# Data Challenges in Estimating the Capacity Value of Solar Photovoltaics

Dhruv Gami, Ramteen Sioshansi, *Senior Member, IEEE,* and Paul Denholm, *Senior Member, IEEE*

***Abstract*—We examine the robustness of solar capacity-value estimates to three important data issues. The first is the sensitivity to using hourly averaged as opposed to subhourly solar-insolation data. The second is the sensitivity to errors in recording and interpreting load data. The third is the sensitivity to using modeled as opposed to measured solar-insolation data.**

**We demonstrate that capacity-value estimates of solar are sensitive to all three of these factors, with potentially large errors in the capacity-value estimate in a particular year. If multiple years of data are available, the biases introduced by using hourly averaged solar-insolation can be smoothed out. Multiple years of data will not necessarily address the other data-related issues that we examine.**

**Our analysis calls into question the accuracy of a number of solar capacity-value estimates relying exclusively on modeled solar-insolation data that are reported in the literature (including our own previous works). Our analysis also suggests that multiple years' historical data should be used for remunerating solar generators for their capacity value in organized wholesale electricity markets.**

***Index Terms*—Capacity value, power system reliability, photovoltaic solar power generation**

## Nomenclature

| | |
|---|---|
| $G_t$ | Time-$t$ generating capacity available |
| $L_t$ | Time-$t$ load |
| $\bar{L}$ | Fixed load added in each time step |
| $p_t$ | Time-$t$ loss of load probability (LOLP) of base system |
| $p_t^S(\bar{L})$ | Time-$T$ LOLP of system with solar generation and $\bar{L}$ MW of load added |
| $S_t$ | Time-$t$ solar generation available |
| $T$ | Time index set |
| $\epsilon$ | Loss of load expectation (LOLE) of base system |
| $\epsilon^S(\bar{L})$ | LOLE of system with solar generation and $\bar{L}$ MW of load added |

## I. Introduction

ELECTRICITY is a commodity for which supply and demand must be instantaneously balanced at all times. A power system delivers energy to end customers. However, demand must be curtailed to ensure system stability if at any time there is insufficient generating capacity to meet real-time demand. Thus, electricity-generating facilities provide value both by supplying energy and through their contribution to reducing the extent, duration, and likelihood of such load curtailments. Capacity value is a standard metric that measures the contribution of a generator or other asset to the power system reliably serving load.

This work was supported by the U.S. Department of Energy through prime contract DE-AC36-08GO28308 and by the Alliance for Sustainable Energy, LLC through subcontract AGN-2-22165-01.

D. Gami and R. Sioshansi are with the Department of Integrated Systems Engineering, The Ohio State University, Columbus, OH 43210, USA (e-mail: gami.3@osu.edu and sioshansi.1@osu.edu).

P. Denholm is with the Strategic Energy Analysis Center, National Renewable Energy Laboratory, Golden, CO 80401, USA (e-mail: paul.denholm@nrel.gov).

System reliability assessments examine the effects of high loads or generator or transmission failures requiring load curtailment. Reliability assessments estimate the probability that such load-curtailment events occur. A resource's capacity value is assessed by estimating the effect that the resource has on reducing load-curtailment probabilities.

Generator failures can occur due to mechanical failures, planned maintenance, or a lack of generating resource. Lack of generating resource is of particular importance to renewable energy resources, such as wind and solar. This is because the ability of such renewable resources to generate electricity depends on climatic conditions, which cannot be controlled, is variable, and may be difficult to forecast. Given this reality, there is a vast and growing literature applying capacity value-assessment techniques to renewable resources. This includes analyses of wind [1]–[19], solar photovoltaic (PV) [20]–[27], and solar thermal [28], [29] plants. These analyses typically use historical load, conventional-generator, and renewable-availability data to estimate the capacity value of the renewable facility being studied. Historical data are used to capture renewable variability and the statistical dependence between renewable availability and load (as load is driven by climatic conditions in many systems).

These analyses find that the capacity value of a renewable facility can range between $5\%$ and $95\%$ of its nameplate capacity. This wide range reflects differences related to geography and technology, which affect load and renewable-availability patterns, and the penetration level of the technology in question. One typical finding is that at low penetration levels, solar generation facilities can have much higher capacity values than wind [27]. This is because solar availability tends to be more coincident with load, especially in summer-peaking power systems, than wind availability is. In many systems, the capacity value of solar is driven by solar availability in about the $10$ highest-load hours of the year [28], [29]. By contrast, the capacity value of wind is driven by wind availability in about the $1000$ highest-load hours of the year.

This difference in the fundamental driver of solar capacity value (compared to wind) raises important questions about the robustness of capacity-value estimates to the underlying data. This is because capacity-value studies often rely on data that are either modeled, recorded at coarse time intervals, or may have other errors. This paper examines the robustness of PV capacity-value estimates to three data issues.

We first examine the use of hourly averaged solar-insolation data (which are typically used) as opposed to one-minute data. The question that we address here is the effect on capacity-value estimates of using hourly averaged data, which can 'smooth' solar-output profiles. We show that hourly averaged data can under- or overestimate the capacity value of solar by up to $10\%$ in an individual year. Having multiple years of data tends to mitigate the errors introduced by the use of hourly averaged data, however.

We next examine the effect of errors in recording load data. There can be ambiguities about load data. Hour-$t$ load data could represent the instantaneous hour-$t$ load measurement, the average load over the hour-ending $t$, the average load over the hour-beginning $t$, or something else. Moreover, utilities may handle daylight savings time differently in reporting load. We bound this effect by examining the sensitivity of PV capacity-value estimates to shifting load profiles forward and backward by one hour. We find that these load shifts can change the capacity value of PV by more than $35\%$ and that the load shifts do not necessarily always affect the capacity-value estimate in the same direction. Moreover, having multiple years of data will not necessarily eliminate the biases introduced by load errors.

We finally examine the effect of using modeled as opposed to measured solar-insolation data to simulate PV output. We find that modeled data can significantly over- or underestimate capacity value. We find that using multiple years of data will not necessarily eliminate these biases either.

The remainder of this paper is organized as follows. Section II details the reliability-based model that we use in all of our capacity-value estimates and Section III discusses the data that are used in our analysis. Section IV summarizes our capacity-value estimates for the PV plants studied and the sensitivity analyses detailed above. We also provide some examples that demonstrate why our capacity-value estimates exhibit the sensitivities that we observe. We provide examples showing the effects of smoothing-out one-minute solar-insolation data by using hourly averaged data instead. We also demonstrate the different types of effects that using modeled as opposed to measured solar-insolation can have on capacity-value estimates. Section V concludes.

