Weather Data for Energy Analytics

From Modeling Outages and Reliability Indices to Simulating Distributed Photovoltaic Fleets

By Jonathan Black, Alex Hofmann, Tao Hong, Joseph Roberts, and Pu Wang

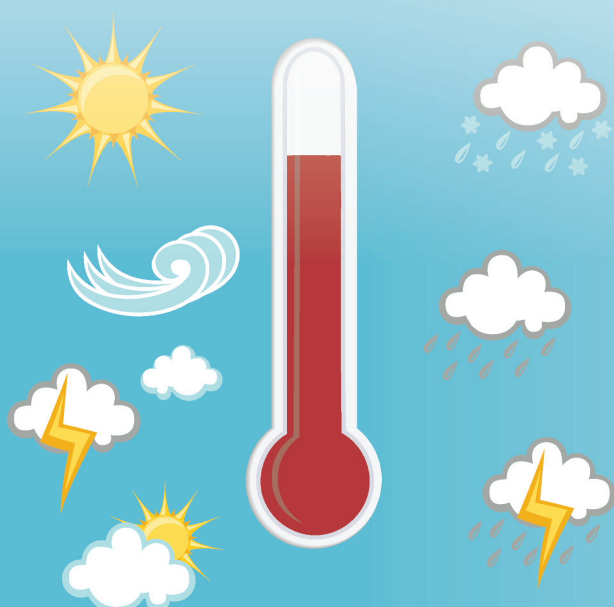

IMAGE LICENSED BY INGRAM PUBLISHING

WEATHER IMPACTS VIRTUALLY ALL FACETS OF our daily life. As a result, many business sectors are affected by weather conditions, and the power industry is no exception. Weather is a major influencer on system reliability and a key driver of both power supply and demand. In this article, we will demonstrate novel uses of weather data for energy analytics via two utility applications. We first use easily accessible weather data together with regression analysis to model distribution outages and construct a probabilistic view of reliability indices that helps reveal a utility's reliability trend. We then use high-resolution, commercial-grade weather data to develop realistic simulations of anticipated behind-the-meter photovoltaic (PV) fleets.

Power Distribution Reliability

Reliability is an important subject in the electric power industry. Outages at distribution systems tie directly to utility revenues and customer satisfaction. A less-reliable grid contributes to greater revenue loss and lower customer satisfaction. As a result, utility reliability is being continuously evaluated by boards, executives, engineers, operators, customers, and regulators. While most stakeholders would not be willing to pay for the large capital, operations, and maintenance expenditures required to keep a super-reliable grid, many would also not be willing to accept poor reliability. Therefore, the electric power industry must continuously balance the costs and benefits of improving system reliability.

Digital Object Identifier 10.1109/MPE.2018.2801442
Date of publication: 18 April 2018

To measure and communicate reliability, utility stakeholders often use reliability indices as a good measure of utility system health. Just as many complex systems have their own level of health measurements, reliability indices let everyone know if the utility system is getting better or worse over time. Reliability indices help power engineers and other operations personnel see and show the interconnected nature of the many independent system components that make up an electric distribution system. To quantify the level of system reliability, power engineers have adopted and invented many reliability indices.

A commonly used index for power distribution systems is System Average Interruption Duration Index (SAIDI), which indicates the total interruption duration for the average customer during a predefined period. The formula to calculate the index is

$$\text{SAIDI} = \frac{\text{Sum of all Customer Interuption Durations}}{\text{Total Number of Customers Served}}.$$

The numerator can be thought of as the total number of outage minutes experienced by all customers for a predefined period, which is typically one year. SAIDI is commonly given in minutes and, generally speaking, represents the average outage minutes experienced for a customer connected to the utility's system. Daily and monthly SAIDI values may be used for a more detailed and timely analysis of system reliability. It should be noted that some utilities do not have enough outages to calculate a SAIDI for every day in the year.

A high SAIDI value indicates poor system reliability. Although SAIDI, together with many other reliability indices, can be used to describe customer experience, it does not offer many actionable insights into the daily operations of utilities. This is because the SAIDI calculation is significantly impacted by severe weather conditions that take out large numbers of customers for an extended period of time.

Let's consider two power companies, P_1 and P_2, each serving 1 million customers. P_1 had 50,000 customers who experienced 300 6-min interruptions within one year. The annual SAIDI value of P_1 can be calculated as $6 \times 300 \times 50{,}000/1{,}000{,}000 = 90$ min. P_2 was hit by a major hurricane that caused a systemwide outage of 90 min, while no other interruptions occurred during the year. The annual SAIDI value of P_2 would be $90 \times 1 \times 1{,}000{,}000/1{,}000{,}000 = 90$ min.

Although both companies share exactly the same SAIDI value, the potential and strategy for them to improve their reliability can be quite different. P_1 can improve its daily operations to reduce the number of interruptions and the duration of each interruption. On the other hand, the major hurricane that hit P_2 could be a rare event that is out of anyone's control. If P_2 chose to do exactly the same practice as before but did not experience another major hurricane in the next year, its SAIDI value would be very small, indicating a very good reliability. If P_2 chose to improve its storm restoration practice, the duration of the interruption caused by a similar hurricane could be shortened.

