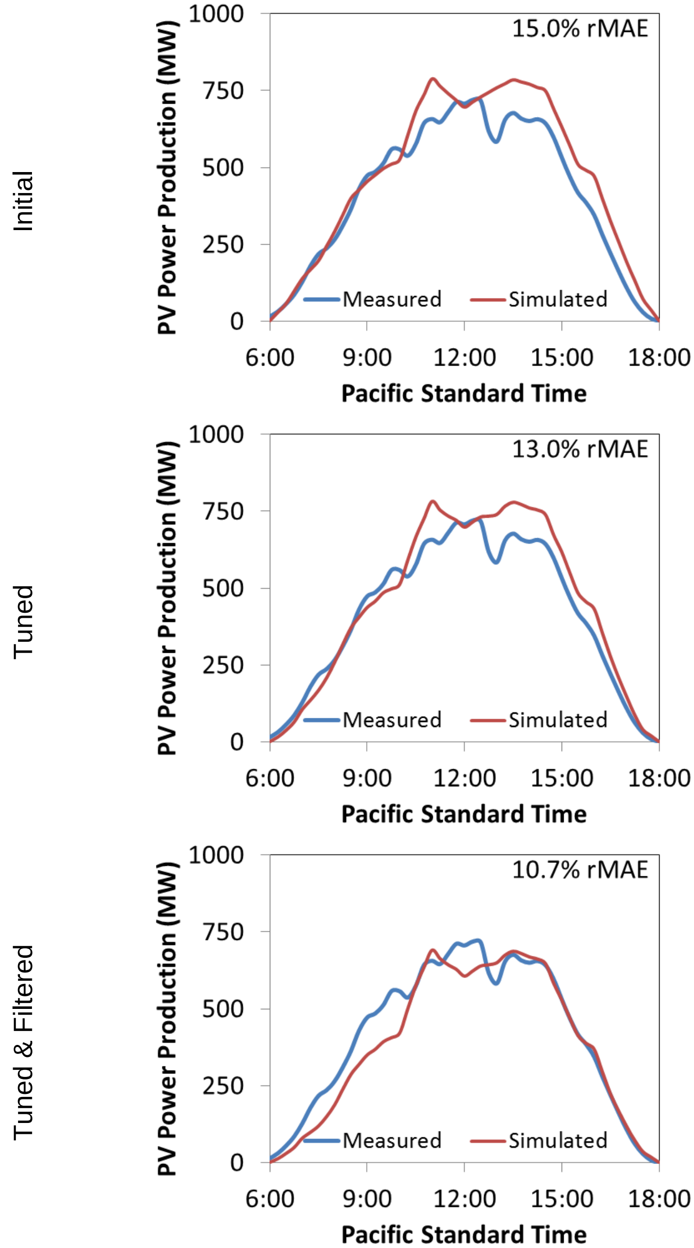


Figure 29. PV fleet production on variable weather day with production issues.



4.14. PV Fleet Simulation Validation (SMUD Data)

Consider, next, results based on data provided by SMUD.

4.14.1. Data Set Correlation

SMUD provided historical PV data for 2,550 distinct PV systems. The data contained a timestamp, measured energy production, duration of the measurement (time increments from 5 minutes up to hourly), and the system's Distributed Generation number (DG number).

PowerClerk® is used as the primary record for all PV systems in SMUD's service territory. PowerClerk contains detailed system specifications, including inverter type and quantity, PV module type and quantity, array tilt, azimuth orientation, and shading. PowerClerk identifies each system by its DG number.

The measured production data set and system specifications data set were linked using the DG number. The systems were assumed to be the same if the DG numbers matched. Random spot checks confirmed that this was a valid assumption.

Matches were obtained for 2,338 of the 2,550 PV systems (i.e., 92 percent of the systems). No DG number match could be found in PowerClerk for 212 of the systems.

4.14.2. PV Production Simulation

Hourly energy was estimated by performing hourly simulations for each system using FleetView by combining system specifications with the SolarAnywhere Enhanced Resolution (1km) hourly data that corresponded to the system's latitude/longitude. (Note: performing the simulation using two half-hour observations rather than one hourly observation would probably improve accuracy).