## II. Capacity-Value Estimation Model

We use a standard reliability-based approach to estimate the capacity value of PV. The specific capacity-value metric that we use is the effective load-carrying capability (ELCC). ELCC is defined as the amount by which a system's load can increase when the output of the PV plant is added to the system, without changing system reliability.

To compute the ELCC of a PV plant, we first assess the reliability of the base system, without the PV. To do this, we compute the loss of load probability (LOLP) of the base system in each time step as:

$$p_t = \text{Prob}\left\{G_t < L_t\right\}, \forall t \in T. \tag{1}$$

The $\text{Prob}\{\cdot\}$ function in (1) can capture generator failures, random load, transmission failures (which can limit the ability of particular generators to serve load), or other random events that can cause a loss of load.

We next compute the loss of load expectation (LOLE) of the base system as the sum of the LOLPs:

$$\epsilon = \sum_{t \in T} p_t. \tag{2}$$

The LOLE represents the expected number of time periods (the duration of the time periods are defined based on the temporal granularity used to compute the LOLPs) over the study period (which is determined by the number of time periods in the set, $T$).

We next compute the LOLPs of the system with the PV plant added to the generator set and an additional $\bar{L}$ MW of load in each time step as:

$$p_t^S(\bar{L}) = \text{Prob}\left\{G_t + S_t < L_t + \bar{L}\right\}, \forall t \in T. \tag{3}$$

The variable, $S_t$, represents the output of the PV plant in these LOLP calculations. The $\text{Prob}\{\cdot\}$ function in (3) captures the same types of events captured by the $\text{Prob}\{\cdot\}$ function in (1). It can also capture the effects of random events (*e.g.*, climatic conditions) on PV output. We then compute the LOLE of the system with the PV and added load as:

$$\epsilon^S(\bar{L}) = \sum_{t \in T} p_t^S(\bar{L}). \tag{4}$$

The final step of the ELCC calculation is to determine a value of $\bar{L}$ that makes the LOLE of the base system and the system with the added load and PV equal. That is, we determine a value of $\bar{L}$ for which:

$$\epsilon^S(\bar{L}) = \epsilon.$$

This is typically done in an iterative fashion, for instance using a bisection method, which is the method employed in our analysis.

## III. Case-Study Data

Our case study examines 10 locations across the United States. More specifically, we examine the locations listed in Table I, for which historical measured solar-insolation data are available at one-minute intervals. The table lists the geographical coordinates of each ground-level measurement station, from which insolation data are taken. It also lists which electric utility's footprint each station is in. We use the local utility to model system loads and capacity available from the remaining generation fleet (*i.e.*, the values of $L_t$ and $G_t$) in the LOLP calculations. Table I also lists the years for which our analysis is conducted for each location. We only conduct ELCC estimates for years for which complete solar-production, load, and generator data are available.

We now detail the solar-production and power system data used in our analysis.

### A. Solar-Production Data

The locations that we study do not have PV plants installed. Thus, we use weather data to simulate the output of a hypothetical PV plant. We now detail the data sources and model used for these simulations.

TABLE I
LOCATIONS STUDIED

| Location | Coordinates | Years |
|---|---|---|
| Nevada Energy, Nevada (NE) | | |
| 1 | 36.09° N, 115.07° W | 2006 |
| 2 | 36.09° N, 115.15° W | 2006 |
| 3 | 36.61° N, 116.03° W | 1999–2002, 2004–2013 |
| Public Service Company of Colorado, Colorado (PSCC) | | |
| 4 | 39.74° N, 105.18° W | 2005–2006 |
| 5 | 40.13° N, 105.23° W | 1999–2002, 2004–2013 |
| NorthWestern Energy, Montana (MPC) | | |
| 6 | 48.33° N, 105.11° W | 1998–1999, 2002–2013 |
| NorthWestern Energy, South Dakota (NWPS) | | |
| 7 | 43.73° N, 96.63° W | 2008, 2010–2013 |
| South Mississippi Electric Power Association, Mississippi (SMEA) | | |
| 8 | 34.25° N, 89.87° W | 1998–2005, 2007–2013 |
| Southern Illinois Power Cooperative, Illinois (SIPC) | | |
| 9 | 40.05° N, 88.39° W | 1998, 2000, 2001, 2003, 2004, 2007 |
| West Penn Power, Pennsylvania (WPP) | | |
| 10 | 40.73° N, 77.95° W | 1999, 2002–2003, 2007–2013 |

*1) Measured Solar-Insolation Data:* Solar-insolation data, which are gathered by ground stations at the locations studied at one-minute intervals, are publicly available from the MIDC[1] and SURFRAD[2] data repositories. These datasets include solar insolation, temperature, and wind speed data, all of which are needed to simulate the electric output of a PV plant.

One of our sensitivity analyses is to compare ELCC estimates using solar-insolation data measured at one-minute intervals to hourly solar-insolation measurements. We model PV output with hourly data by averaging the 60 one-minute solar-insolation measurements corresponding to each hour to obtain an hourly averaged solar-insolation measurement. We then construct a dataset in which each hourly averaged solar-insolation measurement is repeated 60 times. By doing this, we examine the effect of changing the input data from one-minute to hourly measurements without any other changes in the underlying PV-simulation model.

Hourly PV simulations often caluculate the sun position at the midpoint of each hour assuming hourly averaged solar-irradiance data. The method that we employ calculates the correct sun position at each minute. The sun position affects the calculated angle of incidence of irradiance on the PV system. Thus, there may be small differences in energy output between the two methods.

*2) Modeled Solar-Insolation Data:* One of the sensitivity analyses that we conduct is to compare ELCC estimates using measured solar-insolation data to estimates using modeled data. We obtain modeled solar-insolation data from version 2.0.0 of the National Solar Radiation Database (NSRDB) [30], [31] for this comparison. The NSRDB uses satellite images to estimate the amount of solar radiation that impacts the surface of the earth at thirty-minute intervals on a 4 km$^2$ grid of the United States.

We use modeled solar-insolation data for the location in the NSRDB that is closest to each location in the MIDC and SURFRAD datasets. The NSRDB data are reported in UTC. We shift the data to local time using the standard time difference between the local time zone of each location and UTC. Because the NSRDB data are reported at 30-minute intervals, we convert the NSRDB data to hourly averaged solar-insolation data and compare the NSRDB-based ELCC estimates to those using hourly average measured data.

