Inferring PV System Specifications from Net Load and Reconstructing Gross Load

Upama Nakarmi
Clean Power Research
April 2022

Legal Notice from Clean Power Research

This report was prepared by Clean Power Research. This report should not be construed as an invitation or inducement to any party to engage or otherwise participate in any transaction, to provide any financing, or to make any investment.

Any information shared prior to the release of the report is superseded by the report. Clean Power Research owes no duty of care to any third party, and none is created by this report. Use of this report, or any information contained therein, by a third party shall be at the risk of such party and constitutes a waiver and release of Clean Power Research, its directors, officers, partners, employees and agents by such third party from and against all claims and liability, including, but not limited to, claims for breach of contract, breach of warranty, strict liability, negligence, negligent misrepresentation, and/or otherwise, and liability for special, incidental, indirect, or consequential damages, in connection with such use.

Executive Summary

Behind-the-Meter PV Systems are Prevalent and Utilities Need to Predict Their Performance

The US Energy Information Administration (EIA) estimated that residential PV systems accounted for 61 percent of total annual generation from small scale distributed PV systems in 2021.[1] The amount of power coming from these Behind-The-Meter (BTM) PV systems emphasizes the need for utilities and system operators to accurately predict PV system performance. One way to predict BTM PV system performance is by simulating production by combining irradiance data with PV system specifications.

Irradiance Data and PV System Specifications are Crucial for Accurate PV Simulation

Clean Power Research's (CPR's) SolarAnywhere® product is an accurate source of irradiance data.[2] But what is an accurate source of PV system specifications? Some utilities require BTM PV system owners to submit system specifications such as azimuth, tilt, and PV size upon interconnection to the utility. CPR's PowerClerk® has processed about half of the BTM PV systems interconnected in the US. In these cases, system specifications may be available through PowerClerk.[3] Utilities that do not use a tool such as PowerClerk may only have minimal PV system information available.

PV System Specs can be Inferred from Net Load Data

This report describes a low-cost, accurate, and scalable software solution that infers PV system specifications from net load data and PV system location. The inferred system specifications include azimuth, tilt, and PV system size. Net load includes the combined effect of a customer's production and consumption (i.e., gross load). Many utilities use only a single meter for their customers. These utilities collect net load data for billing purposes.

Inference Method Applied to 8,000+ BTM PV Systems in Sacramento Municipal Utility District's Service Territory

We apply the method in this report to 8,324 PV sites in the territory served by the Sacramento Municipal Utility District (SMUD). SMUD customers who own BTM PV systems have two meters. The primary meter measures net load and the additional meter measures PV production. As a result, SMUD represents a valuable test case to validate the method's accuracy and compare it to other methods.

Method Validation

This report presents a detailed comparison between PV system specifications inferred using various methods and data sources, shown in Table 1. In addition to the system specifications inferred using net load, we infer PV system specifications using two other methods. We also use the installer/user reported PV system specifications in PowerClerk. The comparison among these various methods and

---

[1] U.S. Energy Information Administration (EIA) https://www.eia.gov/electricity/monthly/ [Online; accessed 26-April-2022].
[2] SolarAnywhere® https://www.solaranywhere.com. [Available Online]
[3] PowerClerk® https://www.cleanpower.com/products/powerclerk/ [Available Online]

metadata sources establishes a baseline to measure the performance of the developed approach as well as the type of metadata source used.

Table 1: Methods and metadata sources of PV system specifications.

Method Name	Input Data	PV System Specifications
Spec Inference from Net Load	Net load time series data	Minimal metadata
Spec Inference from PV	PV production time series data	Minimal metadata
PowerClerk	Direct reported specifications	Minimal to detailed metadata
Extended Spec Inference from PV	PV production time series data	Detailed metadata

The first and second methods use Spec Inference, a tool developed by CPR to infer PV system specifications using both net load and PV production time series data. Spec Inference infers minimal but key PV system specifications/metadata including azimuth, tilt, and PV size. The third method uses PV system specifications reported in PowerClerk by the installers/users. The reported PV system specifications range from minimal metadata to detailed metadata. The fourth and final method uses the Extended Spec Inference, a tool developed by CPR to infer system specifications using only PV production data. This method does not infer system specifications using net load data. The Extended Spec Inference tool infers detailed PV system metadata including the key parameters such as azimuth, tilt, and PV size as well as additional parameters such as tracking type and shading obstructions. Using PV system specifications from the four different methods and metadata sources, PV production is simulated by combining system specifications with irradiance data.

Report Outline

Sections 2 and 3 present a brief introduction and objective of the case study, respectively. Section 4 discusses Spec Inference, CPR's software-based tool designed to infer PV system specifications using net load data. Section 5 discusses the modifications made to the net load-based Spec Inference, to infer PV system specifications using PV production data. Section 6 discusses the results obtained for PV system specifications inferred using the different methods and metadata sources. Section 7 summarizes the key results and concludes the report.

Simulation Results for One Day

Figure 1 compares hourly simulated PV production and measured PV production on June 22, 2016, for the fleet of 8,324 PV sites using the four different methods and metadata sources. In all four cases, the measured PV represents the PV production data collected by the additional meter at the SMUD sites. In Figure 1a, simulated PV production is estimated using system specifications inferred by Spec Inference from Net Load. Similarly, in Figure 1b, simulated PV production is estimated using system specifications inferred by Spec Inference from PV production. In Figure 1c, simulated PV production is generated using system specifications reported in PowerClerk. Finally, in Figure 1d, simulated PV production is generated using system specifications inferred by Extended Spec Inference from PV production. System specifications from PowerClerk and Extended Spec Inference include shading elevation angles in multiple azimuth directions. The hourly relative Mean Absolute Errors (rMAE's) are 5.9%, 5.2%, 3.0%, and 3.6% for the system specifications obtained via Spec Inference from Net Load, Spec Inference from PV production, PowerClerk, and Extended Spec Inference from PV production, respectively.[4]

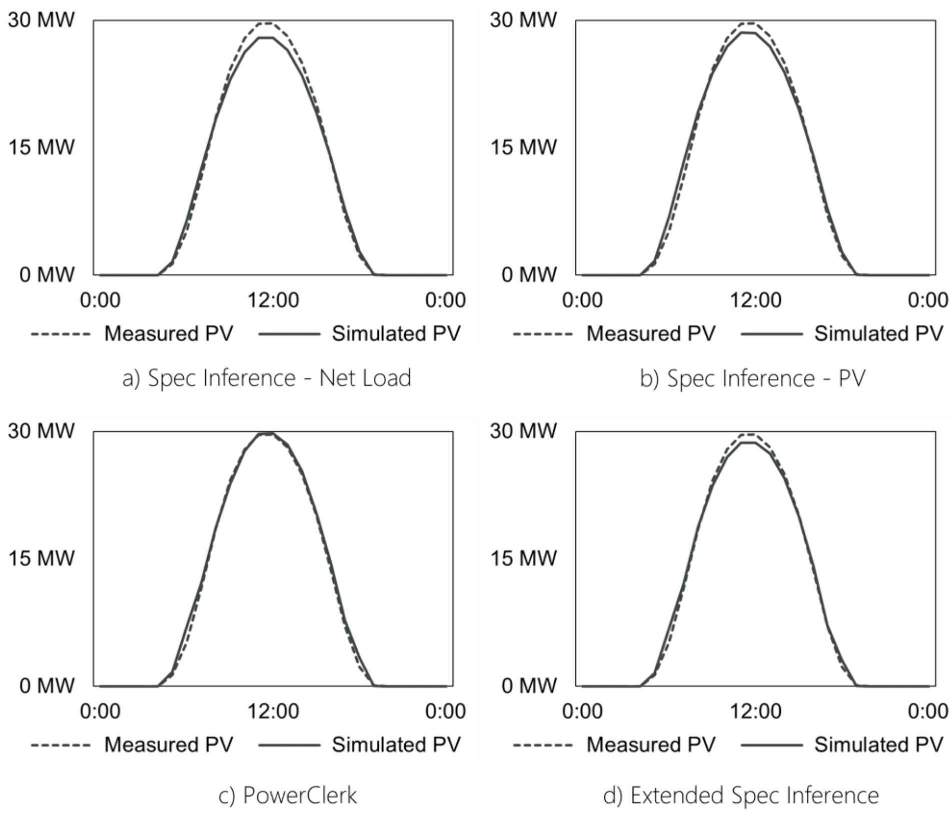


Figure 1: Measured vs simulated production during June 22, 2016, for the fleet of 8,324 sites.

[4] T. E. Hoff, R. Perez, J. Kleissl, D. Renne, and J. S. Stein. "Reporting of irradiance model relative errors." American Solar Energy Society (ASES) annual conference, Denver, CO, USA, pp. 13-17, 2012.

Simulation Results for the Entire Year

Figure 2 compares hourly simulated and measured production during the entire year (2016) for the fleet of 8,324 PV sites using the four different methods and metadata sources. The hourly rMAE's obtained are 9.4%, 9.1%, 7.5%, and 6.0% for the PV system specifications inferred by Spec Inference from Net Load, Spec Inference from PV production, PowerClerk, and Extended Spec Inference from PV production, respectively. The lowest error is obtained using system specifications derived using Extended Spec Inference from PV production. The rMAE's obtained using system specifications inferred by Spec Inference from Net Load and Spec Inference from PV production are comparable.

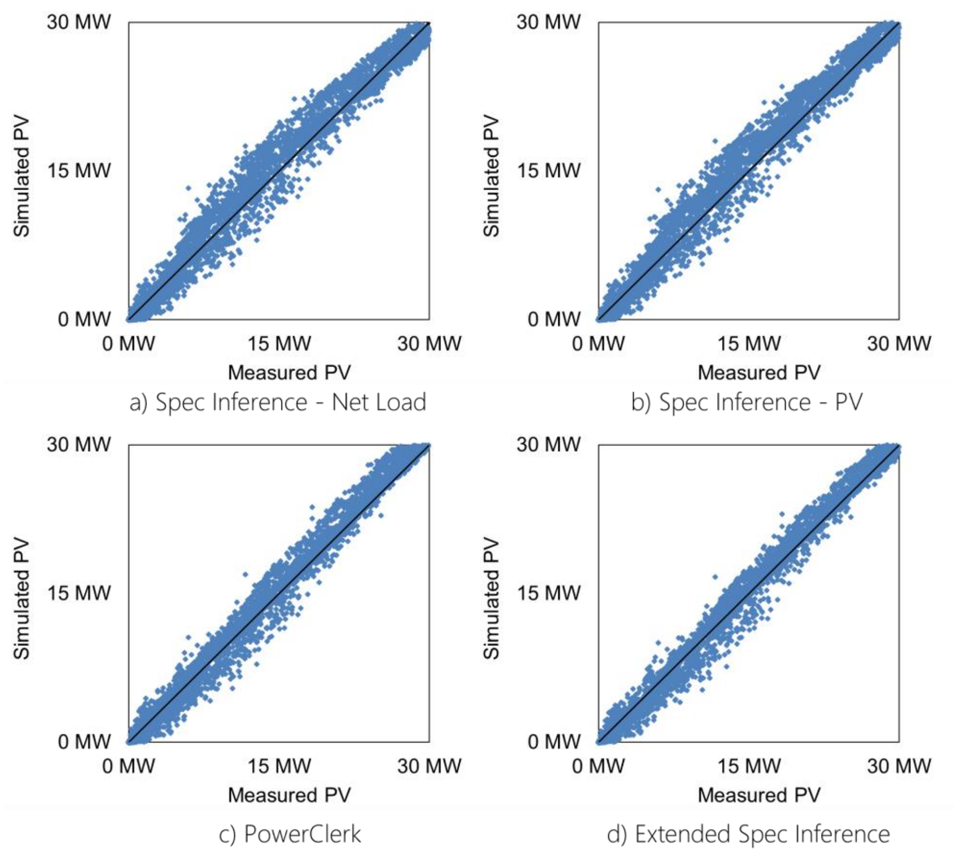


Figure 2: Measured vs simulated production during 2016 for the fleet of 8,324 sites.

Comparison between Fleetwide Simulation Results

Figure 3 summarizes the rMAE's between the hourly measured and simulated production during the entire year (2016) for the fleet of 8,324 PV sites using the four different methods and metadata sources. The figure also compares the rMAE between the hourly measured and simulated production for three additional cases when: 1) shading and soiling losses are not considered, 2) only shading losses are considered, and 3) both shading and soiling losses are considered. For the purposes of this study, we consider soiling losses as the PV sites are located in Sacramento, CA, where rainfall is limited. The blue bars represent rMAE for system specifications inferred by Spec Inference from Net Load and Spec Inference from PV production, when shading and soiling losses are not considered. The yellow bars, for the Spec Inference from Net Load and Spec Inference from PV production, represent the rMAE when a constant horizon shading elevation angle of 10 degrees is considered, while generating the simulated production.

The yellow bars, for the PowerClerk and Extended Spec Inference, represent the rMAE when shading in multiple azimuth directions is considered. For PowerClerk, shading elevation angles in the multiple azimuthal directions are reported for each array separately. For the Extended Spec Inference, shading elevation angles in the multiple azimuthal directions are inferred for the overall system only. For all four cases, the black bars represent the rMAE when shading and soiling are both considered for generating the simulated production. Table 2 shows the hourly rMAE obtained between the measured and simulated production during 2016 for the fleet of 8,324 PV sites.

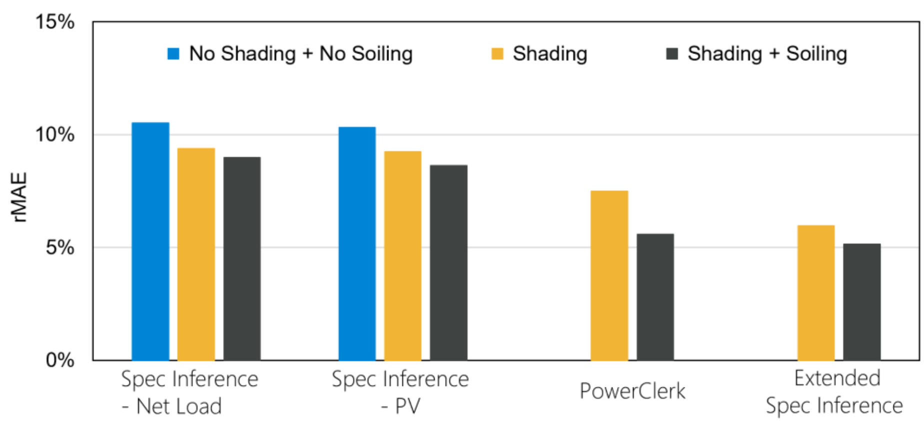


Figure 3: Hourly rMAE between measured and simulated production in 2016 for the fleet of 8,324 sites using system specifications inferred by: 1) Spec Inference from Net Load, 2) Spec Inference from PV production, 3) PowerClerk, and 4) Extended Spec Inference from PV production.

Table 2: Hourly rMAE between measured and simulated production for the fleet of 8,324 sites in 2016.

Method Name	No shading and No Soiling	Shading	Shading and Soiling
	Hourly rMAE	Hourly rMAE	Hourly rMAE
Spec Inference from Net Load	10.5%	9.4%	9.0%
Spec Inference from PV	10.3%	9.2%	8.6%
PowerClerk		7.5%	5.6%
Extended Spec Inference from PV		6.0%	5.1%

As expected, the consideration of shading and soiling losses yields the lowest error for each of the four different methods and metadata sources. Three key observations from Figure 3 and Table 2 are as follows:

1. Among the four different inference methods and metadata sources, system specifications inferred by Extended Spec Inference from PV production data yields the lowest error. This is because the production data is encoded with the real-world conditions under which the PV system operates. Further, Extended Spec Inference considers shading elevation angles in multiple azimuth directions. This contributes to the overall lower error rates. The system specifications obtained from Extended Spec Inference may not be the exact specifications of the PV system. For example, a site with a PV system may have multiple arrays oriented in different directions. One array may face towards the east while another may face towards the west. In such cases, Extended Spec Inference does not infer system specifications for each array separately. It only infers one set of specifications for a site. Regardless of this, these system specifications are the best among the above-mentioned methods. Thus, we recommend using system specifications obtained from Extended Spec Inference if PV production data is available.

2. The system specifications from PowerClerk have lower errors compared to the system specifications inferred by Spec Inference from Net Load or Spec Inference from PV production. We attribute this to the accuracy and detail of the reported system specifications for the SMUD sites, which are collected and reported for each array in a PV system separately. This may not be the case for other electric utilities. Generally, utilities compensate for the lack of PV system metadata by making assumptions based on location of the PV system. For example, PV systems in the Northern hemisphere are assumed to have an azimuth angle of due south and tilt approximately equal to the latitude of the site. Therefore, the reported system specifications may be inaccurate and unreliable.

3. The error rates obtained from the system specifications inferred by Spec Inference from Net Load and Spec Inference from PV production are comparable to each other. This proves that CPR's Spec Inference tool for inferring system specifications from net load is working well. This also indicates the potential for utilities that only collect net load data to infer the system specifications of the PV systems installed in their service territories.

Gross Load Reconstruction

Gross load or total load is the actual demand at the customer sites. This information is useful for utilities to prepare for contingencies and worst-case scenarios. Gross load can be estimated by adding PV production to net load.