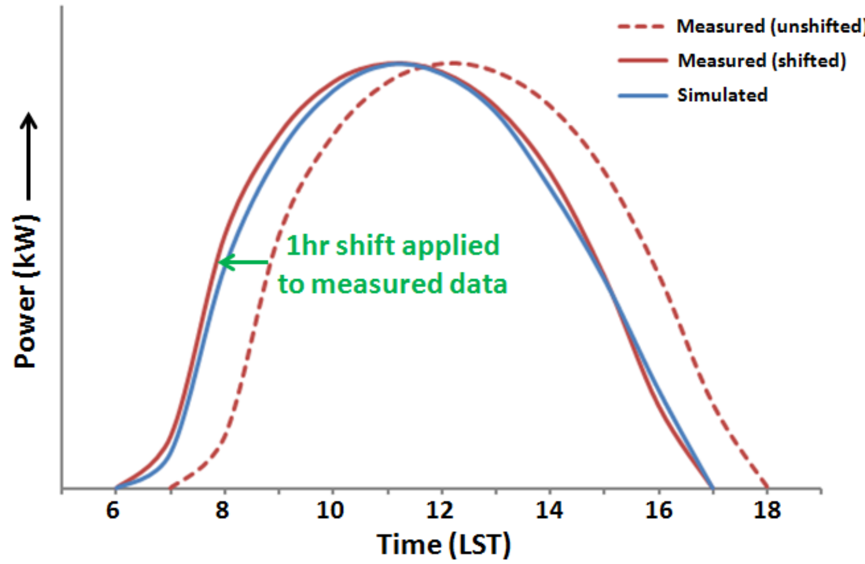
Measured data that contained sub-hourly time intervals were converted to hourly time intervals.

Simulated and measured data were time-correlated (i.e., matched up by date and time). Records were discarded where either the simulated or measured data was missing.

4.14.3. Data Quality Issues

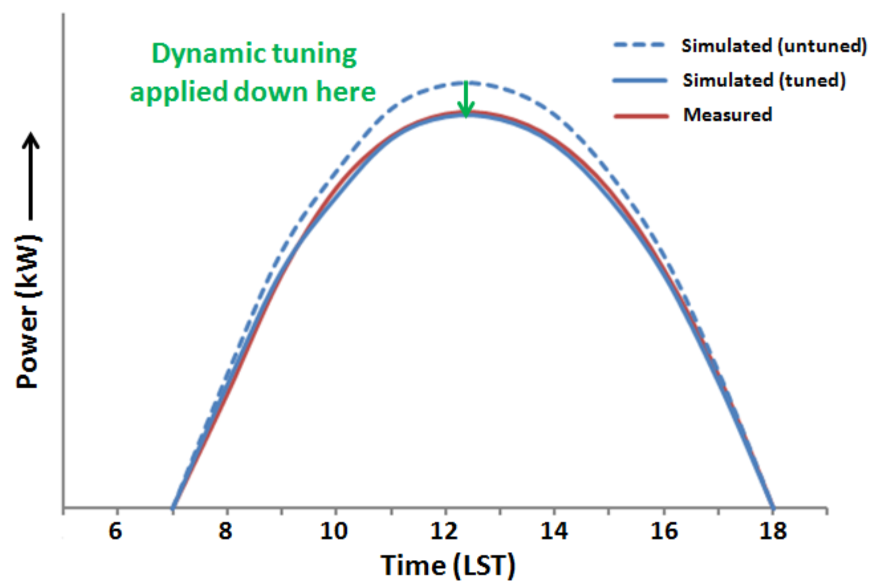
It was determined that some of the measured data did not properly time-correlate with the simulated data. This was corrected by shifting the measured data backward or forward up to 60 minutes in 15 minute intervals (-60,-45,-30, -15, 0, +15, +30, +45,+60). The rMAE was calculated for each time shift. The time shift that resulted in the lowest rMAE was assumed to be the most correct for the measured data. Figure 30 illustrates this measured data time shift procedure for one day for one system.

Figure 30: Illustration of CPR's measured data time shift correction process.



Site-specific tuning was applied to PV simulation results using CPR's dynamic tuning process once the simulated and measured data were time-correlated over the period of examination. A scale factor was selected that minimized certain error characteristics. Figure 31 illustrates the results of the dynamic tuning process for one day for one system.

Figure 31: Illustration of CPR's dynamic tuning methodology.



4.14.4. Results (Fleet of Systems)

Validation results were generated for both individual systems and fleets of systems. This section presents results based on the fleet of systems. Appendix B presents additional details.

CPR completed successful simulations of 2,338 SMUD PV systems of which 132 systems were excluded due to various missing or erroneous measured data issues. Results from the 2,206 remaining PV system simulations are presented here.

The time shift correction (illustrated in Figure 30) was applied to the measured data and the dynamic tuning analysis procedures (illustrated in Figure 31) was applied to simulated results for each PV system. Figure 32 presents the distribution of time shift analysis results for all measured PV systems. The majority of systems required little or no time correction.

Figure 32: Distribution of time shift corrections applied to all PV systems (2206).