*3) Simulated Solar Generation:* The production of a solar plant at each study location is estimated using version 5 of the PVWatts model [32] from the System Advisor Model Software Development Kit.[3]

We assume that each PV plant has a 100 MW dc nameplate capacity and a dc to ac derate factor of 0.81. This derate factor captures energy and efficiency losses associated with converting the dc output of the PV cells to ac. The plants are assumed to be fixed-axis with a 180° azimuth and an optimal tilt equal to the location's latitude.

## B. Power System Data

Modeling LOLPs and LOLEs requires load and generator data. We now detail the data sources used and how specifically LOLPs are computed.

*1) Load Data:* Hourly historical load data are obtained from Form 714 data submitted by each of the utilities studied to the Federal Energy Regulatory Commission.[4] Because our base case analysis uses solar insolation data recorded at one-minute intervals, each hourly load observation is repeated 60 times to obtain time-synchronized load data.

*2) Generator Data:* Historical generator data are obtained from Form 860 data submitted by each generator to the United States Department of Energy's Energy Information Administration.[5] The Form 860 data specify the owner of each generating facility. We only consider generating facilities that were owned by each utility and operational in each year studied in modeling their respective power systems. The Form 860 data also specify the nameplate capacity, generating fuel, and prime mover of each generator.

These data are combined with historical effective forced outage rate (EFOR) data reported in the North American Electricity Reliability Corporation's (NERC's) Generating Availability Data System (GADS).[6] The GADS reports historical EFORs based on generating technology and plant capacity.

*3) LOLP Calculation:* Our base case analysis uses solar insolation data recorded at one-minute intervals. Thus, we compute LOLPs at one-minute intervals. LOLPs are computed assuming that the loads and PV output are fixed based on their historical values. The only randomness modeled are conventional-plant failures. These failures are represented using a simple two-state (online/offline) model. We assume that generator outages are serially and cross-sectionally independent, with each plant's failure probability being the corresponding EFOR obtained from the GADS. Generator EFORs and nameplate capacities are used to compute each utility system's capacity outage probability table [33].

Some of the systems that we model may export or import energy to neighboring regions. As such, the LOLE of the base

[1] https://www.nrel.gov/midc/
[2] http://www.esrl.noaa.gov/gmd/grad/surfrad/
[3] https://sam.nrel.gov/sdk
[4] http://www.ferc.gov/docs-filing/forms/form-714/data.asp
[5] https://www.eia.gov/electricity/data/eia860/
[6] http://www.nerc.com/pa/RAPA/gads/Pages/default.aspx

system may be higher or lower than NERC's one outage-day in 10 years reliability standard [9]. To account for these imports and exports, we proportionally scale the loads so that the base system has an LOLE of 144 minutes (2.4 hours) in each year.

Table II summarizes the range (over the years studied) of unscaled loads in each of the utility systems modeled. As is common in the United States, many of the systems are summer-peaking (due to air conditioning loads). However, the MPC, NWPS, and SMEA systems have winter peaks that are very close to the summer peak. These systems are also winter-peaking in some of the years studied.

TABLE II
UTILITY LOAD DATA [GW]

| Utility | Average | Summer Peak | Winter Peak |
|---|---|---|---|
| NE | 1.9–3.0 | 4.0–6.3 | 2.2–4.0 |
| PSCC | 2.8–4.8 | 4.5–8.2 | 4.0–6.8 |
| MPC | 0.3–1.1 | 0.4–1.5 | 0.4–1.5 |
| NWPS | 1.2–1.2 | 1.6–1.8 | 1.7–1.8 |
| SMEA | 0.2–1.7 | 0.2–2.7 | 0.4–2.9 |
| SIPC | 1.8–5.5 | 3.2–9.4 | 2.5–7.2 |
| WPP | 4.5–8.1 | 6.4–13.8 | 6.8–10.9 |

## IV. ELCC ESTIMATES

This section first summarizes our ELCC estimates in the base case, in which solar-insolation data measured at a ground station at one-minute intervals are used. We then discuss the results of our sensitivity analyses.

### A. Base-Case ELCC Estimates

Table III provides summary statistics of the annual ELCC estimates (measured in MW) for each location studied. This includes the average, standard deviation, minimum, and maximum ELCCs (across the years studied) for each location. Because only a single year's data are available for Locations 1 and 2, only the averages are reported for these two. The table shows that the ELCCs in a given year can range between 8 MW and 41 MW. These values correspond to 10% and 50% of the 83 MW ac nameplate capacity of the PV plants studied.

TABLE III
ELCC ESTIMATES USING ONE-MINUTE MEASURED SOLAR-INSOLATION DATA [MW]

| Location | Average | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|
| 1 | 30.4 | | | |
| 2 | 26.3 | | | |
| 3 | 24.2 | 3.0 | 17.8 | 30.5 |
| 4 | 23.1 | 5.2 | 19.4 | 26.8 |
| 5 | 24.2 | 7.0 | 14.6 | 40.9 |
| 6 | 16.9 | 4.9 | 7.9 | 23.0 |
| 7 | 17.7 | 9.2 | 9.0 | 32.1 |
| 8 | 25.5 | 5.9 | 13.2 | 32.2 |
| 9 | 26.5 | 3.6 | 21.5 | 32.4 |
| 10 | 20.9 | 5.1 | 15.3 | 31.5 |

Solar ELCC is driven by the coincidence between solar insolation and load. For this reason winter-peaking systems tend to see low ELCCs. Locations 6–8 and 10 are in utilities that are winter peaking in some of the years studied. The average ELCC of PV plants in these locations in the winter-peaking years are between 9% and 56% lower than the ELCCs in the summer-peaking years. Our finding that ELCCs are lower in winter-peaking systems is consistent with other capacity-value estimates of solar PV reported in the literature, although our overall ELCC estimates are lower than those reported elsewhere [23].

### B. Sensitivity of ELCCs to Hourly Solar-Insolation Data

Table IV summarizes the sensitivity of the ELCC estimates to using hourly as opposed to one-minute solar-insolation data. The table shows the average and maximum absolute value of differences (across the years studied) in ELCCs when using the two sets of solar-insolation measurements. These values are reported as absolute [MW] differences and as percentages of the average ELCC estimate obtained using one-minute data. Because only a single year's data are available for Locations 1 and 2, maximum differences are not reported for these two. Table IV shows that using hourly averaged data can bias the resulting ELCCs significantly in a single year. For instance, the ELCC based on one-minute solar-insolation data of the PV plant in Location 3 is 18 MW in 2005. This is overestimated by 2 MW (a 9% error) if hourly averaged data are used instead.