Gross Load Results for One Day

Figure 4 compares hourly simulated and measured gross load on June 22, 2016, for the fleet of 8,324 PV sites using the four different inference methods and metadata sources. Figure 4a, Figure 4b, Figure 4c, and Figure 4d estimate simulated gross load by adding the net load to the simulated production estimated using system specifications derived by 1) Spec Inference from Net Load, 2) Spec Inference from PV production, 3) PowerClerk, and 4) Extended Spec Inference from PV production, respectively. The system specifications inferred by Spec Inference from Net Load and Spec Inference from PV production include a constant horizon shading elevation angle of 10-degree. The system specifications from PowerClerk and the Extended Spec Inference include shading elevation angles in multiple azimuth directions. In all four cases, soiling losses are also considered. The hourly rMAE's were 4.9%, 4.2%, 2.2%, and 3.2% for Figure 4a, Figure 4b, Figure 4c, and Figure 4d, respectively.

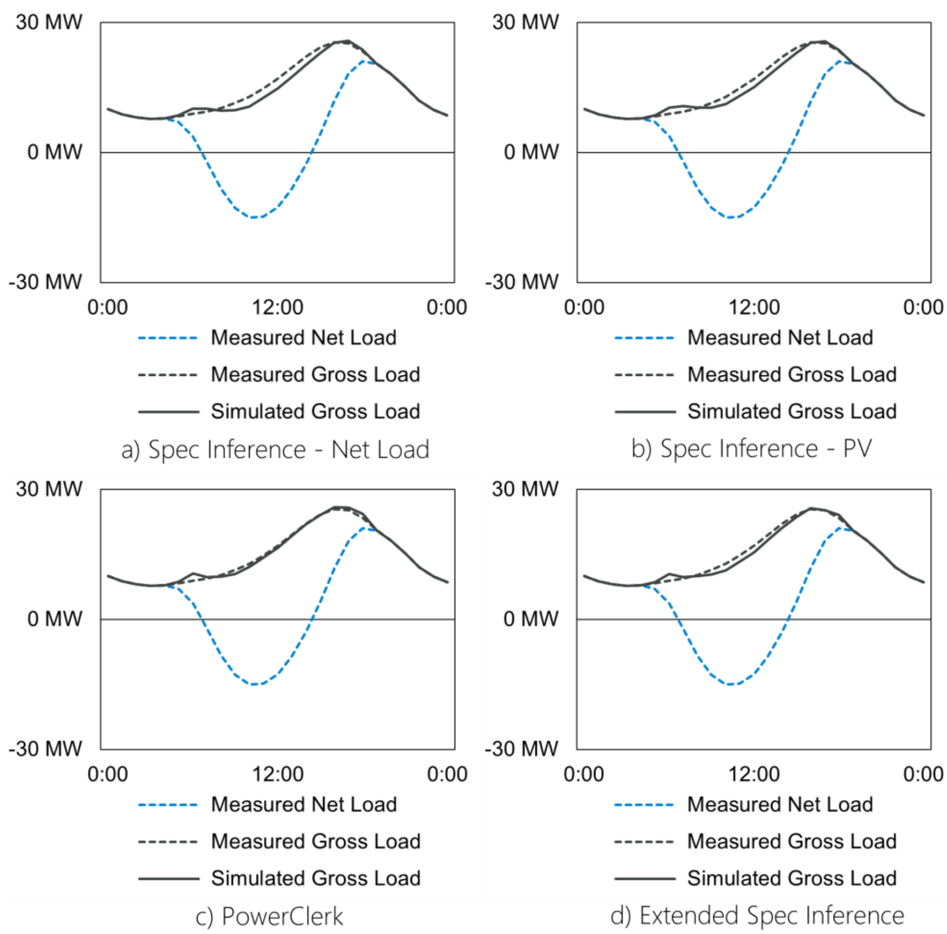


Figure 4: Measured vs simulated gross load during June 22, 2016, for the fleet of 8,324 sites.

Gross Load Results for the Entire Year

Figure 5 compares hourly simulated and measured gross load during the entire year (2016) for the fleet of 8,324 PV sites using the four different inference methods and metadata sources. The system specifications inferred by Spec Inference from Net Load and Spec Inference from PV production include a constant horizon shading elevation angle of 10 degrees. The PowerClerk system specifications and the system specifications from Extended Spec Inference include shading elevation angles in multiple azimuth directions. In all four cases, soiling losses are also considered. The hourly rMAE's obtained were 5.3%, 5.1%, 3.3%, and 3.0% for Figure 5a, Figure 5b, Figure 5c, and Figure 5d, respectively.

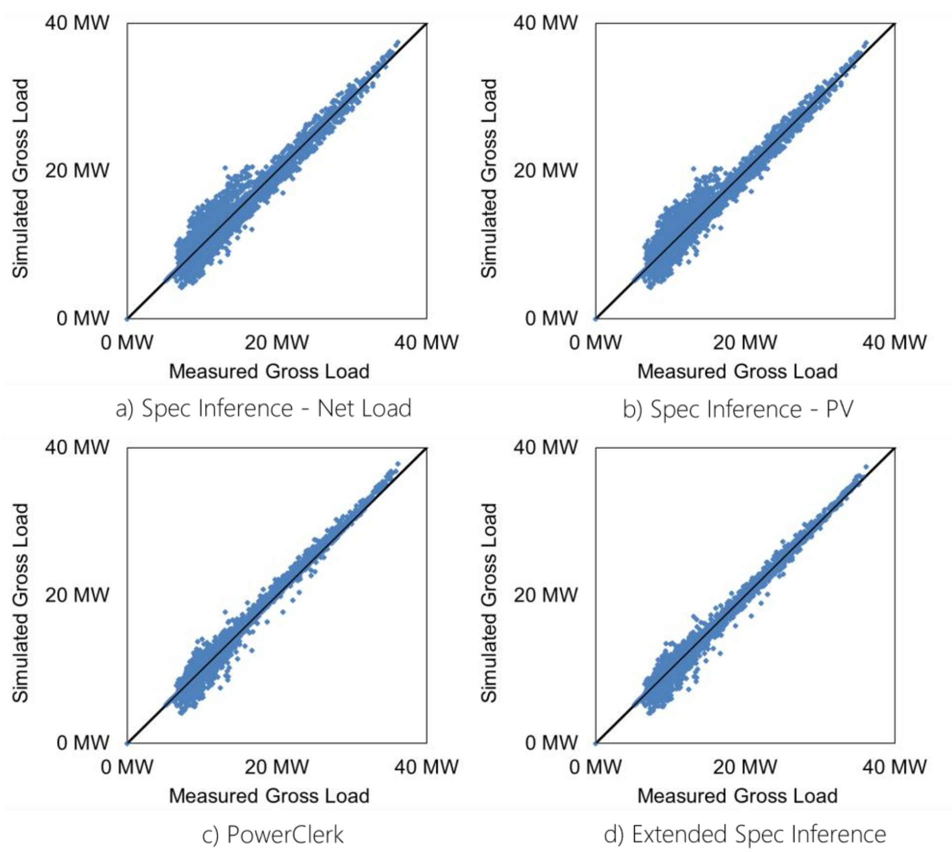


Figure 5: Measured vs simulated production during 2016 for the fleet of 8,324 sites.

PV Specs can be Accurately Inferred from Net Load Data

The key conclusion of this work is that PV system specifications derived using net load data are comparable to system specifications derived using PV production data. This implies that the widely available net load data can be used as a substitute for PV production to infer system specifications. These specifications can then be combined with high-quality solar irradiance data to simulate PV production and predict PV fleet performance.

1 Table of Contents

2 Introduction

A growing number of homes and businesses have installed PV on their buildings. In 2021, residential PV systems accounted for 61% of total annual generation from small-scale distributed PV systems.[1] This share is expected to increase each year. These PV systems are known as Behind-The-Meter (BTM) PV systems and have changed customers from consumers into prosumers. That is, they both consume electricity and produce electricity.

BTM PV systems provide technical and operational challenges to many electric utilities.[5] These utilities have a single electric meter at the customer's location. As a result, the meter measures the combined effect of the customer consuming electricity and the PV system producing electricity. This is referred to as net load (i.e., gross/total load minus PV production). The result is that the utilities lack information about the PV system performance. The ideal case for utilities is the availability of readings of both net load and PV production. An important way that utilities could use such information is to use net load and PV production to reconstruct gross load. Gross load would support the utility in their planning process (i.e., to prepare for worst-case outcomes if the PV system goes off-line). It would also support them in more accurately forecasting loads (i.e., separately forecasting consumption and PV production). Gross load would also enable customers to better understand the efficiency investments that they can make to reduce their electricity consumption.

One way to address this limitation requires a hardware-based solution: adding a second meter to homes with PV systems. This, however, is a costly solution that does not scale well. The costs include the initial capital cost of the meter, the on-going cost of maintaining the meter, and the cost of cleaning up the data if/when gaps in the data occur. An alternative feasible solution is to simulate PV production using irradiance data and PV system specifications. System orientation, including azimuth and tilt, are the fundamental inputs to transposition models that generate plane of array irradiance (POAI) from global horizontal and direct normal irradiance components.[6] Then, the generated POAI can be used in conjunction with system specifications (such as PV size) as input to energy models to simulate system production.[7] Finally, the simulated production can be added to net load to reconstruct gross load.

---

[5] O. Gandhi, D. S. Kumar, C. D. Rodríguez-Gallegos, and D. Srinivasan. "Review of power system impacts at high PV penetration Part I: Factors limiting PV penetration." Solar Energy, vol. 210, pp. 181-201, 2020.
[6] R. Perez, P. Ineichen, R. Seals, J. Michalsky, and R. Stewart. "Modeling daylight availability and irradiance components from direct and global irradiance." Solar energy, vol. 44, no. 5, pp. 271-289, 1990.
[7] Dobos AP. 2014. PVWatts Version 5 Manual.

3 Objective

This report discusses CPR's alternative, low-cost, software-based solution. The solution is to determine PV system specifications (azimuth, tilt, and PV size) from net load data. The system specifications are used to simulate time series PV production. The simulated PV production is combined with net load to reconstruct gross load. In addition to the system specifications inferred using net load, PV system specifications are inferred using two other methods. We also use the installer/user reported PV system specifications in PowerClerk. The comparison among these various methods and metadata sources establishes a baseline to measure the performance of the developed approach as well as the type of metadata source used. This report presents a detailed comparison between PV system specifications inferred using four different methods and metadata sources, as shown in Table 3. Furthermore, this report also examines the effect of considering shading and soiling losses along with the inferred system specifications.

Table 3: Methods and metadata sources of PV system specifications.

Method Name	Input Data	PV System Specifications
Spec Inference from Net Load	Net load time series data	Minimal metadata
Spec Inference from PV	PV production time series data	Minimal metadata
PowerClerk	Direct reported specifications	Minimal to detailed metadata
Extended Spec Inference from PV	PV production time series data	Detailed metadata

The first and second methods use Spec Inference, a tool developed by CPR to infer PV system specifications using both net load and PV production time series data. Spec Inference infers minimal but key PV system specifications/metadata i.e., azimuth, tilt, and PV size. The third method uses PV system specifications reported in PowerClerk by the installers/users. The reported PV system specifications range from minimal metadata to detailed metadata. The fourth and final method uses the Extended Spec Inference, a tool developed by CPR to infer system specifications using only PV production data. This method does not infer system specifications using net load data. The Extended Spec Inference tool infers detailed PV system metadata including the key parameters such as azimuth, tilt, and PV size as well as additional parameters such as tracking type and shading obstructions.

4 Spec Inference from Net Load

CPR's solution for estimating the PV system specifications (azimuth, tilt, and PV size) using net load data involves four steps.

1. Find time-of-day and magnitude of minimum net load.
2. Use CPR's base load identification method to estimate the customer's base load.[8]
3. Produce Plane-of-Array Irradiance (POAI) using CPR's SolarAnywhere clear-sky GHI for various azimuth and tilt combinations.
4. Identify azimuth, tilt, and PV size using the time-of-day and magnitude of minimum net load.

4.1 Find time-of-day and magnitude of minimum net load

The first step involves finding the magnitude of minimum net load and time-of-day when the minimum net load occurs. This is repeated for each month of the year. The selection is based on a test that identifies whether the time-of-day of minimum net load contains only PV production and base loads. Using this test, one representative day per month is selected from the net load data. This day is the clearest day among all clear sky days during the month. Then, the time-of-day and magnitude of minimum net load of these selected days are estimated. Figure 6 shows the net load profile of the selected days for an actual customer. In the figure, for January, the 27th is selected as the representative day for that month. Similarly, for February, the 29th is selected as the representative day for that month, and so on.

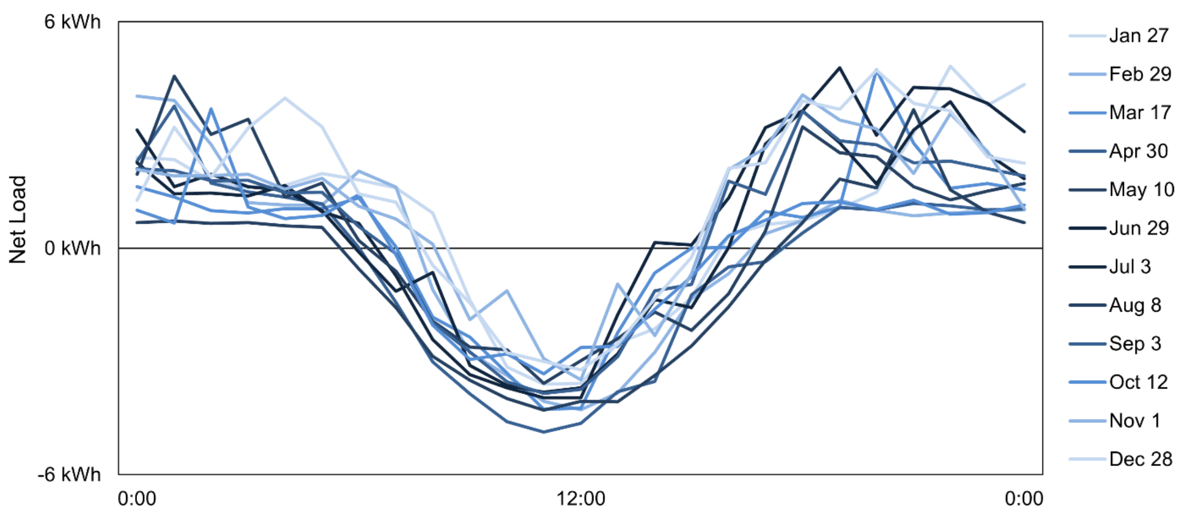


Figure 6: Load profile of selected days per month.

[8] U.S. patent number 10,747.914. Computer-Implemented System and Method for Estimating Electric Baseload Consumption using Net Load Data.

4.2 Estimate base load

The second step is the estimation of the customer's base load using CPR's base load identification method. In this method, frequency distribution of the net load is used to determine the base load. Example, Figure 7 shows the frequency distribution of the net loads for April 2016, for an actual customer. The x-axis and y-axis represent the net loads and their frequency of occurrences, respectively. For this customer, the base load for April 2016 is 1.1 kW.

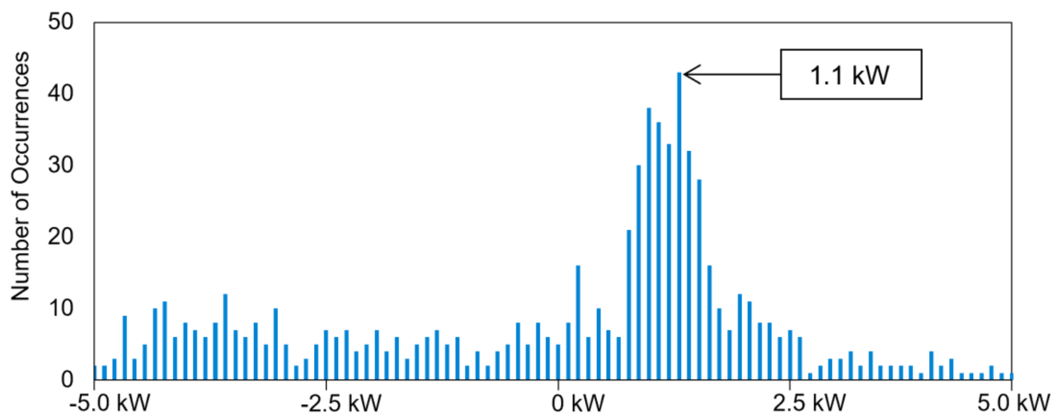


Figure 7: Frequency distribution of loads for April 2016.