The dynamic tuning methodology was applied to each PV system simulation. The distribution of results is presented in Figure 33. While the peak in scaling factors applied is centered about zero, there is strong asymmetry present towards the downscaling side of the distribution. This unevenness in the distribution is likely due to influences which tend to lead to PV system underperformance. These effects can include system soiling, module mismatch and degradation, and enhanced rooftop-related temperature losses. Figure 33 suggests that, in practice, it is more common for a PV system to underperform than to over perform.

Figure 33: Distribution of dynamic scaling factors applied to all locations (2206) derived from the six months of simulated vs. measured production data.

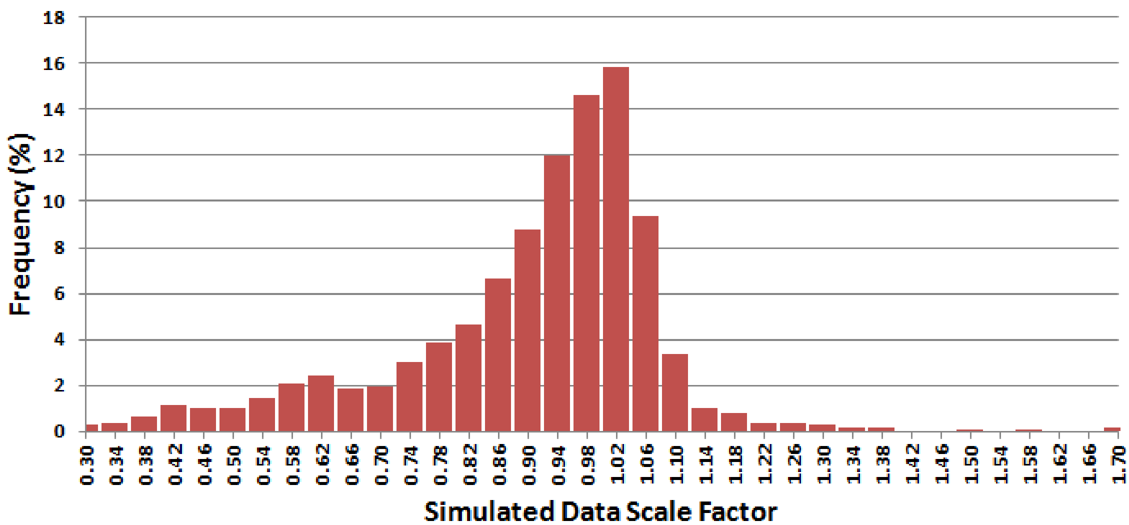


Figure 34 presents 5 days of the aggregate fleet hourly simulation and measured production data for all 2,206 PV systems. Overall, the fleet PV simulations line up with production better than at the individual level due to system wide smoothing effects. As noted before, simulations for clear days tend to line up better with measured data than those for cloudy days. The daily rMAE statistics in Figure 35 confirm that there is lower error on sunny days. There is also less error observed on cloudy days due to aggregating of fleet production.

Figure 36 presents the hourly-averaged MBE. It suggests that at a fleet-level the simulations tend to slightly over predict energy during the morning and late afternoon timeframes while under predicting energy during the peak sunshine part of the day. It is likely that this can be corrected through improvements to the inverter power curve modeling.

Figure 37 illustrates the cloudy vs. clear day simulation aspects of the fleet simulations by breaking down the hourly simulated vs. measured statistics in: (a) all conditions; (b) clear day conditions; and (c) cloudy day conditions. The systematic morning/later afternoon over prediction and midday under prediction tendencies are well illustrated here.

Figure 34: Simulated (red line) and measured (blue line) production for all 2,206 systems.

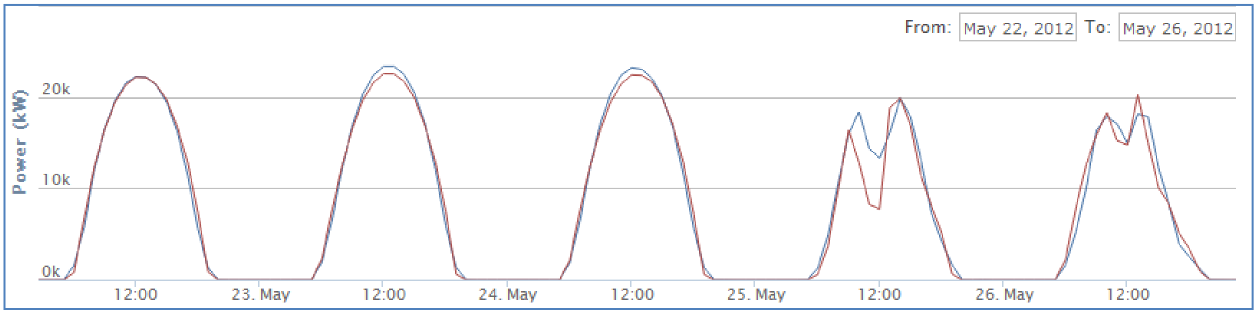


Figure 35: Aggregate daily MAE for all 2,206 systems from 4/16/2013 to 10/10/2013.