TABLE IV
ABSOLUTE VALUE OF DIFFERENCES IN ELCC ESTIMATES USING MEASURED ONE-MINUTE AND HOURLY-AVERAGED SOLAR-INSOLATION DATA

| | Mean Difference | | Maximum Difference | |
|---|---|---|---|---|
| Location | [MW] | [%] | [MW] | [%] |
| 1 | 0.7 | 2.3 | | |
| 2 | 2.2 | 8.2 | | |
| 3 | 0.3 | 1.3 | 1.7 | 9.3 |
| 4 | 1.2 | 5.0 | 1.6 | 6.0 |
| 5 | 0.3 | 1.3 | 1.0 | 3.7 |
| 6 | 0.2 | 1.1 | 0.4 | 2.3 |
| 7 | 0.3 | 1.2 | 0.6 | 2.0 |
| 8 | 0.2 | 0.6 | 0.4 | 2.0 |
| 9 | 0.1 | 0.4 | 0.2 | 0.8 |
| 10 | 0.2 | 1.1 | 0.5 | 2.9 |

Fig. 1 demonstrates how hourly averaged solar-insolation data can 'smooth out' variability in PV output, thereby affecting the ELCC estimate. It shows simulated PV output and loads at Location 6 on 13 July, 2005, which is the peak-load day of the year for the NWE system, in which the PV plant is located. Using the hourly averaged and one-minute solar-insolation data gives the same energy production from the PV plants over the course of the day—472 MWh. However, hourly averaged data slightly overestimates PV generation in the afternoon, when the system load is high, impacting the ELCC calculation. One-minute solar-insolation data estimate 192 MWh of PV output from the beginning of hour 12 to the end of hour 15, as opposed to 196 MWh of output estimated during this period using hourly averaged data. These and other differences result in a 1.0% error in the ELCC estimate.

The effect of 'smoothing out' variability in solar-insolation that is shown in Fig. 1 could also be caused by analyzing the energy production of a solar facility that is spread out over a wide geographic area. Because a ground instrument measures solar-insolation at a single geographic point, the measurement tends to exhibit more noise than a series of PV panels spread out over the ground or multiple rooftops would.

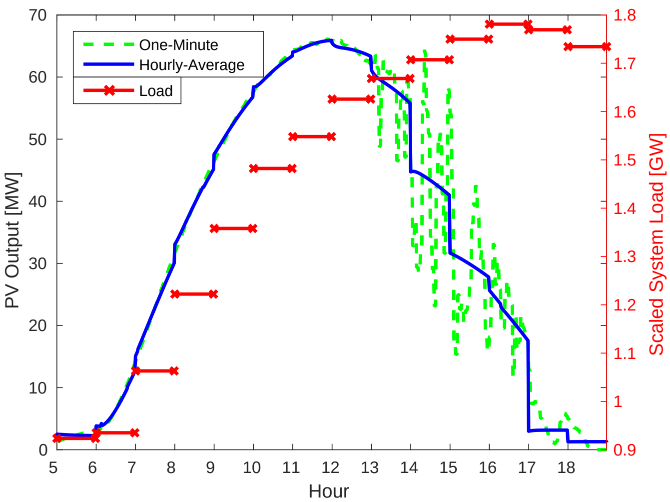


Fig. 1. Simulated PV output using one-minute and hourly averaged solar-insolation data and loads for Location 6 on 13 July, 2005.

Using hourly average data can under- or overestimate ELCCs, even in the same system. However, we find that hourly averaged data overall tends to bias the ELCC estimates upwards. Of the 91 location/year pairs studied, 74 of the ELCCs are overestimated when hourly averaged solar-insolation data are used. This suggests that hourly averaged solar-insolation data tend to overestimate afternoon PV output, giving the upward bias in the estimates.

Using hourly averaged solar-insolation data can introduce significant errors in estimating an ELCC in a particular year. However, the mean absolute differences reported in Table IV show that having multiple years of data tends to smooth out these errors. Locations 3, 5, 6, 8, and 10, for which we have nine to 16 years of data, have average absolute errors in the ELCC estimates between hourly averaged and one-minute solar-insolation data of no more than 1.3%. Locations 2 and 4, for which we have one and two years of data, respectively, have higher average errors of 8.2% and 5.0%, respectively. These findings suggest that although using hourly averaged data can result in biased ELCC estimates, using multiple years of data can mitigate these errors to some extent.

### C. Sensitivity of ELCCs to Load Shifts

Table V summarizes the sensitivity of the ELCC estimates to shifting the load data reported in Form 714 backward and forward one hour. The reason for conducting this analysis is that there is ambiguity in what exactly is reported by utilities in Form 714. The submitted data have no clear documentation indicating what specific load measurements are reported. The hour-$t$ load reported could represent the instantaneous load at hour $t$ or could be an average load over the preceding or following hour. Shifting the loads forward and backward one hour bounds the effect of such errors in reporting and interpreting the load data on ELCC estimates.

For each of backward and forward load shifts, Table V reports the average (over the years studied) differences in the ELCCs compared to using unshifted loads. A positive difference means that using the shifted load gives a higher ELCC estimate compared to using the unshifted loads. The differences are reported as percentages of the average ELCC estimate using unshifted loads. The table also reports the average of the absolute values of the differences.