4.3 Produce POAI using SolarAnywhere clear-sky GHI for various azimuth/tilt combinations

The third step is to produce POAI using CPR's SolarAnywhere clear-sky GHI in the same location as the site for which the net load data was collected. The POAI is produced for various azimuth and tilt combinations. The number of azimuth and tilt combinations in consideration can vary. For example, azimuths from 0° to 355° in 5-degree increments and tilts from 0° to 90° in 5-degree increments suggest 1368 azimuth/tilt combinations. Whereas azimuths from 0° to 350° in 10-degree increments and tilts from 0° to 90° in 5-degree increments suggest 684 azimuth/tilt combinations. Each azimuth/tilt combination along with SolarAnywhere clear-sky GHI is used to produce POAI for the days selected as the representative days from the net load data in the first step. Then, two parameters are extracted from each day: time-of-day of maximum POAI and magnitude of maximum POAI. Figure 8 shows the time-of-day of maximum POAI in terms of deviation from solar noon. The color bar below the image shows the deviation from solar noon in hours. Figure 9 shows magnitude of maximum POAI. The color bar below the image shows the magnitude in W/m^2. The figures are produced for various azimuths (with 5-degree increments) as shown in the x-axis and for a tilt of 25 degrees. Here, 0-degree azimuth represents due north, 90-degree represents due east, 180-degree azimuth represents due south, and 270-degree represents due west.

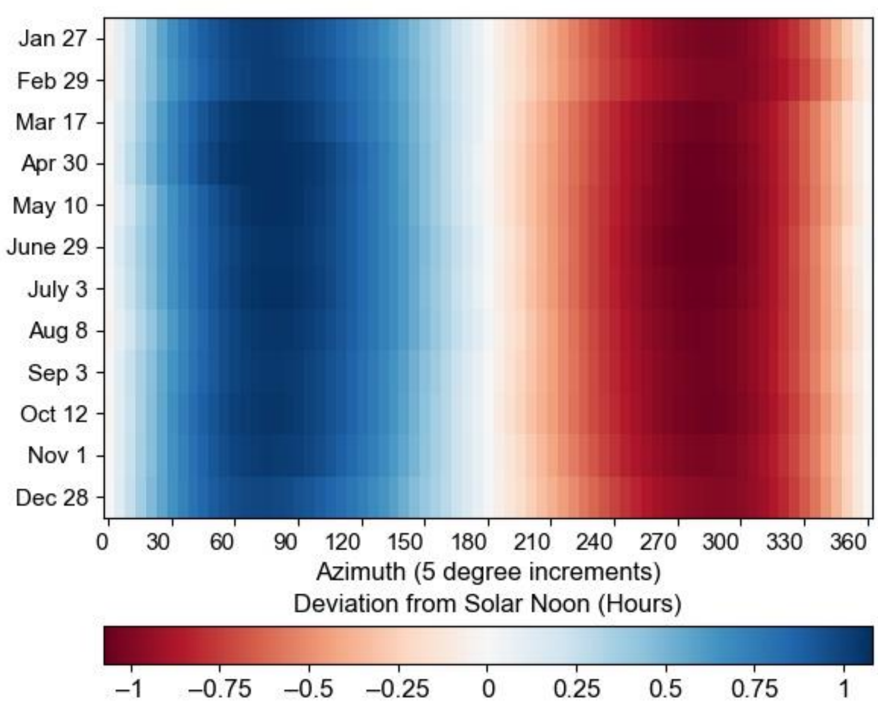


Figure 8: Time of day of maximum POAI for azimuths from 0° to 360° in 5-degree increments and tilt 25°.

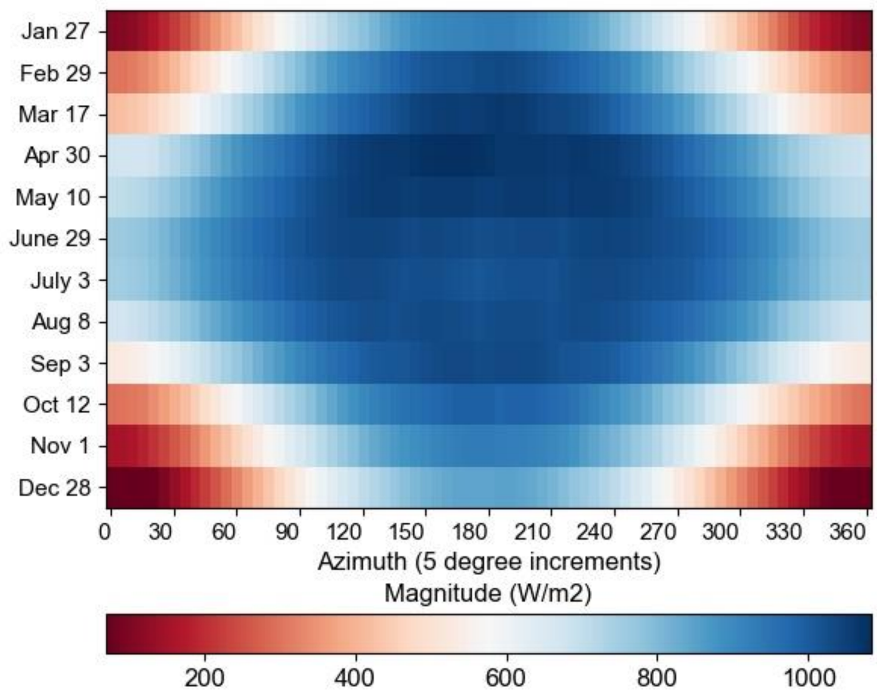


Figure 9: Magnitude of maximum POAI for azimuths from 0° to 360° in 5-degree increments and tilt 25°.

4.4 Identify azimuth, tilt, and PV size

The fourth and final step estimates the optimal azimuth, tilt, and PV size of the customer's PV system using the magnitude of minimum net load and time-of-day when that minimum net load occurs.

This involves the following three sub-steps.

1. First, a PV system size is selected for each azimuth/tilt combination defined during the POAI generation process discussed in the third step. This involves using the base load estimated from net load in the second step.
2. Second, the time-of-day and magnitude of minimum net load are weighted and combined to produce a time-magnitude-based metric for net load. A similar process is undertaken for the time-of-day and magnitude of maximum POAI for the various azimuth/tilt combinations in consideration. Therefore, each azimuth/tilt combination considered in the POAI generation process yields a time-magnitude-based metric.
3. Third, the time-magnitude-based metric for net load and time-magnitude-based metric for POAI are compared and the best-fit optimal azimuth/tilt/PV size combination is identified.

5 Spec Inference from PV Production

CPR's Spec Inference approach for estimating the PV system specifications (azimuth, tilt, and PV size) from net load can be extended to infer the same three key system specifications using PV production data as well. The modifications to Spec Inference from net load involve the following three steps.

1. Find time-of-day and magnitude of maximum PV production.
2. Produce POAI using CPR's SolarAnywhere clear-sky GHI for various azimuth and tilt combinations.
3. Identify azimuth, tilt, and PV rating using the magnitude of maximum PV production and time-of-day when maximum PV production occurs.

The major difference between Spec Inference using Net Load and Spec Inference using PV production is that the latter approach does not require the calculation of base load.

6 Case Study of 8,000+ BTM SMUD PV Sites

This section discusses the PV system specifications (azimuth, tilt, and PV size) obtained for the 8,324 sites in the SMUD service territory. The system specifications are obtained using four different inference methods and metadata sources including 1) Spec Inference from Net Load, 2) Spec Inference from PV production, 3) reported in PowerClerk, and 4) Extended Spec Inference from PV production. The site locations are shown in the Google map snapshot in Figure 10.

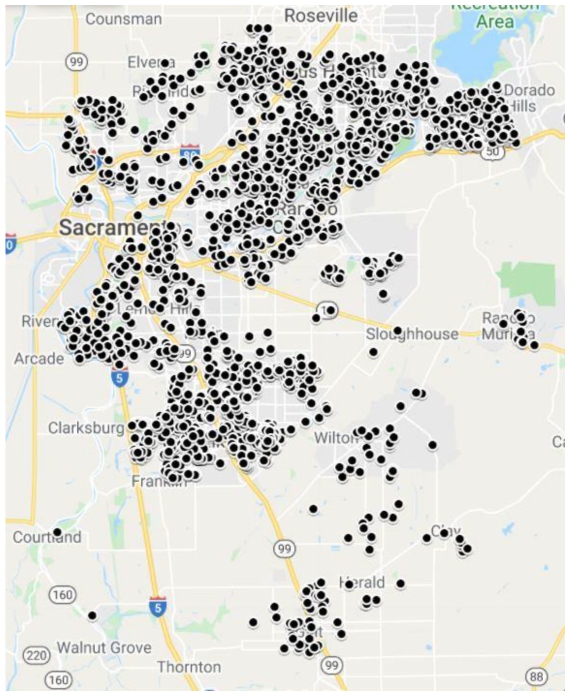


Figure 10: Site locations.

6.1 Individual Site

6.1.1 Comparing DC kW Ratings

DC kW ratings inferred by Extended Spec Inference from PV production are considered as the baseline for comparison with DC kW ratings inferred using the other methods and metadata sources, as shown in Equation 1. This is because the Extended Spec Inference estimates the key system specifications such as azimuth, tilt, and PV size with consideration of the shading losses incurred by the PV system. While the reported system specifications in PowerClerk also contain shading obstruction elevation angles in multiple azimuth directions, they do not model the real-world conditions of the PV system's surroundings that is reflected in the PV production data.

Figure 11 shows the scatter-plot comparison between the DC kW rating inferred by Extended Spec Inference from PV production and a) Spec Inference from Net Load, b) Spec Inference from PV production, and c) reported in PowerClerk (plots are shown for systems whose ratings are lower than 40 kW). Note that DC kW ratings from PowerClerk are reported for each array individually. Thus, the rating of a PV system is found by adding the ratings of the individual arrays. The relative Mean Absolute Error (rMAE) found using Equation 1 between the DC kW rating inferred by Extended Spec Inference from PV production and a) Spec Inference from Net Load is 13%, b) Spec Inference from PV production is 7%, and c) reported in PowerClerk is 14%. Among the three methods, the lowest error is obtained using Spec Inference with PV production. This is logical since Spec Inference and Extended Spec Inference both use PV production as input. DC kW ratings inferred by Spec Inference from net load and ones reported in PowerClerk have comparable rMAE's. Overall, the total DC kW rating of the fleet of 8,324 meters obtained from various methods and metadata sources are shown in Table 4.

Equation 1: rMAE between DC kW ratings.

$$rMAE = \frac{\sum |DC\ Rating_{Other\ Methods} - DC\ Rating_{Extended\ Spec\ Inference}|}{\sum DC\ Rating_{Extended\ Spec\ Inference}} \forall sites$$

Table 4: Total DC kW rating of 8,324 meters.

Method Name	Total DC kW Rating
Spec Inference from Net Load	42,549 kW
Spec Inference from PV	42,223 kW
PowerClerk	47,769 kW
Extended Spec Inference from PV	44,257 kW

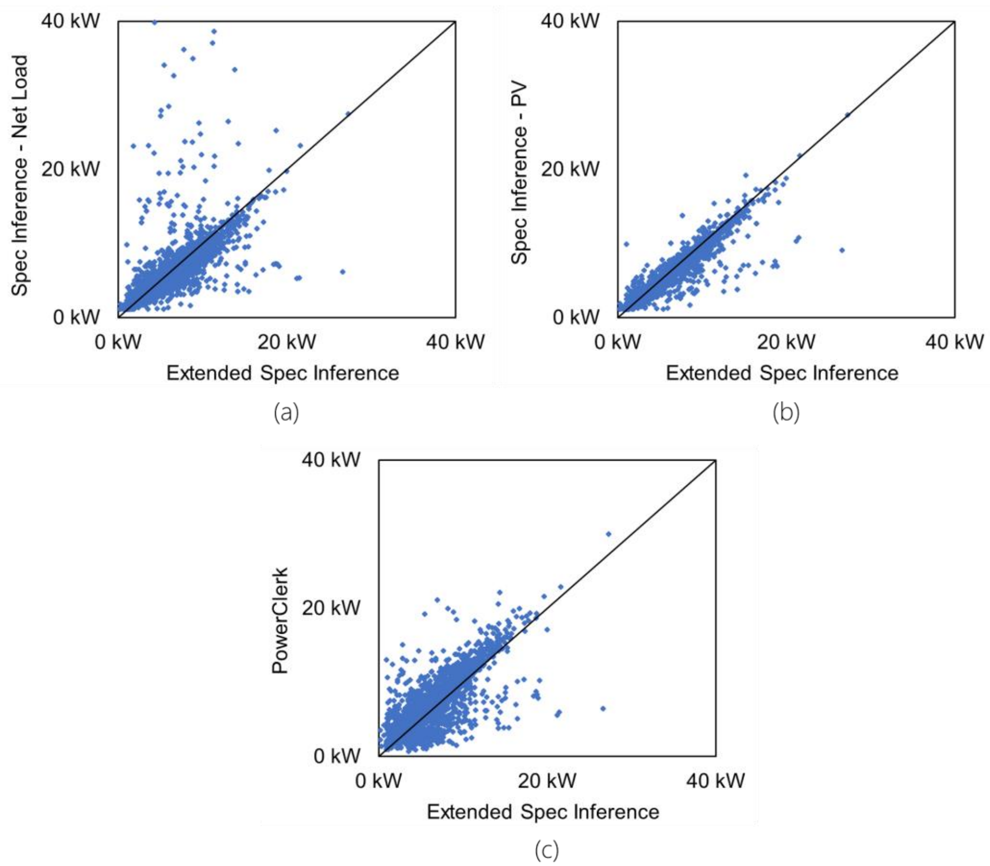
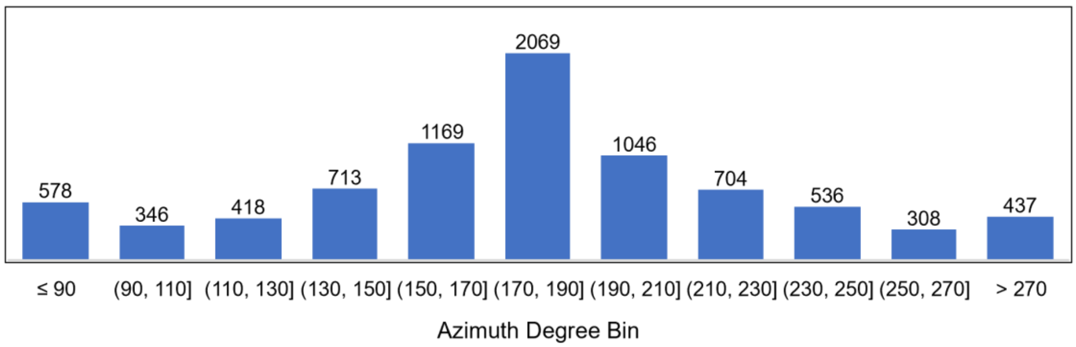


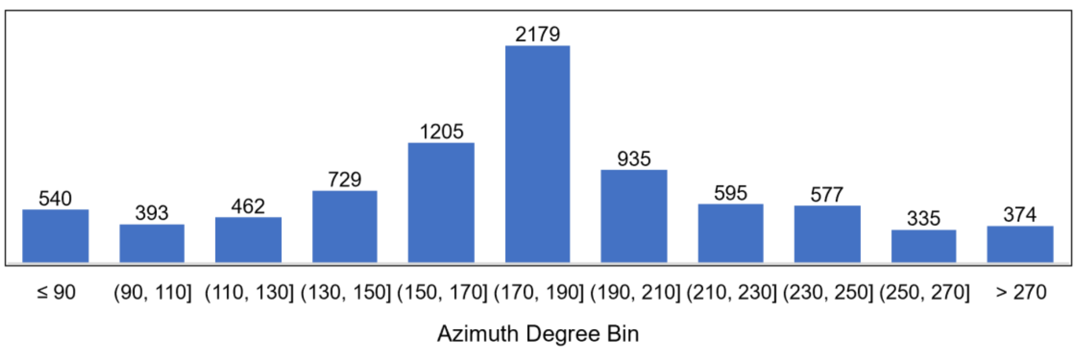
Figure 11: Comparison between DC kW ratings.

6.1.2 Most Commonly Occurring Azimuth

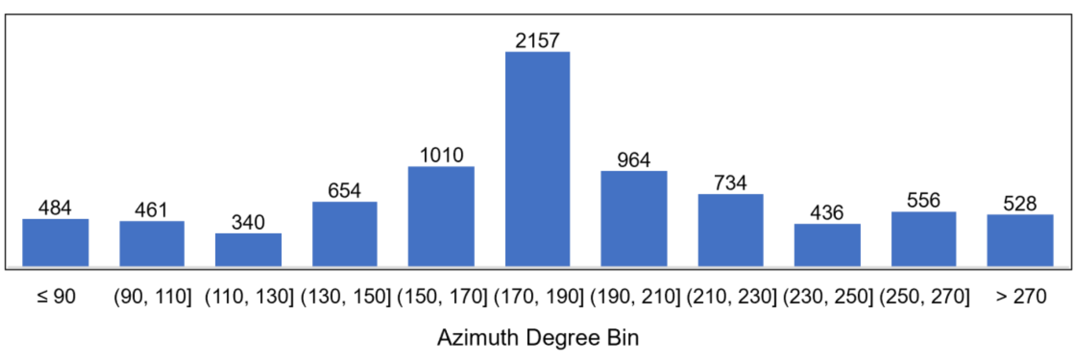
Distributions of azimuths inferred by a) Spec Inference from Net Load, b) Spec Inference from PV production, and c) Extended Spec Inference from PV production for the fleet of 8,324 sites are also examined. The distributions are shown in Figure 12a), Figure 12b), and Figure 12c), respectively. Distributions from PowerClerk are excluded since system specifications are reported for arrays of a PV system such that a single system may have multiple arrays and consequently, multiple azimuths.