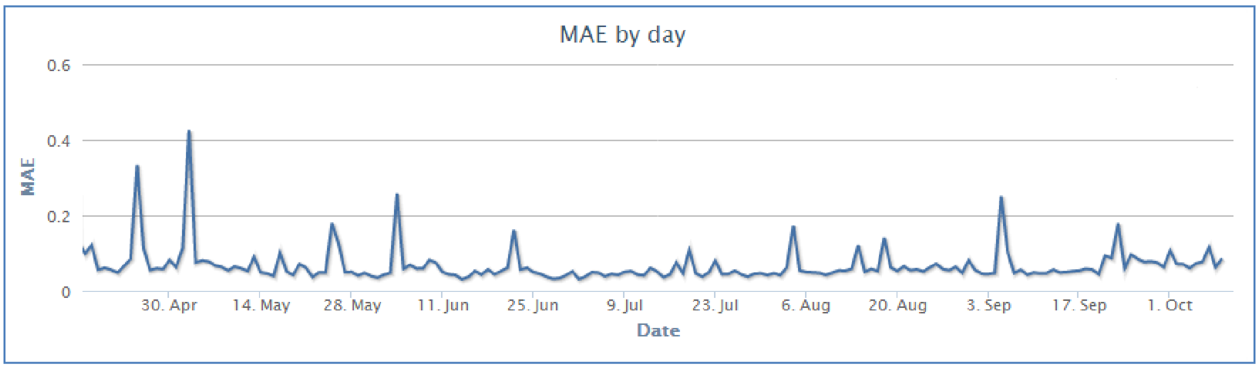


Figure 36: Hourly MBE for all 2,206 systems.

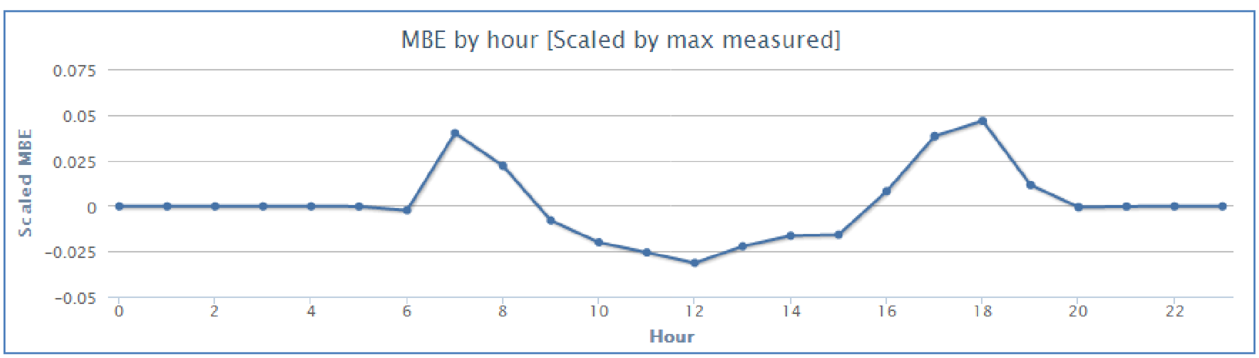


Figure 37: Scatter plot of simulated vs. measured hourly energy production for all 2,206 systems from 4/16/2012 - 10/10/2012 for all day conditions (a), clear days (b) and cloudy days (c).