TABLE V
AVERAGE DIFFERENCES IN ELCC ESTIMATES USING LOAD DATA SHIFTED BACKWARD AND FORWARD ONE HOUR [%]

| | Shifted Backward | | Shifted Forward | |
|---|---|---|---|---|
| Location | Difference | Abs. Value | Difference | Abs. Value |
| 1 | 31.3 | 31.3 | −31.4 | 31.4 |
| 2 | 29.6 | 29.6 | −30.4 | 30.4 |
| 3 | 41.8 | 41.8 | −37.3 | 37.3 |
| 4 | 36.9 | 36.9 | −34.7 | 34.7 |
| 5 | 34.4 | 34.4 | −31.5 | 31.5 |
| 6 | 3.5 | 6.8 | −4.8 | 7.7 |
| 7 | 11.2 | 11.2 | −7.7 | 8.3 |
| 8 | 16.2 | 20.8 | −17.6 | 21.7 |
| 9 | 17.5 | 17.5 | −19.4 | 19.4 |
| 10 | 4.7 | 5.4 | −5.0 | 9.8 |

For all of the systems studied, shifting the loads back one hour gives an average increase in the ELCC estimates. This is because most of these systems are summer-peaking (on average over the years studied). It is common in summer-peaking systems for the peak load to lag peak solar output by a few hours (due to thermal inertia in buildings delaying air conditioning loads). Shifting the loads back improves the coincidence in loads and PV output, increasing ELCCs. There are, however, six, three, and three years, respectively, in which the MPC, SMEA, and WPP systems are winter-peaking. Shifting the loads back one hour decreases the ELCC estimates of PV in these systems (*cf.* Locations 6, 8, and 10) during four, one, and two of these years, respectively. This phenomenon is exhibited by the average differences and the average of the absolute values of the differences in Table V not being the same for Locations 6, 8, and 10.

Table V shows that errors in properly interpreting load data can have a drastic effect on ELCC estimates. Moreover, having multiple years of data does not necessarily smooth out these errors. Locations 3, 5, and 8, for which we have at least 14 years of data, have average (over the years studied) errors of more than 21%. This is because the ELCC differences given by load shifting tend to be persistent over the years and do not smooth out as more data are used for the ELCC analysis.

### D. Sensitivity of ELCCs to Modeled Solar-Insolation Data

Table VI summarizes the sensitivity of the ELCC estimates to using modeled solar-insolation data. Table VI provides summary statistics of the differences in the ELCC estimates using the two datasets. The differences are reported as percentages of the average ELCC estimate obtained using measured solar-insolation data. Because we only have one year's data for Locations 1 and 2, only average values are reported for these two. A positive difference means that the NSRDB-based ELCC is higher than the estimate using measured solar-insolation data.

The table shows that modeled data can introduce significant errors in the ELCC estimates. Moreover, having multiple years of data do not necessarily smooth out these errors, as seen for Locations 6, 8, and 10, for which we have 10–15 year's data.

TABLE VI
DIFFERENCES IN ELCC ESTIMATES USING MEASURED AND MODELED SOLAR-INSOLATION DATA [%]

| Location | Average | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|
| 1 | −1.1 | | | |
| 2 | 10.4 | | | |
| 3 | 0.9 | 5.5 | −17.4 | 4.8 |
| 4 | 3.2 | 1.8 | 1.4 | 5.0 |
| 5 | 0.6 | 7.8 | −17.2 | 12.6 |
| 6 | −7.0 | 10.1 | −24.5 | 15.2 |
| 7 | 26.9 | 10.4 | 15.4 | 41.2 |
| 8 | 12.8 | 27.1 | −5.8 | 112.1 |
| 9 | 7.1 | 5.3 | −0.2 | 12.5 |
| 10 | 4.3 | 4.4 | −2.4 | 11.5 |

Fig. 2 shows the distribution of the differences in the ELCC estimates between using modeled and measured data in each year examined for each location. The differences are given as percentages of the ELCC estimate obtained for each year using measured solar-insolation data. The figure uses the same convention as in Table VI that a positive difference means that the NSRDB-based analysis overestimates the ELCC.

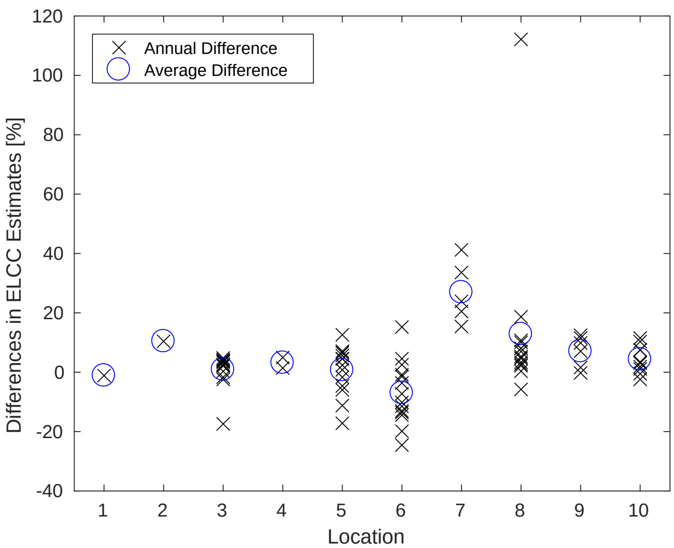


Fig. 2. Annual and average (over years analyzed) differences in ELCC estimates using measured and modeled solar-insolation data.

Modeled solar-insolation data can introduce a number of errors in the ELCC estimates, which are illustrated in Figs. 3 through 6. One is that the modeled data can grossly under- or overestimate solar-insolation. An example of this is given in Fig. 3, which shows simulated PV output using measured and modeled solar-insolation data and loads for Location 3 on 12 July, 2012. This is the highest-load day for the NE utility system, which contains Location 3. The figure shows that the NSRDB data completely underestimate PV output on this day. The measured solar-insolation data yield 493 MWh of generation during the day shown as opposed to 223 MWh using the NSRDB data. As a result, using NSRDB data underestimates the ELCC computed using measured solar-insolation data by 17% in 2012.

We find that at some locations the NSRDB can systematically under- or overestimate PV output over multiple years. As an example of this, the NSRDB overestimates PV output

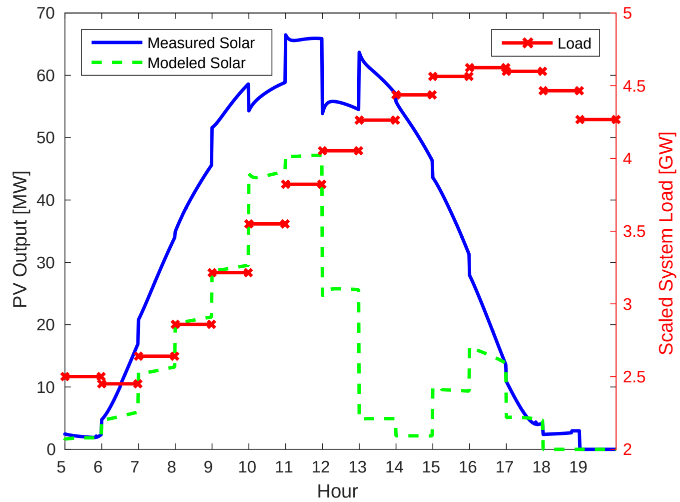


Fig. 3. Simulated PV output using measured and modeled solar-insolation data and loads for Location 3 on 12 July, 2012.

in an individual year by between 1% and 13% (relative to the measured data) for Location 9, giving an average (over the six years studied) PV-output overestimate of 7%. As a result, the NSRDB gives an average ELCC estimate that is 7% greater than that using measured data. Interestingly, the NSRDB slightly underestimates the ELCC for Location 9 in 2007 but overestimates aggregate PV output in that year by 1%. This is because the NSRDB underestimates PV output on the highest-load day of 2007.