(a)



(b)



(c)

Figure 12: Distribution of azimuths inferred by a) Spec Inference from Net Load, b) Spec Inference from PV production, and c) Extended Spec Inference from PV production.

The x-axes represent the 20-degree azimuth bins, and the bars represent the count of the systems that fall in the bin. Bin ≤90 represents azimuths less than or equal to 90 degrees and bin >270 degrees represent azimuths greater than 270 degrees. Note that 0-degree azimuth represents due north, 90-degree represents due east, 180-degree azimuth represents due south, and 270-degree represents due west. For each of the three different methods, we observe that the most common azimuth in the sites belongs to the 170° to 190° group. This suggests that orientation of most PV systems is approximately due south.

6.1.3 Most Commonly Occurring Tilt

Distributions of the tilts inferred by a) Spec Inference from Net Load, b) Spec Inference from PV production, and c) Extended Spec Inference from PV production for the fleet of 8,324 sites are also examined. The distributions are shown in Figure 13a), Figure 13b), and Figure 13c), respectively. Again, distributions of reported tilts from PowerClerk are excluded. For each of the three different methods, the most common tilt in the sites belongs to the 20° to 30° group, which is the approximate optimal tilt for the PV sites based on their locations in the SMUD service territory.

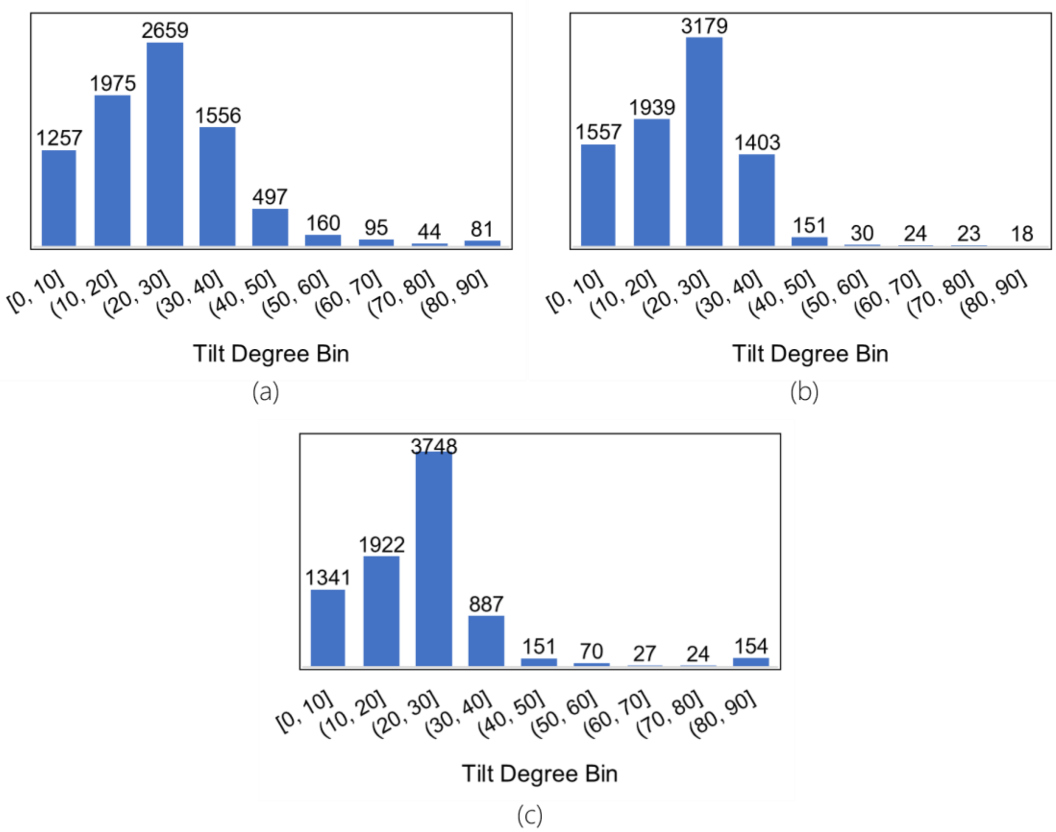


Figure 13: Distribution of tilts inferred by a) Spec Inference from Net Load, b) Spec Inference from PV production, and c) Extended Spec Inference from PV production.

6.2 Fleetwide

Fleetwide production for the 8,324 PV sites is estimated by simulating PV production for each site individually and then, summing the individual simulated PV productions. The simulated production is generated using CPR's SolarAnywhere software for the four different cases: 1) PV system specifications (azimuth, tilt, size) inferred by Spec Inference from Net Load, 2) PV system specifications (azimuth, tilt, size) inferred by Spec Inference from PV production, 3) PV system specifications reported in PowerClerk, and 4) PV system specifications from Extended Spec Inference from PV production. The simulated production is then compared with the measured PV production collected by the utility using the additional PV meters.

6.2.1 Fleetwide PV Production using Spec Inference from Net Load

In this method, PV system specifications (azimuth, tilt, and size) for each site are inferred using Spec Inference from Net Load. The four-step approach is discussed in Section 4. The inferred system specifications are then used along with CPR's SolarAnywhere software to generate simulated production for each site. Finally, the simulated production for the fleet of systems is estimated by summing all individual simulated productions. Similarly, the measured PV production of the fleet of systems is also estimated by summing up the individual PV productions. Next, we describe in detail the results obtained for: 1) four selected days in the spring, summer, fall, and winter seasons, 2) hourly results for the entire year, and 3) daily results for the entire year. Note that relative Mean Absolute Error (rMAE) between the measured production $P_{measured}(t)$ and simulated production $P_{simulated}(t)$ over time t is estimated using Equation 2.[4]

Equation 2: rMAE between measured and simulated production.

$$rMAE = \frac{\sum |P_{measured}(t) - P_{simulated}(t)|}{\sum P_{measured}(t)} \forall t$$

6.2.1.1 Simulation Results for One Day

Figure 14 illustrates the results for four selected days through different seasons. Figure 14a) shows the result for a day in the spring season: March 22, 2016. The hourly relative Mean Absolute Error (rMAE) between the measured PV production and simulated PV production for the fleet of 8,324 sites for the day is 5.4%. Similarly, Figure 14b), Figure 14c), and Figure 14d) show the results for a day in the summer season (June 22, 2016), fall season (September 22, 2016), and winter season (December 22, 2016), respectively. The hourly rMAE's for the selected days are 5.8%, 13.1%, and 26.3%, respectively. The hourly rMAE for the day in the winter season is the highest compared to days in other seasons. Further, in Figure 14d), for December 22, 2016, we observe that the simulated production is over-estimated with respect to the PV production. This is likely due to the lack of consideration of shading losses which have a prominent loss effect on PV production during the winter season, when the sun is mostly lower in the sky.

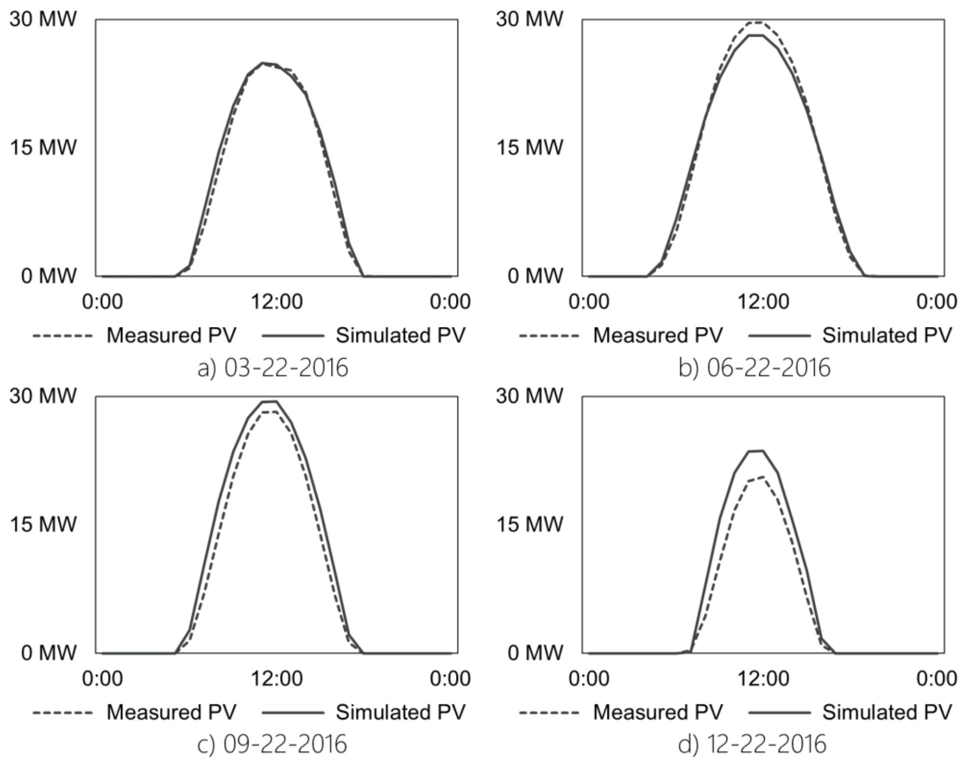


Figure 14: Fleetwide hourly production of selected days.

6.2.1.2 Hourly Simulation Results for Entire Year

Next, production is estimated for the fleet of 8,324 sites for the entire year. Simulated and measured production are compared on a per hour basis, as shown in Figure 15. The hourly rMAE between the two is 10.5%. From Figure 15, we observe that the simulated production is slightly higher compared to the corresponding measured PV production. This is likely due to the lack of consideration of shading and soiling losses that reduce the production of the system.

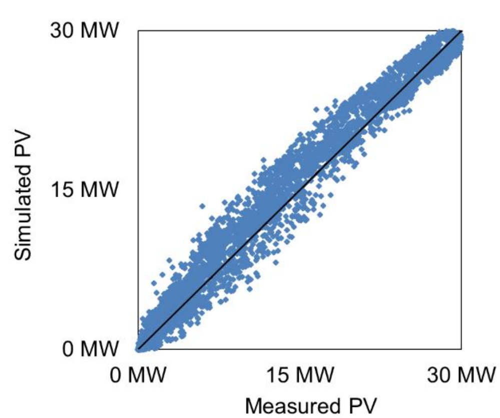


Figure 15: Fleetwide hourly production for 2016.

6.2.1.3 Daily Simulation Results for Entire Year

Finally, hourly production for the fleet of 8,324 sites for the entire year is aggregated on a per day basis and the per day simulated production is compared with the per day measured PV production. The daily rMAE between the simulated and measured PV production for the entire year is 8.1%, shown in Figure 16. Again, we attribute the overestimated simulated production to the lack of consideration of shading and soiling losses.

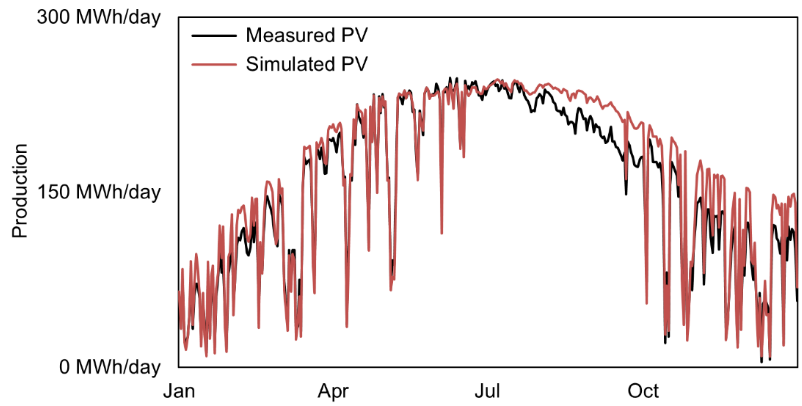


Figure 16: Fleetwide daily production for 2016.

6.2.2 Fleetwide Shaded PV Production using Spec Inference from Net Load

As discussed in the previous Section 6.2.1, the lack of consideration of shading and soiling losses incurred by the PV system overestimates the simulated production of the PV systems compared to its corresponding measured production. In this section, shading losses are accounted for by employing a data-driven method of estimating a constant horizon shading elevation angle using production data from a small number of sites.

While PV production data may not be available for most electric utilities, for the purposes of this case study, we assume that measured production is available for a small number of sites. We also assume that the small number of sites are representative of the entire territory-wide shading obstructions. To estimate the optimal constant horizon shading elevation angle, PV production is simulated using system specifications inferred by Spec Inference from Net Load in Section 6.2.1 along with various values of the additional parameter i.e., constant horizon shading elevation angle. We compare the hourly simulated production with the hourly measured production for the fleet of considered systems. Finally, the constant horizon shading elevation angle that yields the lowest hourly rMAE is assumed to be the optimal shading elevation angle.

We used 100 sites in the SMUD service territory for estimating the constant horizon shading elevation angle and found that the optimal shading elevation angle was 10°, as shown in Figure 17. Figure 17a) shows the hourly rMAE for the entire year (2016) between the simulated production and measured production for the 100 sites in increments of 10 degrees and Figure 17b) shows the hourly rMAE between the simulated production and measured production for the 100 sites in increments of 1 degree. In the upcoming sections, we will discuss the results of considering shading losses incurred by the fleet of 8,324 PV systems with a constant horizon shading elevation angle of 10°.

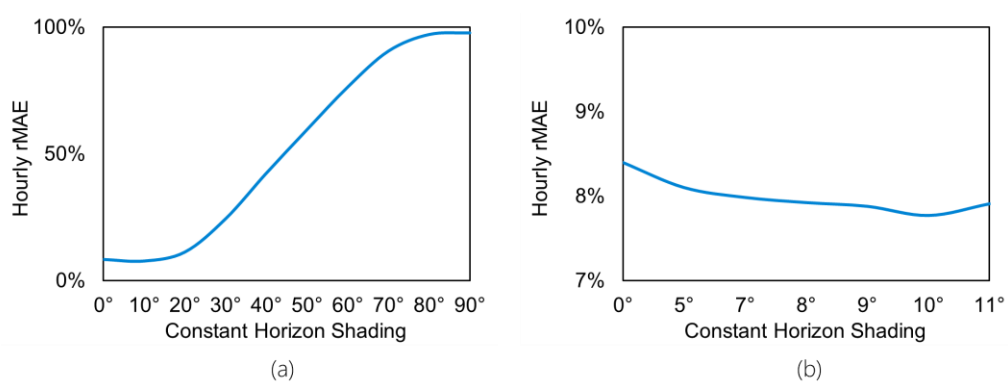


Figure 17: Hourly rMAE for the entire year (2016) between measured and simulated production for various constant horizon shading elevation angles. a) In increments of 10 degrees, b) In increments of 1 degree.

6.2.2.1 Simulation Results for One Day

Figure 18a), Figure 18b), Figure 18c), and Figure 18d) show the PV production for a day in the spring (March 22, 2016), summer (June 22, 2016), fall (September 22, 2016), and winter (December 22, 2016) seasons, respectively. The hourly rMAE between the measured production and simulated production for the four days for the fleet of 8,324 sites is 5.0%, 5.9%, 11.5%, and 19.6%, respectively. Compared to the hourly rMAE of 5.4%, 5.8%, 13.1%, and 26.3% obtained for the same four days for spring, summer, fall, and winter seasons, respectively without the inclusion of shading losses (in Section 6.2.1.1), there is a significant decrease in hourly rMAE for the days in the fall and winter seasons with the inclusion of shading losses. This result proves that approximating constant horizon shading elevation angles from a small number of sites (100 sites) is still effective in significantly reducing the hourly rMAE for the fleet of systems (8,324 sites) for days in the fall and winter seasons.

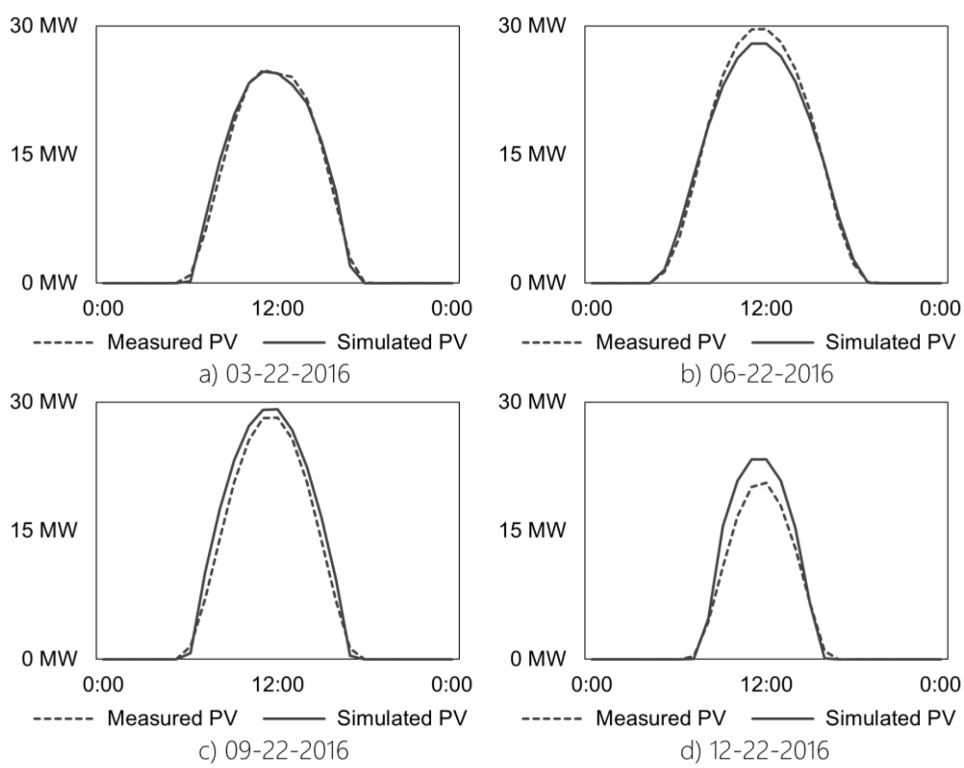


Figure 18: Fleetwide hourly production of selected days.