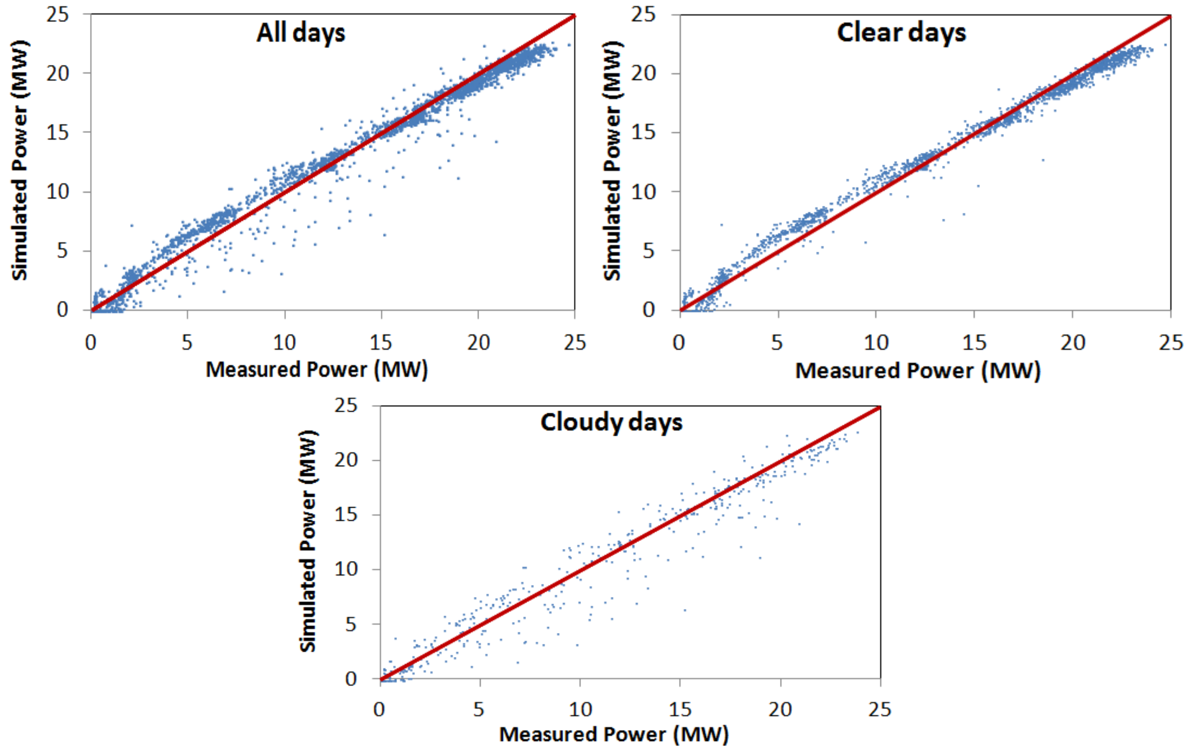


Table 1 presents error statistics for the fleet of 2,206 systems over a six-month period from 4/16/2012 to 10/10/2012. Overall rMAE is 6.2 percent during this observational period under all conditions. This error drops to 5.4 percent when only clear days are included due to the exclusion of higher error prone cloudy days which exhibit 11.1 percent error on their own.

Table 1: rMAE for all 2,206 systems.

	Clear Days	Cloudy Days	All Days
rMAE	5.4%	11.1%	6.2%
Ave Daily Energy	185.8 MWh	143.9 MWh	178.2 MWh
Number of Days	145 days	32 days	177 days

4.14.5. Fleet of Well-Behaved PV Systems

Further full fleet PV system simulation results are presented now. PV systems were removed with reported six-month MAE statistics higher than 10 percent to filter out some of the noise present in the fleet simulation process. This reduced the simulation pool to 1,102 systems.

Figure 38 presents examples of the trimmed down aggregate fleet hourly simulation and measured production data. Good partly cloudy day alignment can be see with mostly cloudy days still presenting challenges. The daily rMAE statistics in Figure 39 confirm the presence of lower error on sunny days with less error also observed on cloudy days due to the aggregation of fleet PV production. The highest noted daily rMAE error day (May 3) is presented in Figure 38. Heavy overcast cloud conditions dominated the SMUD-footprint region on May 3 which resulted in lower energy simulations due to the under prediction of surface irradiance.

The improvement in fleet error statistics is further illustrated in the hourly-averaged MBE presented in Figure 40. There is less morning and afternoon error while the previously noted midday under prediction error almost disappears. Figure 41 further illustrates the cloudy vs. clear day simulation aspects of the fleet simulations by breaking down the hourly simulated vs. measured statistics during: (a) all conditions; (b) clear day conditions; and (c) cloudy day conditions.

Figure 38: Simulated (red line) and measured (blue line) production for 1,102 well-behaved systems over a four day period in May.

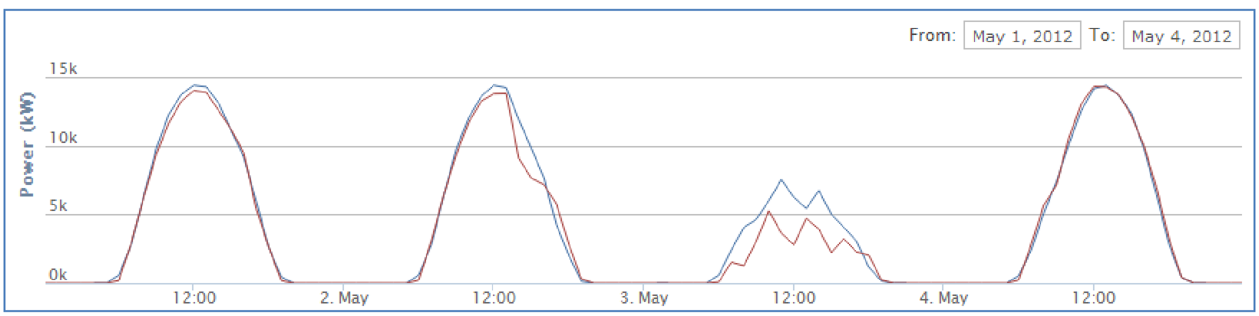


Figure 39: Aggregate daily MAE for 1,102 well behaved systems from 4/16/2012 to 10/10/2012.