Figs. 4 and 5 demonstrate a more extreme case of this for Location 5 in 2004. The NSRDB overestimates PV output over the year by 0.5% relative to the measured solar-insolation data. Moreover, as Fig. 4 shows, the average (over the year) simulated diurnal PV-output profile using measured and NSRDB data are almost identical. Fig. 5 shows, however, that although the average performance of the NSRDB is virtually identical to the measured data for 2004, the NSRDB underestimates PV output on the highest-load day of the year in the NE system, which is 14 July. As a result, the NSRDB underestimates the ELCC by 11% relative to using measured data. This finding, again, highlights the extreme sensitivity of the ELCC estimate to output during the small number of peak-load hours in the study period.

At other locations, the NSRDB can both under- or overestimate PV output across multiple years. In the case of Location 6, the NSRDB under- or overestimates PV output in an individual year by between 12% and 11% and underestimates PV output by 6% over the 14 years studied. As a result, the average ELCC is underestimated by 7% relative to using measured data.

Fig. 6 illustrates another type of error that modeled solar-insolation data can introduce. It shows simulated PV output using measured and modeled solar-insolation data and loads for Location 8 on 27 June, 2013, the highest-load day of the year. The two solar-insolation data sets give very similar aggregate PV output profiles on this day—469 MWh and 455 MWh using measured and modeled data, respectively. The modeled data, however, underestimate PV production in

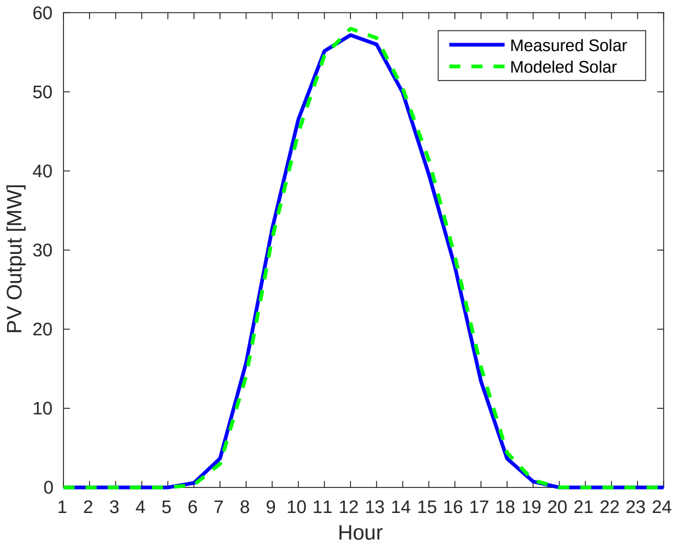


Fig. 4. Average (over the year) simulated diurnal PV-output profile using measured and modeled solar-insolation data for Location 5 in 2004.

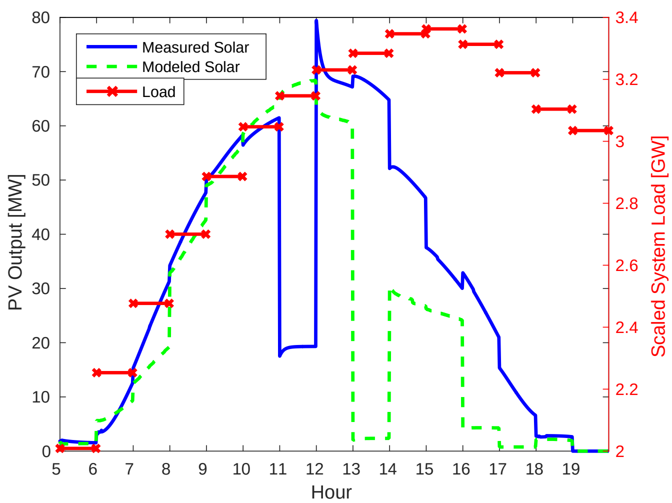


Fig. 5. Simulated PV output using measured and modeled solar-insolation data and loads for Location 5 on 14 July, 2004.

the morning hours and overestimate it in the afternoon. As a result, the modeled data give an ELCC estimate that is 19% greater than that using measured solar-insolation data.

## V. Conclusions

This paper examines the sensitivity of solar PV ELCC estimates to a number of potential data issues. Figs. 1, 3, 5, and 6 give specific examples showing how using hourly averaged or modeled solar-insolation can introduce significant errors in capacity-value estimates. The errors shown in these figures are not the only ones possible. However, these examples convey the major types of errors that we observe in our analysis.

Our analysis suggests that if a sufficient number of years' data are used in a capacity-value analysis, hourly solar-insolation data should provide relatively robust results. However, errors in recording or interpreting load data cannot necessarily be fixed through more data. More troubling, our analysis shows that modeled solar-insolation can introduce large

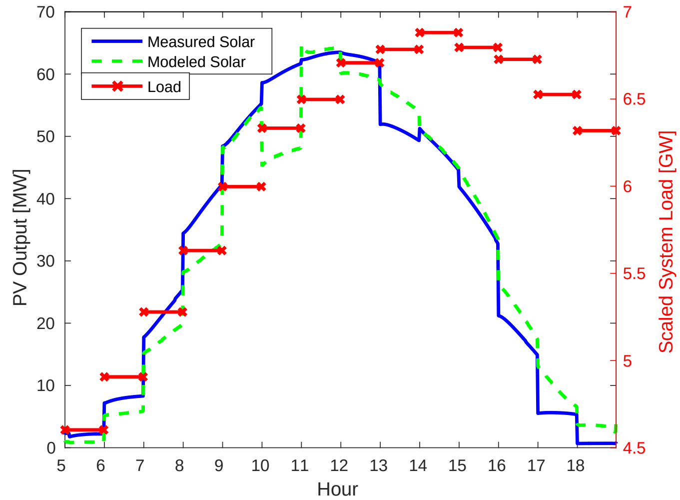


Fig. 6. Simulated PV output using measured and modeled solar-insolation data and loads for Location 8 on 27 June, 2013.

errors in capacity-value estimates. This finding is problematic because measured solar-insolation data are not available at the spatial and temporal granularity that modeled data are. Even at locations with a ground-based measurement station, 'holes' in the data appear due to instrument failures. For instance, the measured solar-insolation data for Location 10 show zero solar insolation for the highest-load day of two of the years studied (these years are, obviously, excluded from our analysis).