6.2.2.2 Hourly Simulation Results for Entire Year

Figure 19 shows a scatter plot comparison of the simulated PV production with shading losses and the measured PV production with an hourly rMAE of 9.4%. This is a 10.5% decrease from the hourly rMAE (of 10.5%) in Section 6.2.1.2, where shading losses were excluded. Significant improvement is observed for days in the fall and winter seasons, where the effect of shading is much more prominent due to the lower position of the sun compared to days in the spring and summer seasons.

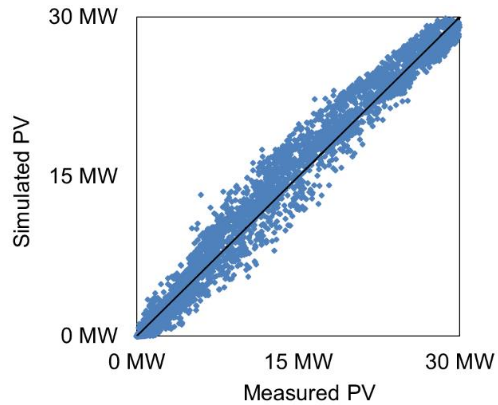


Figure 19: Fleetwide hourly production for 2016.

6.2.2.3 Daily Simulation Results for Entire Year

The daily rMAE between the simulated production with shading losses and measured PV production for the entire year is 6.3%, shown in Figure 20. This is a 22% decrease from the daily rMAE of 8.1% obtained in Section 6.2.1.3, when shading losses are not considered.

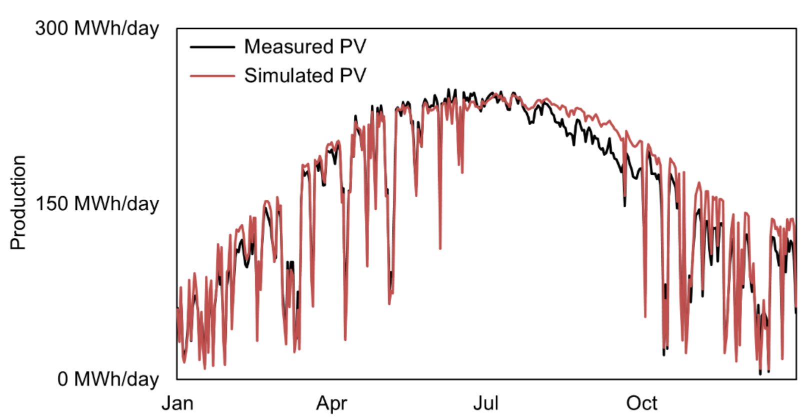


Figure 20: Fleetwide daily production for 2016.

6.2.3 Fleetwide PV Production using Spec Inference from PV Production

While the Spec Inference approach discussed in Section 4 is used for inferring PV system specifications using net load data, the same approach with modifications can also be used for inferring system specifications using PV production data. These modifications have been discussed in Section 5 and applied to PV production data measured using the additional PV meters at the sites. The inferred PV system specifications include azimuth, tilt, and system size which are used along with CPR's SolarAnywhere simulation software to generate simulated production for each site. Finally, simulated production for the fleet of systems is estimated by summing all individual simulated productions. A similar approach is taken to estimate the fleet measured production. Next, the results obtained for: 1) four selected days in the spring, summer, fall, and winter seasons, 2) hourly results for the entire year, and 3) daily results for the entire year are discussed in detail.

6.2.3.1 Simulation Results for One Day

Figure 21a), Figure 21b), Figure 21c), and Figure 21d) show the PV production for a day in the spring (March 22, 2016), summer (June 22, 2016), fall (September 22, 2016), and winter (December 22, 2016) seasons, respectively. Hourly rMAE between measured production and simulated production for the four days for the fleet of 8,324 sites are 6.0%, 5.2%, 13.0%, and 23.3%, respectively. These rMAE's are comparable to that obtained from system specifications inferred using the net load data (in Section 6.2.1.1).

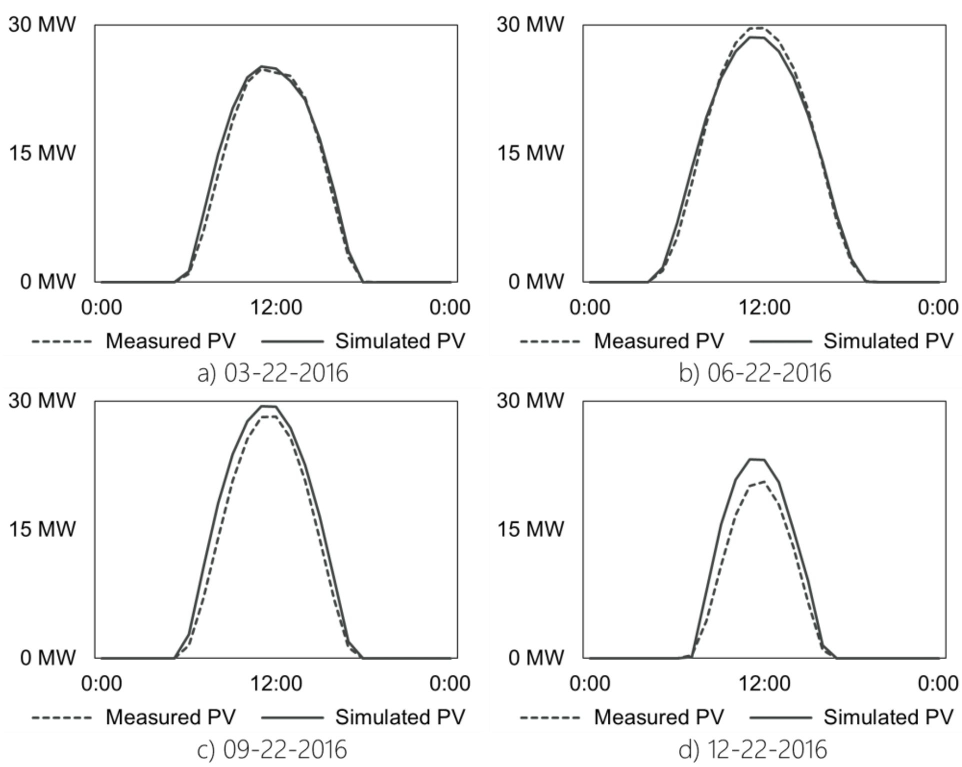


Figure 21: Fleetwide hourly production of selected days.

Furthermore, the hourly rMAE for the day in the winter season is the highest compared to days in other seasons. This is also in line with the results obtained from system specifications inferred from net load data without considering shading losses, discussed in Section 6.2.1.

6.2.3.2 Hourly Simulation Results for Entire Year

Figure 22 shows a scatter plot comparison of the simulated PV production and the measured PV production for the fleet of 8,324 sites for the entire year of 2016. The hourly rMAE between the two is 10.3%. Similar to the scatter plot for the specifications inferred by Spec Inference from Net Load without consideration of shading losses, shown in Figure 15, we observe that simulated production is slightly higher compared to the corresponding measured production.

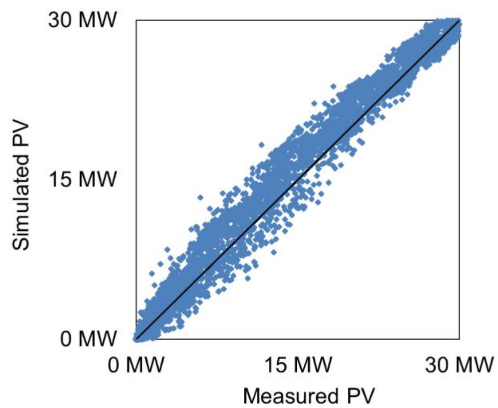


Figure 22: Fleetwide daily production for 2016.

6.2.3.3 Daily Simulation Results for Entire Year

The daily rMAE, between the simulated and measured PV production for the fleet of 8,324 sites, shown in Figure 23, for the entire year of 2016 is 8.1%.

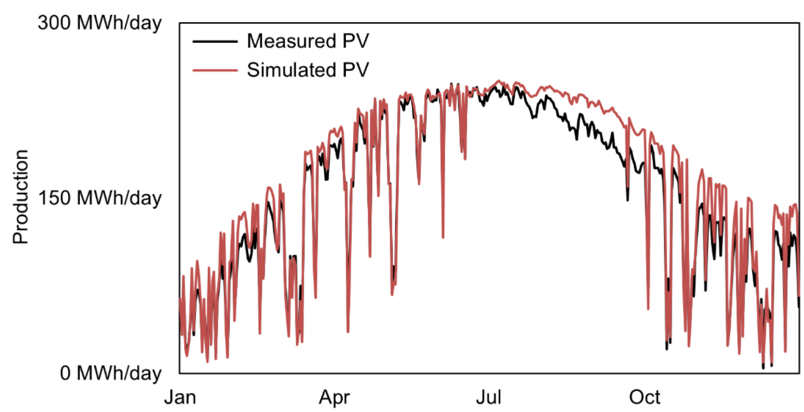


Figure 23: Fleetwide daily production for 2016.

6.2.4 Fleetwide Shaded PV Production using Spec Inference from PV Production

6.2.4.1 Simulation Results for One Day

Figure 24a), Figure 24b), Figure 24c), and Figure 24d) show the PV production for a day in the spring (March 22, 2016), summer (June 22, 2016), fall (September 22, 2016), and winter (December 22, 2016) seasons, respectively and the hourly relative Mean Absolute Error (rMAE) between the measured production and simulated production with the inclusion of shading losses for the four days for the fleet of 8,324 sites are 5.6%, 5.2%, 13.0%, 17.8%, respectively. These rMAE's are comparable to that obtained from system specifications inferred using the net load data with the inclusion of shading losses discussed in Section 0.

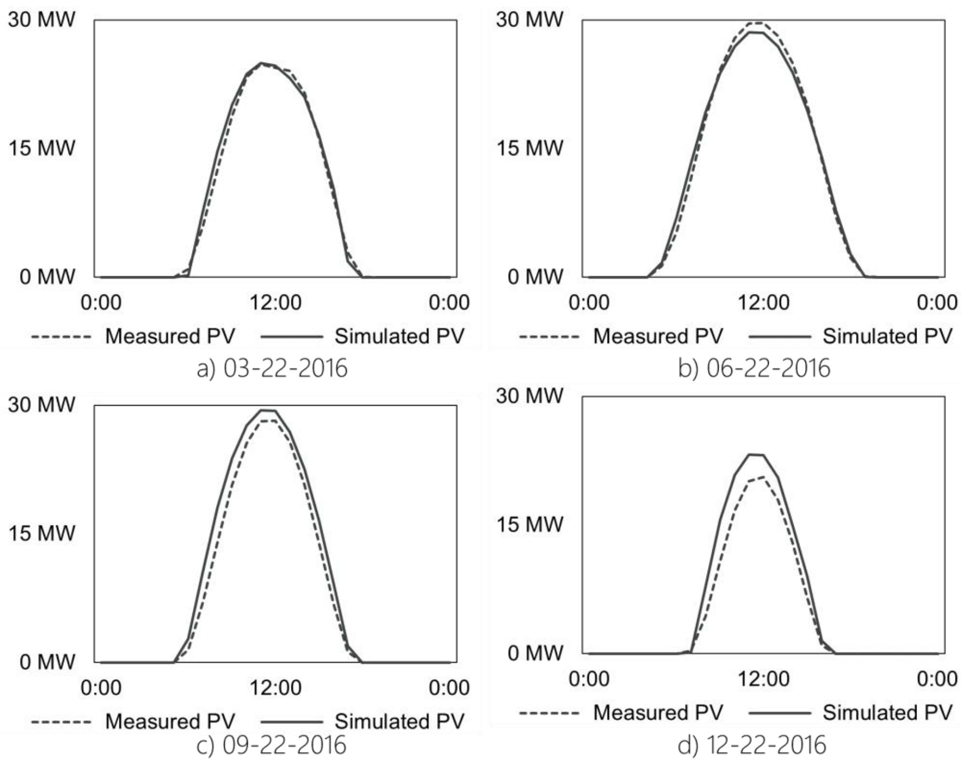


Figure 24: Fleetwide hourly production of selected days.

6.2.4.2 Hourly Simulation Results for Entire Year

Figure 25 shows a scatter plot comparison of the simulated PV production with the inclusion of shading losses and the measured PV production for the fleet of 8,324 sites. The hourly rMAE between the two is 9.2%.

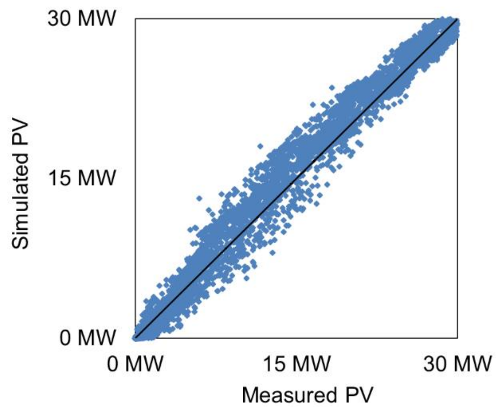


Figure 25: Fleetwide hourly production for 2016.

6.2.4.3 Daily Simulation Results for Entire Year

The daily rMAE, between the simulated PV production with the inclusion of shading losses and measured PV production for the fleet of 8,324 sites, shown in Figure 26, for the entire year of 2016 is 6.2%.

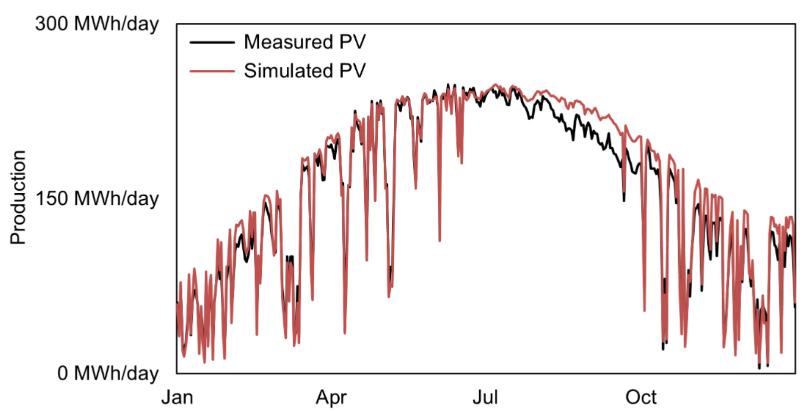


Figure 26: Fleetwide daily production for 2016.

6.2.5 Fleetwide Shaded PV Production using PowerClerk

Installer/user reported PV system specifications are accessed through PowerClerk, CPR's software tool that facilitates utilities with capabilities to effectively manage the Distributed Energy Resources' (DER) interconnection and incentive applications.[3] The system specifications of each site are collected by the utility. These include detailed system specifications for individual PV arrays that create a PV system as well as detailed system specifications for individual inverters. The PV system parameters included in the reported system specifications for the PV arrays are as follows:

1. Manufacturer
2. Model
3. Quantity
4. Nameplate DC rating
5. PTC rating
6. Azimuth
7. Tilt
8. Tracking
9. East shading obstruction
10. East SouthEast shading obstruction
11. South SouthEast shading obstruction
12. South shading obstruction
13. South SouthWest shading obstruction
14. West SouthWest shading obstruction
15. West shading obstruction

The system parameters included in the reported system specifications for the inverters are as follows:

1. Manufacturer
2. Model
3. Quantity
4. Nameplate DC rating
5. Inverter efficiency

These system specifications are used as input to CPR's SolarAnywhere simulation software to generate simulated production for each site. And the simulated production for the fleet of systems is estimated. Next, the results obtained for: 1) four selected days in the spring, summer, fall, and winter seasons, 2) hourly results for the entire year, and 3) daily results for the entire year are discussed in detail.