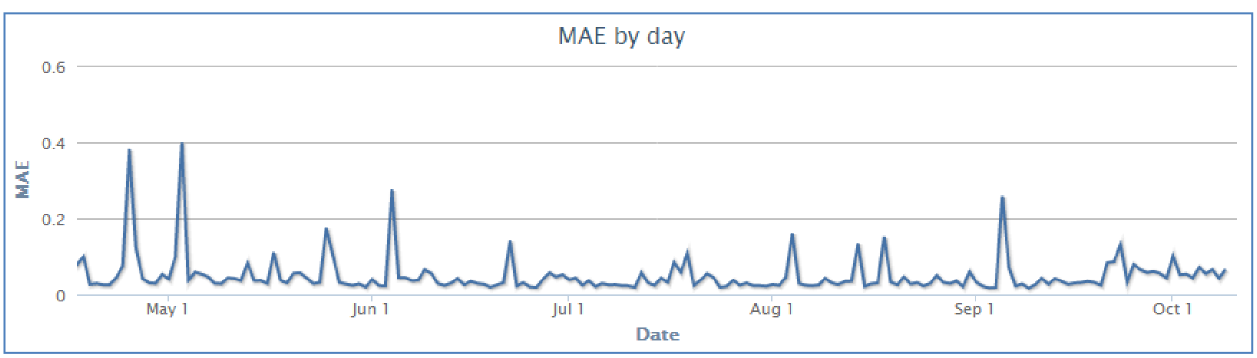


Figure 40: Hourly MBE for 1,102 well behaved systems.

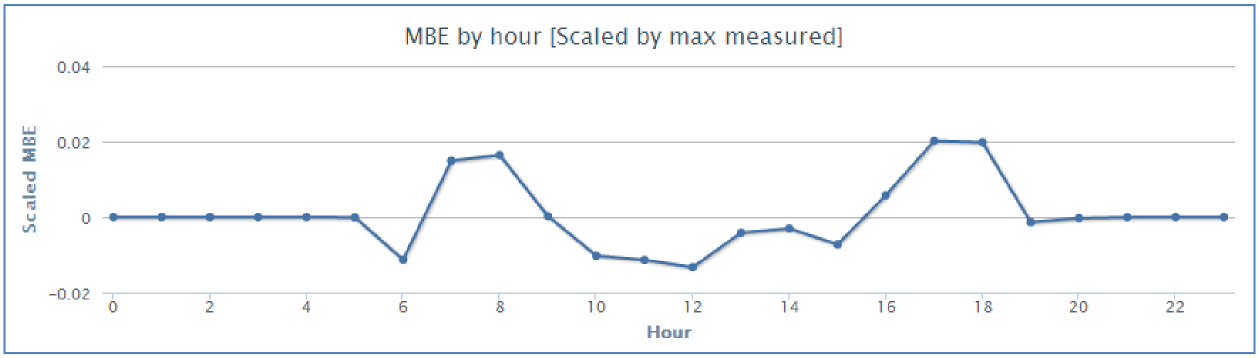


Figure 41: Scatter plot of simulated vs. measured hourly energy production for 1,102 well-behaved sites from 4/16/2012 to 10/10/2012 for all day conditions (a), clear days (b) and cloudy days (c).