The results of this paper call into question the accuracy of photovoltaic capacity-value estimates in the literature (including our own previous works) that rely on modeled solar-insolation and publicly reported load data. It is important to stress, however, that we only analyze a single modeled solar-insolation dataset (the NSRDB). There are numerous other modeled solar-irradiance datasets and a comparison of the performance of the NSRDB to these datasets is an important future research topic. The NSRDB is the only publicly available dataset that uses a physics-based approach (specifically, using the Fast All-Sky Radiation Model for Solar applications or FARMS) to modeling solar irradiance. It is not known how this would perform compared to other modeling techniques (*e.g.*, approaches based on time series). It should also be stressed that our analysis only compares the use of the NSRDB to measured solar-insolation data for purposes of capacity-value estimation. Capacity-value estimatation is not the sole use of solar-insolation data, and our analysis should not be construed as making any claims about the accuracy of other types of solar analyses using the NSRDB.

Despite these concerns, these data issues can be addressed with time. As PV systems become more common and widely deployed, measured *solar-output* (as opposed to only solar-insolation) data could become more widely available. This may reduce our reliance on modeled data. Other data-quality issues, for instance related to load reporting, can be addressed by FERC and other agencies that collect such data by developing detailed standards on data reports (*e.g.*, what load measurements to report and how time zone and daylight savings are accounted for).

Our analysis also has some important implications for remunerating solar generators for capacity value. Many organized electricity markets have mechanisms that compensate solar generators based on historical production data. Our analysis suggests that these schemes can, on average, get capacity payments correct if they use hourly metered solar production, so long as a sufficient number of years of historical data are included in determining payments. Otherwise, if only a small number of years' data are used, payments based on hourly metered solar output may not properly remunerate solar generators for the reliability benefits that they provide the system.

An interesting topic for future research is to better understand the tradeoffs between the number of years of solar-insolation data used in capacity-value estimation and the temporal granularity of those data. We only compare hourly averaged and one-minute data. Moreover, our analysis only goes as far as finding that using at least nine years of hourly average solar-insolation data results in relatively good ELCC estimates (compared to using less than two years of data). This result raises the question of understanding the marginal value of having additional years of hourly averaged solar-insolation data for the purposes of capacity-value estimation. Another related question is how many years of data would be needed if solar-insolation data are recorded at other temporal granularities (*e.g.*, 10-minute data). Such analyses will require more data covering more locations and years than we currently have access to, in order to ensure that the findings are robust. As more measured solar-output data become available, such analyses should become feasible.

## Acknowledgment

The authors thank D. Heimiller and A. Weekley for invaluable help in gathering the data required for this analysis. A. Weekley also provided a great deal of assistance with data-quality and PVWatts support. A. Sorooshian, A. Lopez, A. Dobos, the editors, and two anonymous reviewers provided some very helpful discussions, suggestions, and comments. Any errors in this work, of course, fall squarely on the authors' shoulders.

## References

[1] J. Haslett and M. Diesendorf, "The capacity credit of wind power: A theoretical analysis," *Solar Energy*, vol. 26, pp. 391–401, 1981.

[2] L. Söder and J. Bubenko, "Capacity credit and energy value of wind power in hydro-thermal power system," in *Proceedings of the 9th Power Systems Computation Conference*, Cascais, Portugal, 30 August-4 September 1987.

[3] M. R. Milligan, "Measuring Wind Plant Capacity Value," National Renewable Energy Laboratory, Tech. Rep. NREL/TP-441-20493, 1996.

[4] M. R. Milligan and B. Parsons, "Comparison and Case Study of Capacity Credit Algorithms for Intermittent Generators," National Renewable Energy Laboratory, Tech. Rep. NREL/CP-440-22591, March 1997.

[5] ——, "Comparison and Case Study of Capacity Credit Algorithms for Wind Power Plants," *Wind Engineering*, vol. 23, pp. 159–166, May 1999.

[6] M. R. Milligan and T. Factor, "Optimizing the Geographic Distribution of Wind Plants in Iowa for Maximum Economic Benefit and Reliability," *Wind Engineering*, vol. 24, pp. 271–290, July 2000.

[7] M. R. Milligan, "Modeling Utility-Scale Wind Power Plants Part 2: Capacity Credit," National Renewable Energy Laboratory, Tech. Rep. NREL/TP-500-29701, March 2002.

[8] M. A. H. El-Sayed, "Substitution potential of wind energy in Egypt," *Energy Policy*, vol. 30, pp. 681–687, June 2002.

[9] E. P. Kahn, "Effective Load Carrying Capability of Wind Generation: Initial Results with Public Data," *The Electricity Journal*, vol. 17, pp. 85–95, December 2004.

[10] M. R. Milligan and K. Porter, "The Capacity Value of Wind in the United States: Methods and Implementation," *The Electricity Journal*, vol. 19, pp. 91–99, March 2006.

[11] H. Shiu, M. R. Milligan, B. Kirby, and K. Jackson, "California Renewables Portfolio Standard Renewable Generation Integration Cost Analysis Mulit-Year Analysis Results and Recommendations–Final Report," California Energy Commission, PIER Publication number CEC-500-2006-064, 2006.

[12] K. R. Voorspools and W. D. D'haeseleer, "Critical evaluation of methods for wind-power appraisal," *Renewable and Sustainable Energy Reviews*, vol. 11, pp. 78–97, January 2007.

[13] C. D'Annunzio and S. Santoso, "Noniterative Method to Approximate the Effective Load Carrying Capability of a Wind Plant," *IEEE Transactions on Energy Conversion*, vol. 23, pp. 544–550, June 2008.

[14] M. R. Milligan and K. Porter, "Determining the Capacity Value of Wind: An Updated Survey of Methods and Implementation," National Renewable Energy Laboratory, Tech. Rep. NREL/CP-500-43433, June 2008.