6.2.5.1 Simulation Results for One Day

Figure 27a), Figure 27b), Figure 27c), and Figure 27d) show the PV production for a day in the spring (March 22, 2016), summer (June 22, 2016), fall (September 22, 2016), and winter (December 22, 2016) seasons, respectively and the hourly relative Mean Absolute Error (rMAE) between the measured production and simulated production for the fleet of 8,324 sites is 1.8%, 3.0%, 10.0%, and 4.1%, respectively. In contrast to the rMAE's obtained by Spec Inference from Net Load and Spec Inference from PV production for the four selected days, the rMAE's obtained using the reported system specifications from PowerClerk yield much lower error. Since the reported system specifications from PowerClerk include detailed system specifications for each array as well as shading obstruction information in multiple azimuth directions, the resulting rMAE's for the selected days are lower.

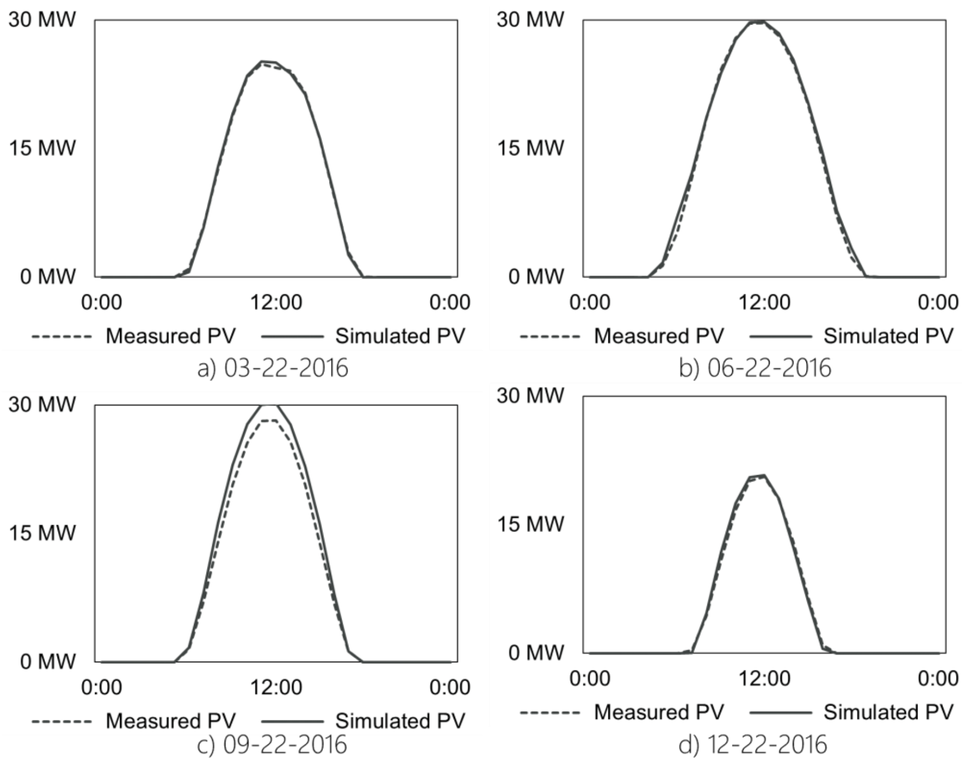


Figure 27: Fleetwide hourly production of selected days.

6.2.5.2 Hourly Simulation Results for Entire Year

The hourly rMAE between the simulated PV production and the measured PV production for the fleet of 8324 sites for the entire year, shown in Figure 28 is 7.5%. This is lower than that of 9.4% and 9.2%

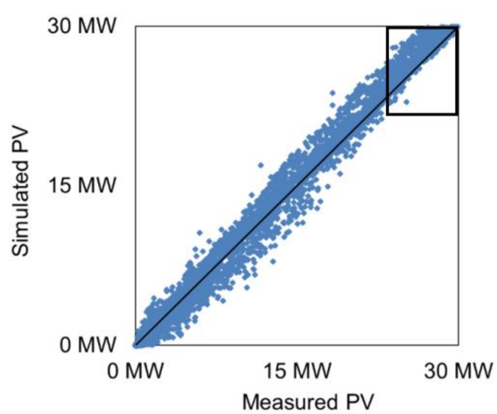


Figure 28: Fleetwide hourly production for 2016.

obtained by Spec Inference from Net Load and Spec Inference from PV production with the inclusion of shading losses, shown in Section 6.2.2.2 and Section 6.2.4.2, respectively. While this reduction in hourly rMAE is significant compared to the other two methods, we observe overestimated simulated production compared to its corresponding measured production for larger fleet productions. For example, from Figure 28, we observe that the simulated production is overestimated compared to the measured production for productions above 25 MWh/hr (shown by the inset black box). To further investigate the overestimated simulated productions compared to the corresponding measured production, we separate the hourly productions of the entire year into the four seasons: spring (March 19 – June 19), summer (June 20 – September 21), fall (September 22 – December 20), and winter (December 21 – March 18), as shown in Figure 29, where the hourly rMAE's are 5.1%, 8.7%, 8.7%, and 8.3%, respectively. As seen from Figure 29, during the spring, summer, and fall seasons, simulated productions above 25 MWh/hr are over-estimated compared to their corresponding measured productions. In contrast, during the winter season, the simulated productions lower than 15 MWh/hr are underestimated compared to their corresponding measured productions. We also observe that most simulated productions during the summer season are overestimated compared to the measured productions. Since the shading obstructions in multiple azimuth directions are already included in this analysis, we postulate that the sites may be heavily affected by other external parameters. Particularly, during the summer and fall seasons, the production of these sites may have been affected by soiling (due to less or no rainfall in these sites) and as such, including soiling losses would likely reduce the over-estimated simulated productions.

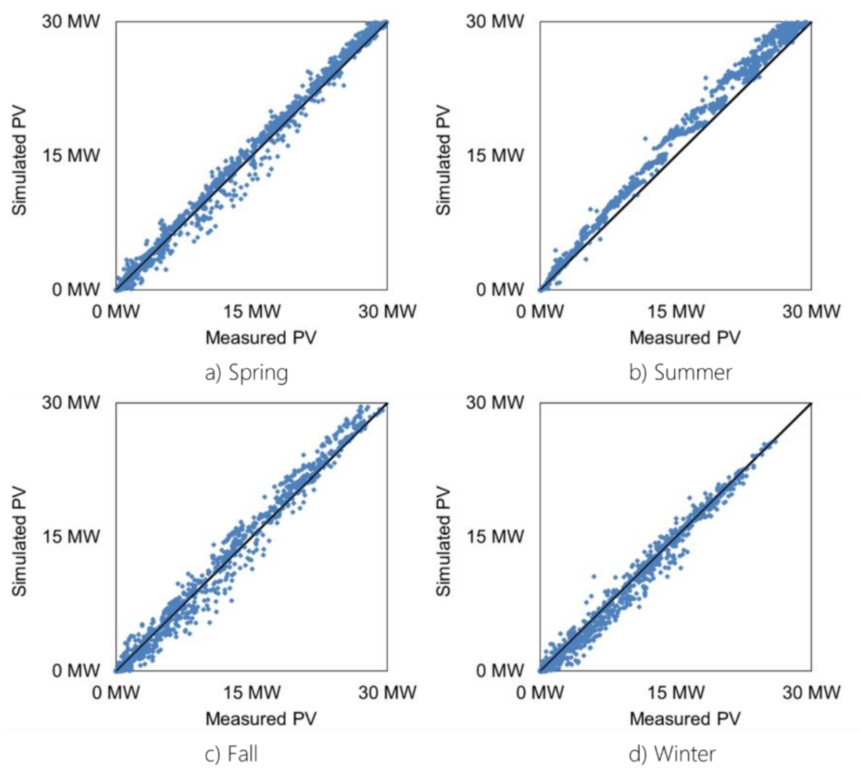


Figure 29: Fleetwide hourly production for a) Spring, b) Summer, c) Fall, and d) Winter for the fleet of 8,324 sites in 2016.

6.2.5.3 Daily Simulation Results for Entire Year

The daily rMAE between the simulated and measured PV production, shown in Figure 30, for the fleet of 8,324 sites for the entire year is 6.3%. As discussed in the previous Section 6.2.5.2, the over-estimated simulated productions (shown by the red line in Figure 30) mostly occur during the summer and fall seasons and can be attributed to less or no rainfall events that accumulate soil in the PV panels during these seasons at the specific SMUD sites.

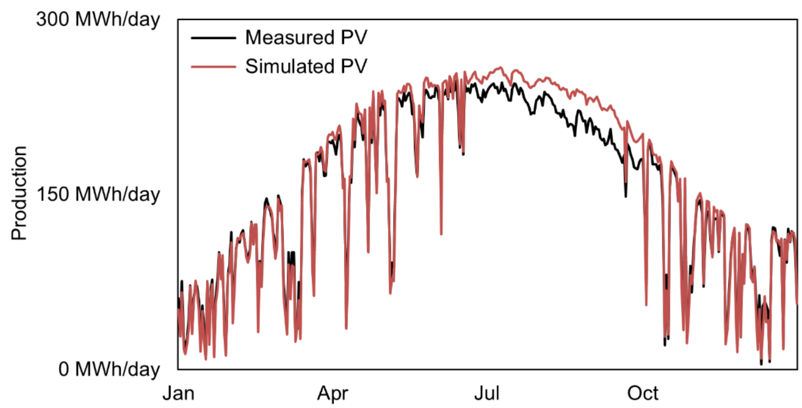


Figure 30: Fleetwide daily production for 2016.

6.2.6 Fleetwide Shaded PV Production using Extended Spec Inference from PV Production

The final method of comparing simulated production with measured production is by using the system specifications inferred by Extended Spec Inference from PV production along with CPR's SolarAnywhere simulation software. The inferred PV system specifications are as follows:

1. Azimuth
2. Tilt
3. PTC rating
4. Tracking
5. East shading obstruction
6. East SouthEast shading obstruction
7. South SouthEast shading obstruction
8. South shading obstruction
9. South SouthWest shading obstruction
10. West SouthWest shading obstruction
11. West shading obstruction

These system specifications are used as input to CPR's SolarAnywhere simulation software to generate simulated production for each site. Then, simulated production for the fleet of systems is estimated. Next, the results obtained for: 1) four selected days in the spring, summer, fall, and winter seasons, 2) hourly results for the entire year, and 3) daily results for the entire year are discussed in detail.

6.2.6.1 Simulation Results for One Day

Figure 31a), Figure 31b), Figure 31c), and Figure 31d) show the PV production for a day in the spring (March 22, 2016), summer (June 22, 2016), fall (September 22, 2016), and winter (December 22, 2016) seasons, respectively and the hourly rMAE between the measured production and simulated production for the fleet of 8324 sites are 2.3%, 3.6%, 6.5%, and 5.5%, respectively. This method performs the best among all four methods: Spec Inference from Net Load, Spec Inference from PV production, and reported system specifications in PowerClerk. These three methods have been discussed extensively in Sections 6.2.1.1, 6.2.3.1, and 6.2.5.1, respectively. Reported system specifications obtained from PowerClerk include detailed system specifications for each PV array and inverter. These specifications along with accurate irradiance data from SolarAnywhere are used to simulate production. However, these specifications do not account for real-world conditions (such as soiling losses) under which the PV system operates. The higher accuracy obtained by using Extended Spec Inference from PV production can be attributed to the use of measured production data, which is encoded with information about all physical conditions under which the PV system operates.

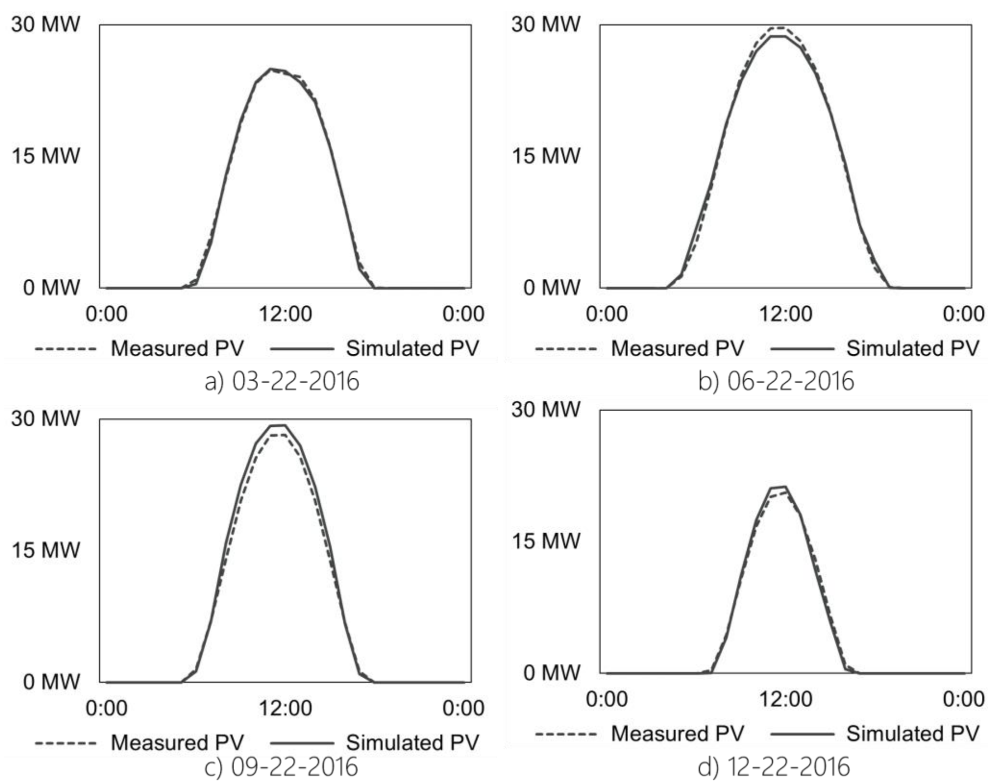


Figure 31: Fleetwide hourly production of selected days.

6.2.6.2 Hourly Simulation Results for Entire Year

The hourly rMAE between the simulated PV production and the measured PV production for the fleet of 8,324 sites, shown in Figure 32 is 6.0%, which is the lowest compared to that of 9.4%, 9.2%, and 7.5% obtained by Spec Inference from Net Load including shading losses, Spec Inference from PV production including shading losses, and PowerClerk in Sections 6.2.2.2, 6.2.4.2, and 6.2.5.2, respectively. The error in the fleet output is approximately equal to the error in the underlying satellite derived irradiance data.[9]

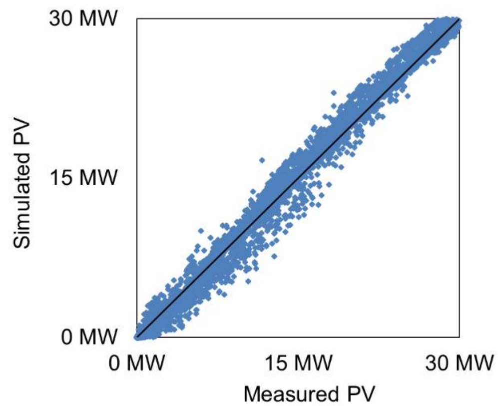


Figure 32: Fleetwide hourly production for 2016.

6.2.6.3 Daily Simulation Results for Entire Year

The daily rMAE between the simulated and measured PV production for the fleet of 8,324 sites, shown in Figure 33, for the entire year is 4.2%. As discussed in the previous sections, the over-estimated simulated productions (shown by the red line in Figure 33) mostly occur during the summer and fall seasons and can be attributed to less or no rainfall events that lead to soiling losses.

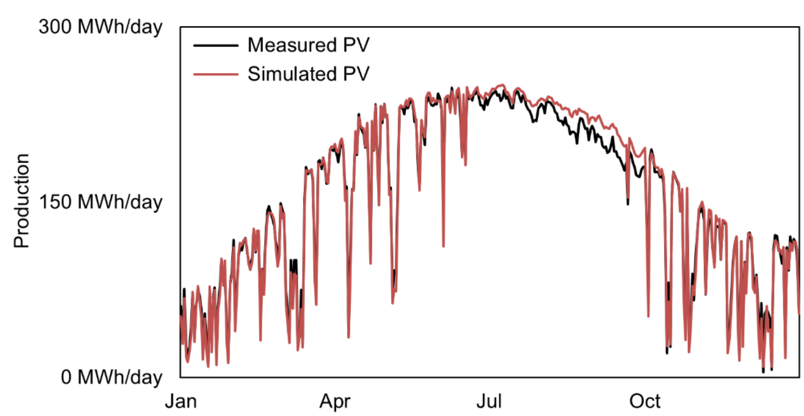


Figure 33: Fleetwide daily production for 2016.

[9] https://www.solaranywhere.com/resources/data-validation/

6.2.7 Soiling Losses

The daily fleet production, shown in Figure 33 of Section 6.2.6.3 is shown again in Figure 34. The plot shows the comparison between the measured and simulated production for the fleet of 8,324 sites, where the simulated production is generated using system specifications obtained from Extended Spec Inference from PV production (along with shading losses). The rMAE between the simulated and measured production is 4.2%. We observed that the simulated production is overestimated compared to the measured production for parts of the summer and fall seasons shown by the inset black box in Figure 34. We attribute this to soiling losses experienced by the sites in the Sacramento, CA location. These sites experience long dry seasons due to no rainfall during parts of summer and fall seasons. We assume that the no rain period occurs from May 20, 2016, to October 2, 2016. Then, we estimate the per day soiling loss of the fleet of 8,324 sites by finding a per day soiling loss factor that reduces the daily rMAE between the measured and simulated production. This factor is shown in Figure 35, where the per day soiling loss is estimated to be around 0.06%. Figure 36 shows the comparison between the measured production and updated simulated production with soiling losses of 0.06% per day. The daily rMAE between the measured and simulated production with soiling losses is 2.9%, which is a 31% reduction from the daily rMAE of 4.2% without soiling losses.