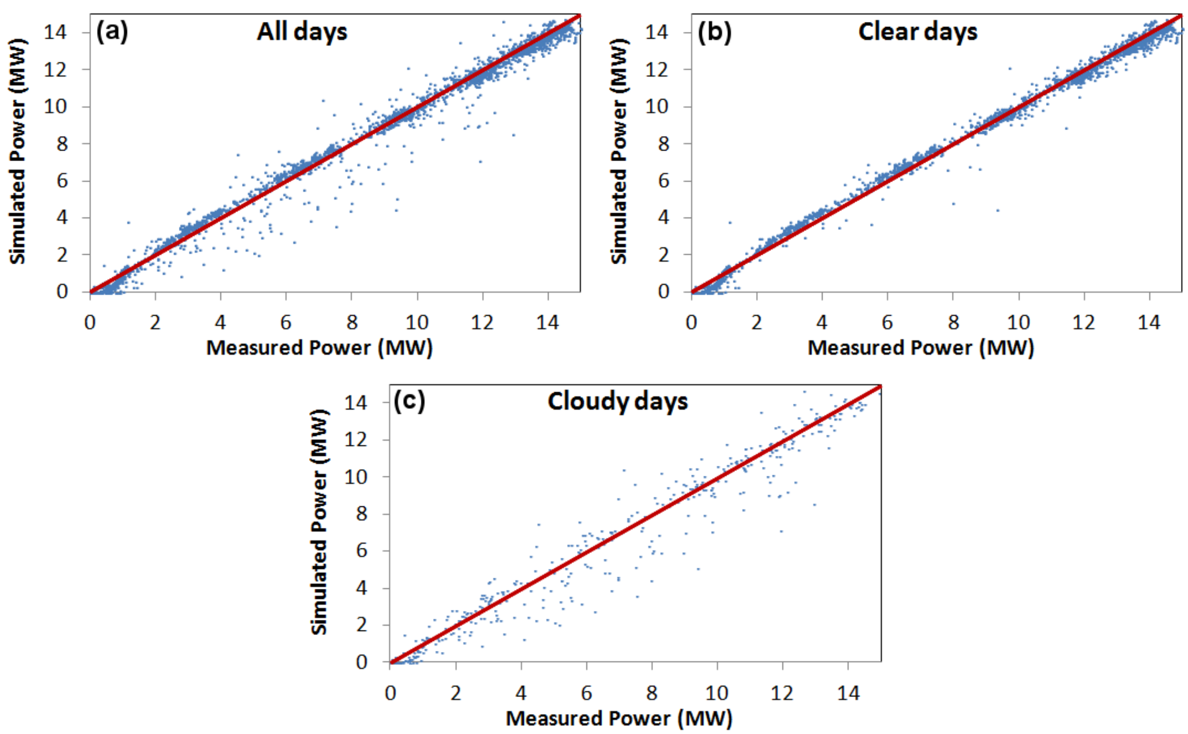


Table 2 presents error statistics for the fleet of 1,102 well-behaved systems over a six-month period from 4/16/2012 to 10/10/2012. Overall rMAE is 4.5 percent during this observational period under all conditions. This error drops down to 3.5 percent when only clear days are included due to the exclusion of higher error prone cloudy days which exhibit 10.0 percent error on their own.

Table 2: rMAE for 1,102 well-behaved systems.

	Clear Days	Cloudy Days	All Days
rMAE	**3.5%**	**10.0%**	**4.5%**
Ave Daily Energy	**112.8 MWh**	**88.3 MWh**	**108.4 MWh**
Number of Days	**145 days**	**32 days**	**177 days**

5. Technology Transfer

A substantial amount of work was performed under the grant agreement. In order to make the results as useful as possible, the results were extensively documented. The documentation includes six patent applications, a peer reviewed journal article, seven conference papers, and a state-of-the art solar resource database for all of California.

5.1.1. Pending Patents

The six patent applications include computer-implemented methods for:

- Tuning photovoltaic power generation plant forecasting.
- Bounding accuracy on a forecast of photovoltaic fleet power generation.
- Inferring operational specifications of a photovoltaic power generation system.
- Correlating satellite imagery for use in photovoltaic fleet output estimation.
- Estimating photovoltaic energy generation for use in photovoltaic fleet operation.
- Bounding accuracy on correlated overhead sky clearness for use in photovoltaic fleet output estimation.

5.1.2. Journal Article

The journal article includes:

- "Using Satellite Insolation Data to Calculate PV Power Output Variability." Paper published in Photovoltaics International, Second Quarter, May 2013, pages 94-99.

5.1.3. Conference Presentations

The conference presentations include:

- "Behind-the-Meter PV Fleet Forecasting." Presentation and paper presented at ASES Solar World 2013. Baltimore, MD, April 2013.
- "Behind-the-Meter PV Fleet Forecasting: Results for 130,000 PV Systems in California" Presentation at SEPA Utility Solar Conference. Portland, Ore., April 2013. Presentation available at: http://www.cleanpower.com/wp-content/uploads/SPI-USC-2013-04-03.pdf
- "Integrating PV Into Utility Planning and Operation Tools." Presentation at DOE/CPUC High Penetration Solar Forum. San Diego, CA, Feb., 2013. Presentation available at: http://www.cleanpower.com/wp-content/uploads/SolarForum2013.pdf
- "Forecasting Output for 130,000 PV Systems in California." Presentation at Utility Wind Integration Group (UVIG) Workshop on Variable Generation Forecasting Applications to Power System Planning and Operations. Salt Lake City, UT, Feb., 2013. Presentation available at: http://www.cleanpower.com/wp-content/uploads/Forecasting-Output-for-130000-PV-Systems-in-California.pdf.
- "Behind-the-Meter PV Fleet Forecasting." Presentation at Utility Wind Integration Group (UVIG) Fall Technical Workshop. Omaha, NE, Oct., 2012.