[15] L. Söder and M. Amelin, "A review of different methodologies used for calculation of wind power capacity credit," in *2008 IEEE Power and Energy Society General Meeting*. Pittsburgh, PA, USA: Institute of Electrical and Electronics Engineers, 20-24 July 2008.

[16] C. Ensslin, M. R. Milligan, H. Holttinen, M. O'Malley, and A. Keane, "Current Methods to Calculate Capacity Credit of Wind Power, IEA Collaboration," in *2008 IEEE Power and Energy Society General Meeting*. Pittsburgh, PA, USA: Institute of Electrical and Electronics Engineers, 20-24 July 2008.

[17] M. Amelin, "Comparison of Capacity Credit Calculation Methods for Conventional Power Plants and Wind Power," *IEEE Transactions on Power Systems*, vol. 24, pp. 685–691, May 2009.

[18] B. Hasche, A. Keane, and M. O'Malley, "Capacity Value of Wind Power, Calculation, and Data Requirements: the Irish Power System Case," *IEEE Transactions on Power Systems*, vol. 26, pp. 420–430, February 2011.

[19] A. Keane, M. R. Milligan, C. J. Dent, B. Hasche, C. D'Annunzio, K. Dragoon, H. Holttinen, N. Samaan, L. Söder, and M. O'Malley, "Capacity Value of Wind Power," *IEEE Transactions on Power Systems*, vol. 26, pp. 564–572, May 2011.

[20] R. Perez, R. Seals, and R. Stewart, "Assessing the load matching capability of photovoltaics for US utilities based upon satellite-derived insolation data," in *Conference Record of the Twenty Third IEEE Photovoltaic Specialists Conference*. Louisville, KY, USA: Institute of Electrical and Electronics Engineers, 10-14 May 1993, pp. 1146–1151.

[21] R. Perez, R. Margolis, M. Kmiecik, M. Schwab, and M. Perez, "Update: Effective Load-Carrying Capability of Photovoltaics in the United States," National Renewable Energy Laboratory, Tech. Rep. NREL/CP-620-40068, June 2006.

[22] R. Perez, M. Kmiecik, J. Schlemmer, L. Root, K. Moore, and P. Stackhouse, "Evaluation of PV generation capacity credit forecast on day-ahead utility markets," in *Solar 2007, Including Proceedings from the 36th ASES Annual Conference, 32nd National Passive Solar Conference, and 2nd Renewable Energy Policy and Marketing Conference "Sustainable Energy Puts America to Work"*, R. Campbell-Howe, Ed. American Solar Energy Society, September 2007, pp. 75–80.

[23] R. Perez, M. Taylor, T. Hoff, and J. P. Ross, "Reaching Consensus in the Definition of Photovoltaics Capacity Credit in the USA: A Practical Application of Satellite-Derived Solar Resource Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, pp. 28–33, March 2008.

[24] T. Hoff, R. Perez, J. P. Ross, and M. Taylor, "Photovoltaic Capacity Valuation Methods," Solar Electric Power Association, SEPA Report 02-08, May 2008.

[25] S. Pelland and I. Abboud, "Comparing Photovoltaic Capacity Value Metrics: A Case Study for the City of Toronto," *Progress in Photovoltaics: Research and Applications*, vol. 16, pp. 715–724, December 2008.

[26] *An Effective Load Carrying Capability Analysis for Estimating the Capacity Value of Solar Generation Resources on the Public Service Company of Colorado System*, Xcel Energy Services, Inc., February 2009.

[27] S. H. Madaeni, R. Sioshansi, and P. Denholm, "Comparing Capacity Value Estimation Techniques for Photovoltaic Solar Power," *IEEE Journal of Photovoltaics*, vol. 3, pp. 407–415, January 2013.

[28] ——, "Estimating the Capacity Value of Concentrating Solar Power Plants: A Case Study of the Southwestern United States," *IEEE Transactions on Power Systems*, vol. 27, pp. 1116–1124, May 2012.
[29] ——, "Estimating the Capacity Value of Concentrating Solar Power Plants with Thermal Energy Storage: A Case Study of the Southwestern United States," *IEEE Transactions on Power Systems*, vol. 28, pp. 1205–1215, May 2013.
[30] M. Sengupta, A. Habte, P. Gotseff, A. Weekley, A. Lopez, C. Molling, and A. Heidinger, "A Physics-Based GOES Satellite Product for Use in NREL's National Solar Radiation Database," National Renewable Energy Laboratory, Tech. Rep. NREL/CP-5D00-62237, July 2014.
[31] M. Sengupta, A. Habte, P. Gotseff, A. Weekley, A. Lopez, M. Anderberg, C. Molling, and A. Heidinger, "A Physics-Based GOES Product for Use in NREL's National Solar Radiation Database," National Renewable Energy Laboratory, Tech. Rep. NREL/CP-5D00-62776, September 2014.
[32] A. P. Dobos, "PVWatts Version 5 Manual," National Renewable Energy Laboratory, Tech. Rep. NREL/TP-6A20-62641, September 2014.
[33] R. Billinton and R. N. Allan, *Reliability Evaluation of Power Systems*. Boston: Pitman Advanced Publishing Program, 1984.

**Dhruv Gami** holds a B.S. in mechanical engineering from the University of Mumbai and an M.S. in industrial and systems engineering from The Ohio State University.

His research focuses on capacity value assessments of solar and wind generators.

**Ramteen Sioshansi** (M'11–SM'12) holds the B.A. degree in economics and applied mathematics and the M.S. and Ph.D. degrees in industrial engineering and operations research from the University of California, Berkeley, and an M.Sc. in econometrics and mathematical economics from The London School of Economics and Political Science.

He is an associate professor in the Integrated Systems Engineering Department at The Ohio State University, Columbus, OH. His research focuses on renewable and sustainable energy system analysis and the design of restructured competitive electricity markets.

**Paul Denholm** (M'11–SM'12) holds the B.S. degree in physics from James Madison University, the M.S. in instrumentation physics from the University of Utah, and the Ph.D. in land resources/energy analysis and policy from the University of Wisconsin-Madison.

He is a senior analyst in the Strategic Energy Analysis Center at the National Renewable Energy Laboratory, Golden, CO. His research interests are in the effects of large-scale renewable energy deployment in electric power systems, and renewable energy enabling technologies such as energy storage and long distance transmission.