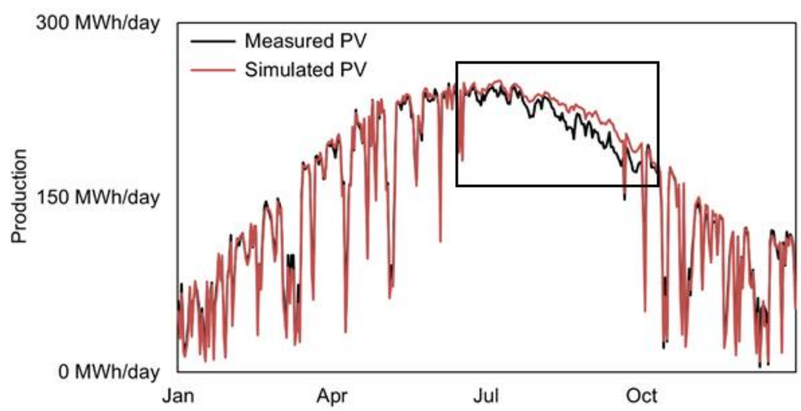


Figure 34: Fleetwide daily production for 2016.

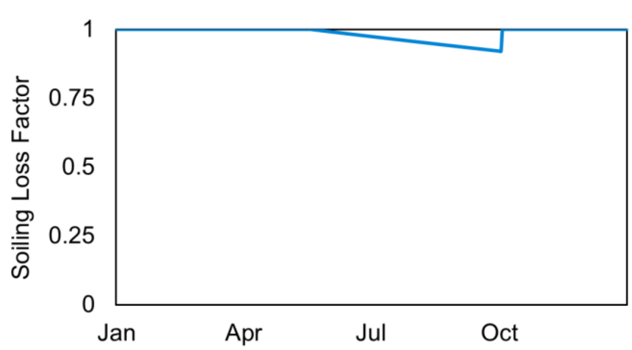


Figure 35: Soiling loss factor during 2016 for the fleet of 8324 sites.

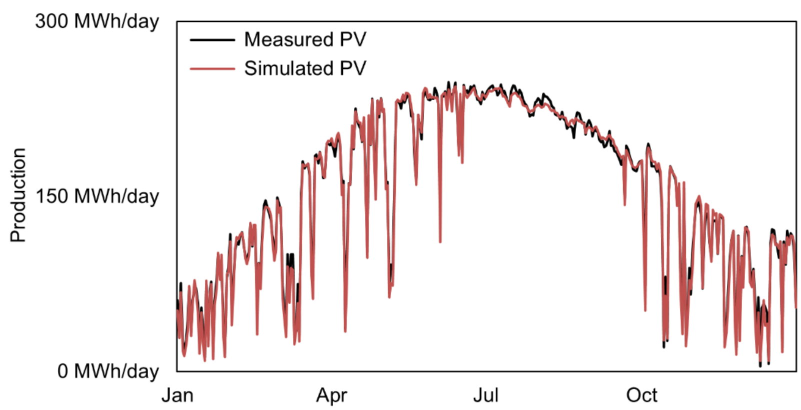


Figure 36: Fleetwide daily production including shading losses during 2016.

Figure 37 shows the monthly simulated production of Figure 36 (represented by the blue bars) with the corresponding soiling losses (represented by the additional black bars on top). The black bars show the amount of production lost each month due to soiling, which is estimated by subtracting the monthly simulated productions with and without soiling. Overall, for 2016, the fleet of 8,324 sites lost 2.1% of production due to soiling. The highest amount of soiling loss occurred during the month of September due to cumulative losses starting from May 20 (due to no rainfall). The rainfall event on October 2 is assumed to have cleaned the panels and reset the production to the maximum.

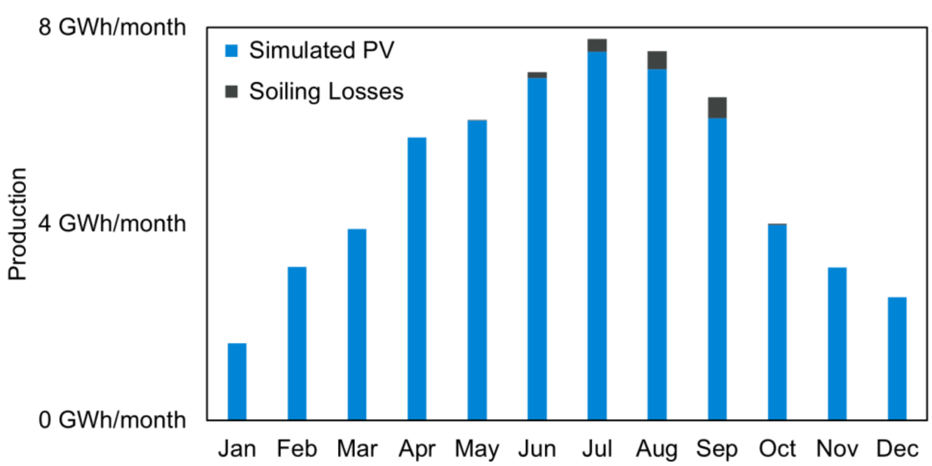


Figure 37: Fleetwide monthly production and soiling losses during 2016.

For 2016, the hourly and daily rMAE between the measured and simulated production obtained for the fleet of 8,324 systems after applying soiling losses to simulated production obtained using system specifications inferred by 1) Spec Inference from Net Load including shading losses, 2) Spec Inference from PV production including shading losses, 3) PowerClerk with shading losses, and 4) Extended Spec Inference from PV production including shading losses are shown in Table 5.

Table 5: Hourly and daily rMAE between measured and simulated production during 2016 for the fleet of 8,324 sites with the inclusion of shading and soiling losses.

Method Name	Hourly rMAE	Daily rMAE
Spec Inference from Net Load	9.0%	5.4%
Spec Inference from PV	8.6%	4.9%
PowerClerk	5.6%	4.2%
Extended Spec Inference from PV	5.1%	2.9%

6.2.8 Summary of Fleetwide PV Production

Figure 38 shows the hourly and daily rMAE's between the measured and simulated production during the entire year (2016) for the fleet of 8,324 PV sites using the four different methods and metadata sources. The figure also compares the hourly and daily rMAE between the measured and simulated production for three additional cases when: 1) shading and soiling losses are not considered, 2) only shading losses are considered, and 3) both shading and soiling losses are considered. For the purposes of this study, we consider soiling losses as the PV sites are located in Sacramento, CA, where rainfall is limited. The blue bars represent hourly and daily rMAE for system specifications inferred by Spec Inference from Net Load and Spec Inference from PV production, when shading and soiling losses are not considered. The yellow bars, for the Spec Inference from Net Load and Spec Inference from PV production, represent the rMAE when a constant horizon shading elevation angle of 10degrees is considered while generating the simulated production. The yellow bars, for the PowerClerk and Extended Spec Inference, represent the rMAE when shading in multiple azimuth directions is considered. For PowerClerk, shading elevation angles in the multiple azimuthal directions are reported for each array separately. For the Extended Spec Inference, shading elevation angles in the multiple azimuthal directions are inferred for the overall system only. For all four cases, the black bars represent the rMAE when shading and soiling are both considered for generating simulated production. Table 6 shows the hourly and daily rMAE obtained between the measured and simulated production during 2016 for the fleet of 8,324 PV sites.

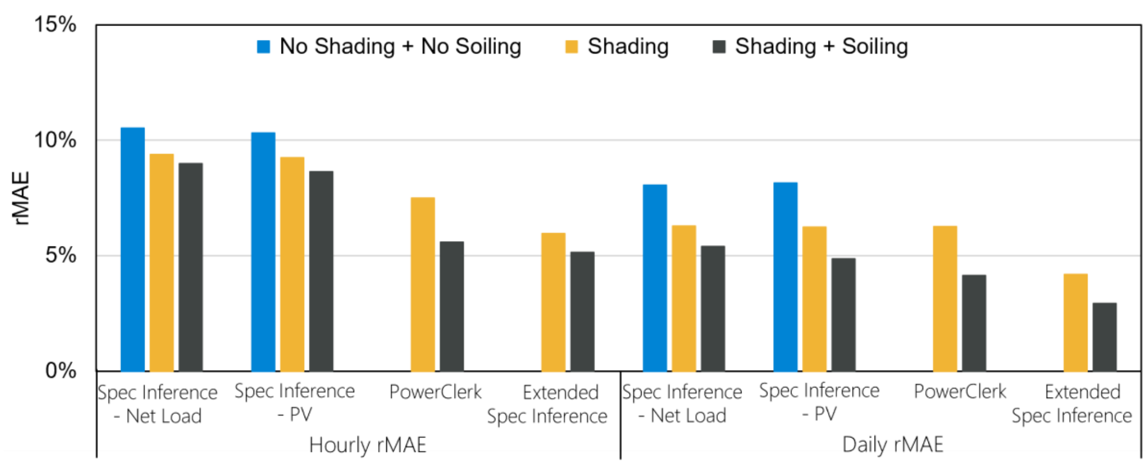


Figure 38: Hourly and daily rMAE between measured production and simulated production during 2016 for the fleet of 8,324 sites using system specifications: 1) inferred by Spec Inference from Net Load, 2) inferred by Spec Inference from PV production, 3) reported in PowerClerk, and 4) inferred by Extended Spec Inference from PV production.

Table 6: rMAE between measured and simulated production during 2016 for the fleet of 8,324 sites.

Method Name	No shading and No Soiling		Shading		Shading and Soiling	
	Hourly rMAE	Daily rMAE	Hourly rMAE	Daily rMAE	Hourly rMAE	Daily rMAE
Spec Inference from Net Load	10.5%	8.1%	9.4%	6.3%	9.0%	5.4%
Spec Inference from PV	10.3%	8.1%	9.1%	6.2%	8.6%	4.9%
PowerClerk			7.5%	6.3%	5.6%	4.2%
Extended Spec Inference from PV			6.0%	4.2%	5.1%	2.9%

As expected, consideration of shading and soiling losses yields the lowest error in each of the four different methods and metadata sources. Three key observations from Figure 38 and Table 6 are:

1. Among the four different inference methods and metadata sources, system specifications inferred using Extended Spec Inference from PV production data yields the lowest error. This is because the production data is encoded with the real-world conditions under which the PV system operates. Further, Extended Spec Inference considers shading elevation angles in multiple azimuth directions. This contributes to the overall lower error rates. The system specifications obtained from Extended Spec Inference may not be the exact specifications of the PV system. For example, a site with a PV system may have multiple arrays oriented in different directions. One array may face towards the east while another may face towards the west. In such cases, Extended Spec Inference does not infer system specifications for each array separately. It only infers one set of specifications for a site. Regardless of this, these system specifications are the best among the above-mentioned methods. Thus, we recommend using system specifications obtained from Extended Spec Inference if PV production data is available.

2. The system specifications from PowerClerk have lower errors compared to the system specifications inferred by Spec Inference from Net Load or Spec Inference from PV production. We attribute this to the accuracy and detail of the reported system specifications for the SMUD sites, which are collected and reported for each array in a PV system separately. This may not be the case for other electric utilities. Generally, utilities compensate for the lack of PV system metadata by making assumptions based on location of the PV system. For example, PV systems in the Northern hemisphere are assumed to have an azimuth due south and tilt approximately equal to the latitude of the site. Therefore, the reported system specifications may be inaccurate and unreliable.

3. The error rates obtained from the system specifications inferred by Spec Inference from Net Load and Spec Inference from PV production are comparable to each other. This proves that CPR's Spec Inference tool for inferring system specifications from net load is working well. This also indicates the potential for utilities that only collect net load data to infer the system specifications of the PV systems installed in their service territories.

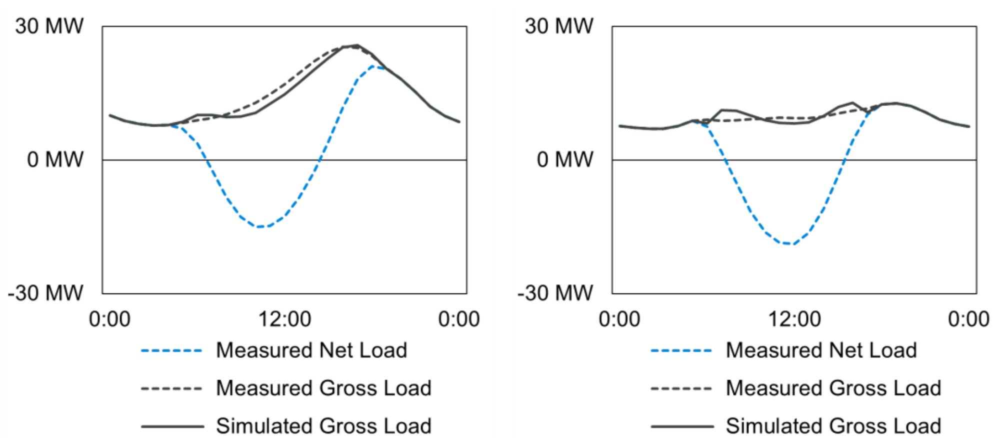
6.2.9 Gross Load Reconstruction

Simulated production and measured net load can be used to estimate gross load by adding the simulated production with the net load. Gross load is the total load that reflects the actual electricity demand at the customer sites. This information is useful for utilities to prepare for contingencies and worst-case scenarios such as cases when all PV systems suddenly go offline. In the upcoming sub-sections, gross load is estimated by adding net load with simulated production generated using system specifications inferred by 1) Spec Inference from Net Load, 2) Spec Inference from PV production, 3) PowerClerk, and 4) Extended Spec Inference from PV production. In each of the four cases, gross load is reconstructed using the simulated production that considers shading and soiling losses, discussed in Section 6.2.7.

6.2.9.1 Net Load

6.2.9.1.1 Simulation Results for One Day

Figure 39a) and Figure 39b) show the gross load for a day in the summer (June 22, 2016) and fall (September 22, 2016) seasons, respectively. The hourly rMAE between the measured production and simulated production for the days in summer and fall for the fleet of 8,324 sites is 4.9% and 6.1%, respectively.



a) 06-22-2016 b) 09-22-2016

Figure 39: Fleetwide hourly gross load of selected days.

6.2.9.1.2 Hourly Simulation Results for Entire Year

The hourly rMAE between the simulated and measured gross load during 2016 for the fleet of 8,324 sites, shown in Figure 40 is 5.3%.

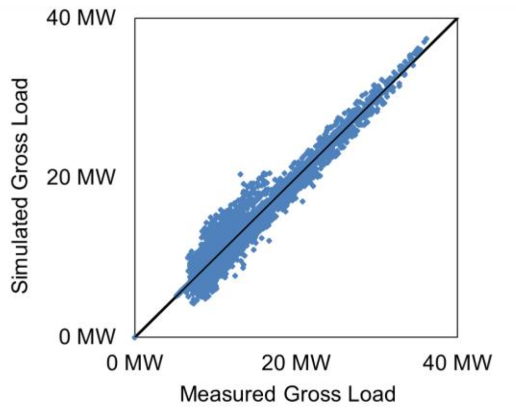


Figure 40: Fleetwide hourly gross load for 2016.

6.2.9.1.3 Daily Simulation Results for Entire Year

The daily rMAE between the simulated and measured gross load during 2016 for the fleet of 8,324 sites, shown in Figure 41 is 3.2%.

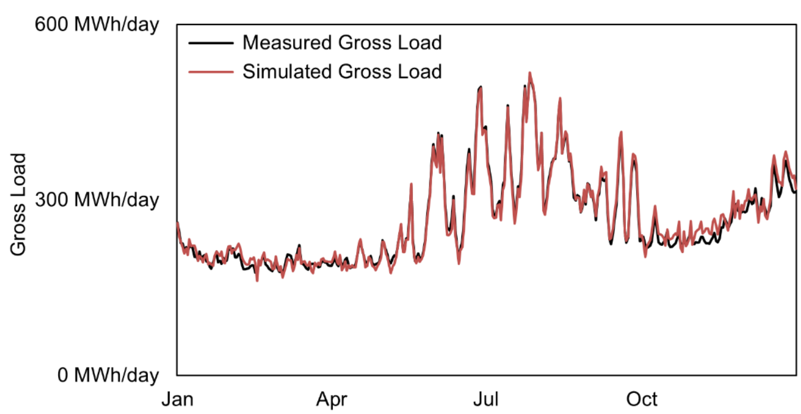


Figure 41: Fleetwide daily gross load for 2016.

6.2.9.2 PV Production

6.2.9.2.1 Simulation Results for One Day

Figure 42a) and Figure 42b) show the gross load for a day in the summer (June 22, 2016) and fall (September 22, 2016) seasons, respectively and the hourly rMAE between the measured production and simulated production for the days in summer and fall for the fleet of 8,324 sites is 4.2% and 6.0%, respectively.