- "Accuracy of Solar Modeling & Forecasting." Presentation at Solar Power International. Orlando, FL, Sep. 2012.
- "Solar Monitoring, Forecasting, and Variability Assessment at SMUD." Presented at WREF 2012 (SOLAR 2012). Denver, CO, May 2012. Paper available at. http://www.cleanpower.com/wp-content/uploads/SMUD-Solar-Assessment_2012.pdf.

5.1.4. Solar Data

The solar data was made publicly available:

- SolarAnywhere Enhanced Resolution (1 km x 1 km, half-hour) freely available to general public at www.solaranywhere.com.
- SolarAnywhere High Resolution (1 km x 1 km, one-minute) data used to forecast PV fleet production for CAISO.

6. Conclusions

6.1. Key Findings

The California Solar Initiative (CSI) has a goal of installing 3,000 MW of new solar electricity by 2016. CSI has identified that one potential barrier to accomplishing this goal is planning and modeling for high-penetration PV grid integration issues. A team led by Clean Power Research (CPR) received approval from the California Public Utilities (CPUC) for a grant titled, "Integrating PV into Utility Planning and Operation Tools."

The project accomplished the following grid-integration tasks:

1. Extend the SolarAnywhere Enhanced Resolution solar resource database, create high resolution (1 km, 1-minute resolution) solar resource data, and benchmark data accuracy.
2. Validate previously developed PV fleet simulation methodologies using measured ground data from fleets of PV systems connected to California's grid.
3. Integrate PV fleet simulation methodologies into utility software tools for use in activities ranging from distribution planning to balancing area operation using CAISO as a test case.

Key conclusions from this work are:

- High resolution solar resource data can be accurately produced.
- This solar resource data can be combined with PV system specifications to accurately simulate PV fleet production.
- The simulation process can be performed quickly enough to support even the challenging application of forecasting production for hundreds of thousands of systems while meeting forecasting time horizon requirements using the appropriate computing resources and underlying system architecture.

6.2. Benefits to California Ratepayers

This project has provided a number of benefits to the state of California.

6.2.1. Solar Resource Data

The first task was to extend SolarAnywhere. SolarAnywhere Enhanced Resolution provides 1 km x 1 km spatial resolution with half-hour temporal resolution irradiance data. It is beneficial in that it is comprehensive for all of California and is freely available at www.SolarAnywhere.com. California's project developers are also leveraging the increased Enhanced Resolution data accuracy to obtain lower financing rates because of reduced project risk; this lowers the cost of solar and increases the penetration of PV in the state. SolarAnywhere High Resolution extends the Enhanced Resolution to one-minute temporal resolution. The High Resolution data is used in PV penetration and variability studies as well as in solar forecasting for the CAISIO as described below.

6.2.2. PV Fleet Simulation Validation

The second task was to validate PV fleet simulation methodologies using measured ground data from fleets of PV systems connected to California's grid. It is critical to the utilities and balancing area authorities responsible to run the grid that they validate models using real-world data. The validation provides public benefits because grid operators need to gain confidence in the models intended to inform grid operation prior to their use.

6.2.3. PV Fleet Simulation Integration Into Utility Software Tools

The third task was to integrate PV fleet simulation methodologies into utility software tools. CAISO has the responsibility of maintaining reliability and accessibility for California's utility grid. As such, they are concerned with the effect of power production from customer-owned PV systems on the balancing area. Prior to this contract, CAISO did not have visibility into the performance of behind-the-meter PV systems. CPR has been providing behind-the-meter PV fleet forecasts every 30 minutes to CAISO for almost one year. This is beneficial to California in that CAISO has visibility into behind-the-meter PV performance when none existed prior to this grant. It has the additional benefit of being a valuable case study for California's IOUs as they consider using the same approach for their needs.

6.3. Potential Next Steps

The next steps of this work could be as follows:

- Continue to improve high resolution solar resource forecasting accuracy.
- Implement a streamlined interconnection process to simultaneously collect PV system specifications in order to continue to be able to define the PV fleet.
- Transition PV fleet forecasting from R&D to an operational environment and integrate into utility tools.
- Design and implement probabilistic/ramp event PV fleet forecasting system.
- Continue to validate results.

The tools and data streams developed as part of this work will be made available to California utilities, ISOs and others to help cost-effectively and reliably integrate distributed PV into the grid.