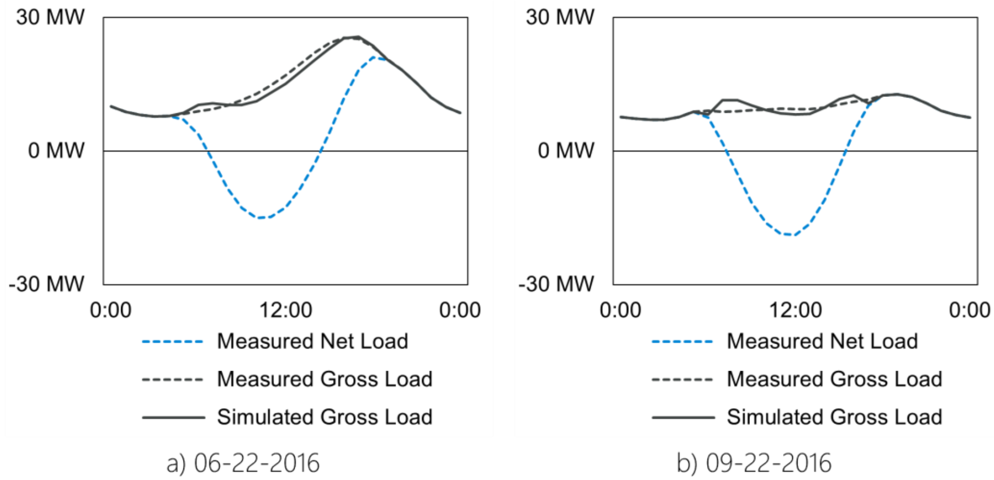


Figure 42: Fleetwide hourly gross load of selected days.

6.2.9.2.2 Hourly Simulation Results for Entire Year

The hourly rMAE between the simulated and measured gross load during 2016 for the fleet of 8,324 sites, shown Figure 43 is 5.1%.

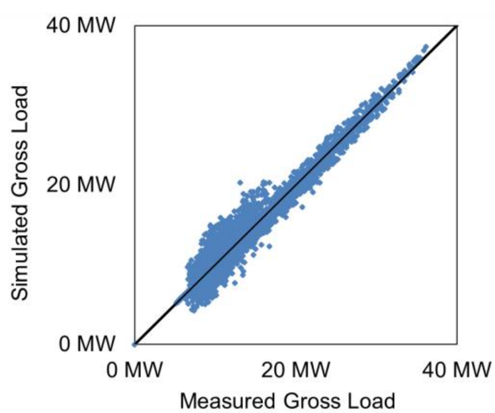


Figure 43: Fleetwide hourly gross load for 2016.

6.2.9.2.3 Daily Simulation Results for Entire Year

The daily rMAE between the simulated and measured gross load during 2016 for the fleet of 8,324 sites, shown in Figure 44 is 2.9%.

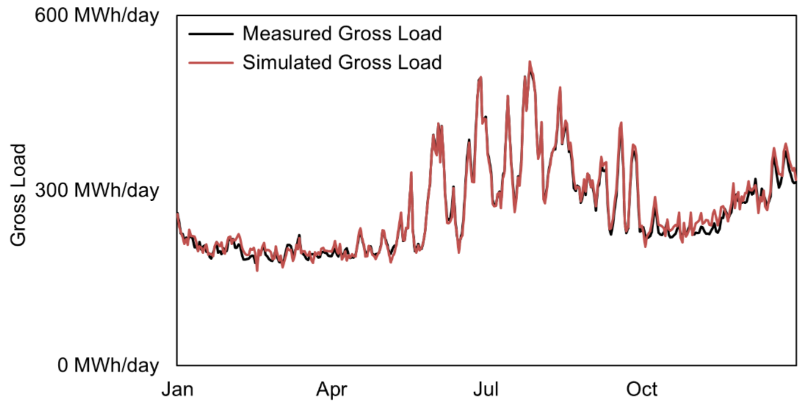


Figure 44: Fleetwide daily gross load for 2016.

6.2.9.3 PowerClerk

6.2.9.3.1 Simulation Results for One Day

Figure 45a) and Figure 45b) show the gross load for a day in the summer (June 22, 2016) and fall (September 22, 2016) seasons, respectively. The rMAE between the measured production and simulated production for the days in summer and fall for the fleet of 8,324 sites are 2.2% and 2.1%, respectively.

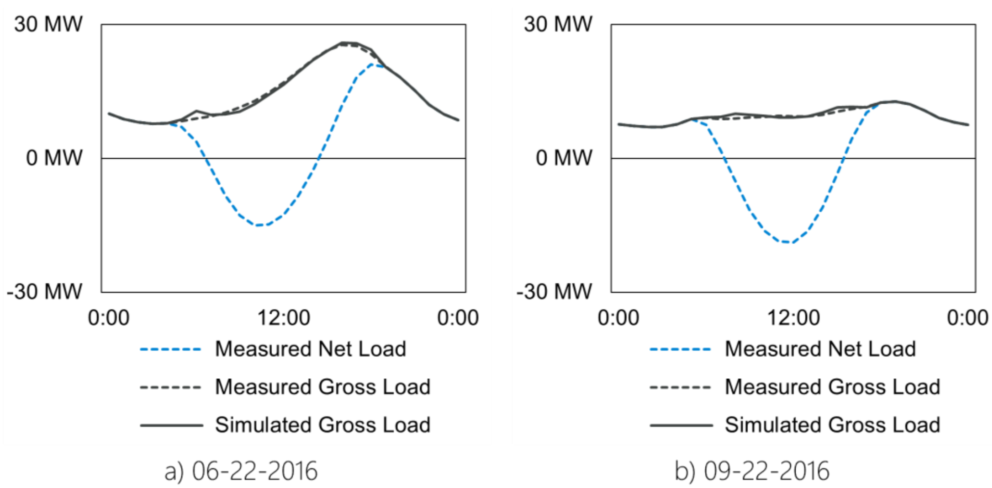


Figure 45: Fleetwide hourly gross load of selected days.

6.2.9.3.2 Hourly Simulation Results for Entire Year

The hourly rMAE between the simulated and measured gross load during 2016 for the fleet of 8,324 sites, shown in Figure 46 is 3.3%.

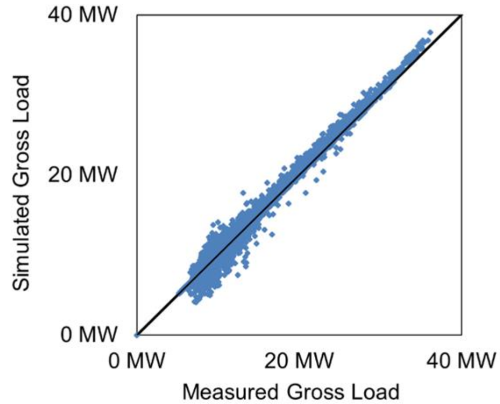


Figure 46: Fleetwide hourly gross load for 2016.

6.2.9.3.3 Daily Simulation Results for Entire Year

The daily rMAE between the simulated and measured gross load during 2016 for the fleet of 8,324 sites, shown in Figure 47 is 2.5%.

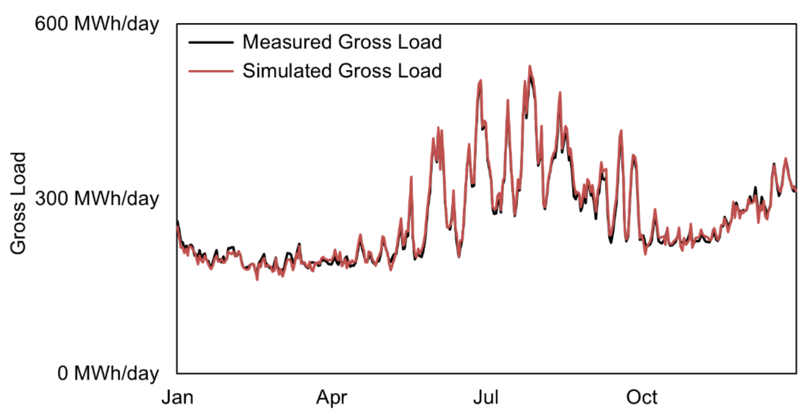
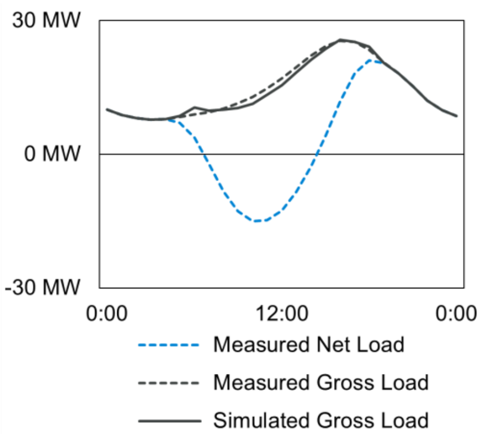


Figure 47: Fleetwide daily gross load for 2016.

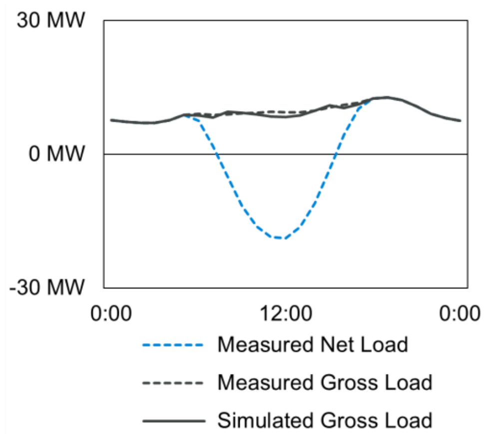
6.2.9.4 Spec Engine

6.2.9.4.1 Simulation Results for One Day

Figure 48a) and Figure 48b) show the gross load for a day in the summer (June 22, 2016) and fall (September 22, 2016) seasons, respectively. The hourly rMAE between the measured production and simulated production for the days in summer and fall for the fleet of 8,324 sites are 3.2% and 2.7%, respectively.



a) 06-22-2016



b) 09-22-2016

Figure 48: Fleetwide hourly gross load of selected days.

6.2.9.4.2 Hourly Simulation Results for Entire Year

The hourly rMAE between the simulated and measured gross load during 2016 for the fleet of 8,324 sites, shown in Figure 49 is 3.0%.

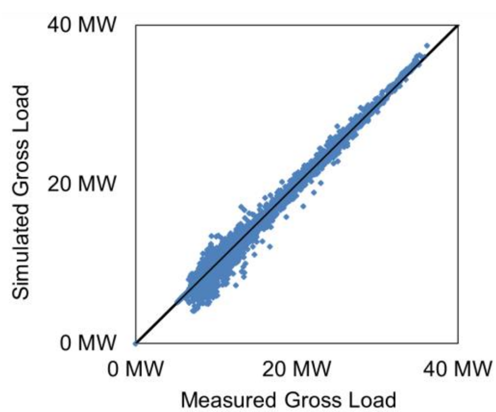


Figure 49: Fleetwide hourly gross load for 2016.

6.2.9.4.3 Daily Simulation Results for Entire Year

The daily rMAE between the simulated and measured gross load during 2016 for the fleet of 8,324 sites, shown in Figure 50 is 1.7%.

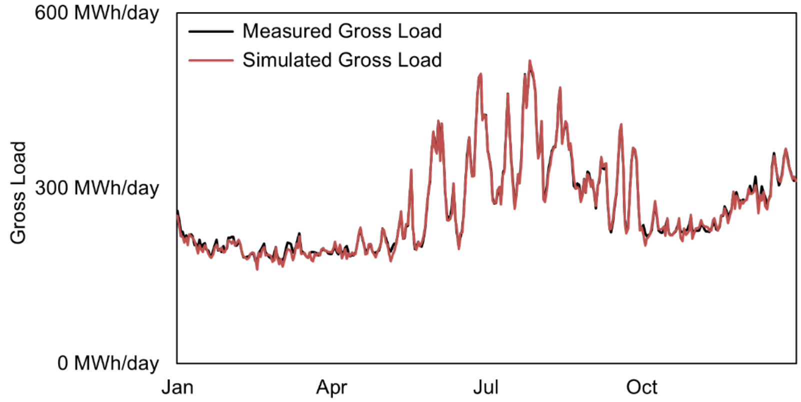


Figure 50: Fleetwide daily gross load for 2016.

6.2.10 Summary of Fleetwide Gross Load

Table 7 and Figure 51 show the hourly and daily rMAE between the measured gross load and simulated gross load during 2016 for the fleet of 8,324 sites using the system specifications inferred by: 1) Spec Inference from Net Load, 2) Spec Inference from PV production, 3) reported from PowerClerk, and 4) Extended Spec Inference from PV production. Additionally, three different case studies are undertaken as follows: a) no consideration of shading and soiling losses, b) consideration of shading losses, and c) consideration of both shading and soiling losses. Note that for system specifications from PowerClerk and Extended Spec Engine, the simulated production considers shading in multiple azimuth directions. For the system specifications from Spec Inference from Net Load and Spec Inference from PV production, the simulated production only considers a constant horizon shading elevation angle of 10 degrees. For the case of consideration of soiling losses, a per day soiling of 0.06% is considered.

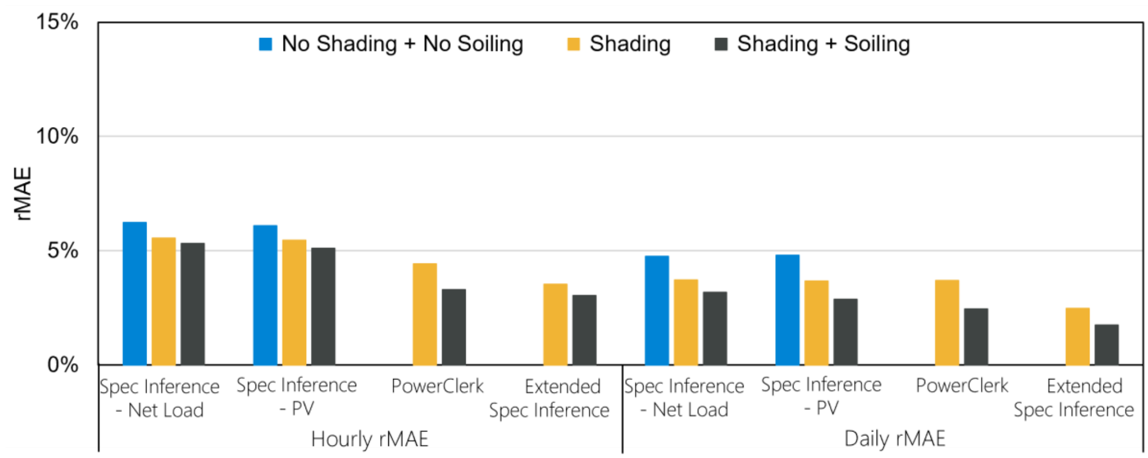


Figure 51: Hourly and daily rMAE between measured gross load and simulated gross load during 2016 for the fleet of 8,324 sites using system specifications inferred by: 1) Spec Inference from Net Load, 2) Spec Inference from PV production, 3) PowerClerk, and 4) Extended Spec Inference from PV production.

Table 7: rMAE between measured and simulated gross load during 2016 for the fleet of 8,324 sites.

Method/Data Source	No shading and No Soiling		Shading		Shading and Soiling	
	Hourly rMAE	Daily rMAE	Hourly rMAE	Daily rMAE	Hourly rMAE	Daily rMAE
Spec Inference from Net Load	6.2%	4.8%	5.5%	3.7%	5.3%	3.2%
Spec Inference from PV	6.1%	4.8%	5.5%	3.7%	5.1%	2.9%
PowerClerk			4.4%	3.7%	3.3%	2.5%
Extended Spec Inference from PV			3.5%	2.5%	3.0%	1.7%

7 Conclusion

We comprehensively discussed the software tool, named Spec Inference, developed by CPR to infer PV system specifications using net load. Spec Inference is also applicable to PV production. We compared the PV system specifications inferred by Spec Inference from Net Load to 1) system specifications inferred by Spec Inference from PV production, 2) reported system specifications in PowerClerk, and 3) system specifications inferred by Extended Spec Inference from PV production. Spec Inference method was applied to net load and PV production dataset of a fleet of 8,324 PV sites in the Sacramento Municipal Utility District. Optimal PV system specifications were inferred for each PV system. The resulting rMAE of 9.0% between the hourly measured and simulated PV production, with the inclusion of shading and soiling losses, verified that the key system specifications inferred using Spec Inference from Net Load are close to the actual system specifications of the sites. While the best set of PV system specifications were obtained using Extended Spec Inference from PV production (proven by the hourly rMAE of 5.1% with the inclusion of shading and soiling losses), PV production data may not be available to most utilities. Thus, CPR's Spec Inference from Net Load is the alternative and low-cost approach beneficial to utilities and system operators that only collect net load data. The resultant system specifications can be used to simulate time-series PV production and reconstruct gross load by combining the net load and simulated PV production.

For more information, contact us:

www.cleanpower.com/contact-us/

www.cleanpower.com