The Beta Method from IEEE Standard 1366

To study major events separately from daily operations and to reveal trends in daily operations that can be hidden by the statistical effect of major events, the IEEE Working Group on Distribution Reliability included the section "Major Event Day Classification" in IEEE Standard 1366, *IEEE Guide for Electric Power Distribution Reliability Indices.* According to this standard, the daily SAIDI value is assumed to follow a log-normal distribution. After taking the natural logarithm transformation of the daily SAIDI values, the days falling beyond the 2.5 standard deviations to the right of the mean are classified as major event days. The mean and standard deviation are denoted by α and β, respectively. This method is also known as the Beta method, the 2.5β method, or IEEE 2.5 Beta.

To appeal to a broad audience, the Beta method keeps its simplicity by avoiding any predictive modeling effort in the SAIDI reporting process other than taking the logarithm transformation and calculating the mean and standard deviation. The Beta method is better at "revealing trends in daily operations that would be hidden by the large statistical effect of major events" than the original all-inclusive SAIDI calculation. However, due to the exclusion of weather information, it is unable to answer some important and frequently asked what-if questions, such as "What would be the SAIDI values if the weather this year was the same as last year?" Consequently, the Beta method has limited capability to reveal trends in reliability, including any improvement or degradation that could be hidden by weather conditions not classified as major events. Moreover, the assumption that daily SAIDI always follows the lognormal distribution is yet to be verified.

Have utility operations been improving over the last few years? The power industry has been trying to answer this question since the late 20th century. While IEEE Standard 1366 has made great progress by bringing a simple tool to the reporting process, there is still a long way to go. In this article, we will demonstrate a predictive modeling approach that can provide a probabilistic view of reliability indices to help further reveal a utility's reliability trend. The case study is based on a small local distribution company (LDC) in the United States.

Descriptive Outage and Reliability Analytics

The LDC serves approximately 30,000 customers. The case study uses nine years of outage information collected from this LDC. Interruptions occurred in 1,964 of the 3,288 days covered by the data set. After taking the log transformation on the daily SAIDI values, we obtain the mean (α) of these nine years of log transformed daily SAIDI values as –2.80 with a standard deviation (β) 1.88. Note that the unit of the

table 1. Summary statistics of daily SAIDI values of a local distribution company, 2008–2016.

	Original			Simplified Beta			IEEE Beta			
Year	Sum	Max	Days	Sum	Max	Days	Threshold	Sum	Max	Days
2008	70	12.2	271	58	5.7	270	1.31	38	1.9	266
2009	70	10.0	221	60	5.4	220	1.31	41	3.4	216
2010	76	22.4	232	54	4.7	231	1.47	49	3.1	230
2011	2447	1793.0	224	47	3.3	216	1.44	47	3.3	216
2012	105	56.7	215	48	6.4	214	1.81	42	3.4	213
2013	51	7.5	216	43	6.3	215	1.77	37	4.0	214
2014	83	14.4	186	69	4.9	185	1.81	69	4.9	185
2015	214	133.6	205	53	4.2	202	1.95	53	4.2	202
2016	167	91.2	194	55	5.1	191	2.09	63	7.3	192
Trend	−26.0	—	—	−0.1	—	—	—	2.8	—	—

Notes
Original: statistics derived from the original data without taking out major event days.
Simplified Beta: statistics derived from the sample data with major event days excluded using the simplified Beta method.
IEEE Beta: statistics derived from the sample data with major event days excluded using the Beta method of IEEE Standard 1366.
Sum: annual SAIDI (in minutes) calculated as the sum of daily SAIDI values.
Max: maximum daily SAIDI (in minutes) during the period being included in the calculation.
Days: number of days being included in the calculation.
Threshold: a rolling number ($\alpha + 2.5\beta$) by year used to exclude major event days as guided by IEEE Standard 1366.
Trend: the trend (in minutes/year) of annual SAIDI during the nine-year period.

original daily SAIDI value (in minutes) is gone after taking the log transformation. Twenty days of the nine-year period fall beyond 2.5 standard deviations. We denote this method of calculating the threshold based on the entire data set (a nine-year period in this case) as the simplified Beta method. We also did the exact calculation as guided by the Beta method of IEEE Standard 1366, resulting in a rolling threshold (e.g., $\alpha + 2.5\beta$) calculated once a year based on the five most recent years of historical values of daily SAIDI values. When fewer than five years of data is available, the entire history was used. For instance, to calculate the threshold for 2011, we used three years of history (2008–2010).

Table 1 lists the sum and maximum of daily SAIDI values and the number of days with interruptions for each year based on the three methods. We also calculate the trends of the annual SAIDI (sum of the daily SAIDI values by year). When the major event days are not excluded in the annual SAIDI calculation, the reliability shows a very strong decreasing trend (−26 min/year). On the other hand, the simplified Beta method shows a flat trend (−0.1 min/year), while the IEEE Beta method presents a moderate increasing trend (2.8 min/year). Again, due to the lack of weather input in the calculation of either method, the trend resulting from each one does not indicate the improvement or degradation of this LDC's daily operations.

The surge of outages in 2011 was due to a major hurricane. Although other hurricanes affected the region during the nine-year period, some of which also resulted in extended interruptions, none were as damaging to the power grid as the one in 2011. Because of this surge, the major event day threshold was bumped up to 1.81 in 2012. According to the Beta method, this surge stays in the calculation of α and β for five years, so the thresholds of 2012–2016 are higher than those of 2008–2011. The aforementioned strong increasing trend in the annual SAIDI calculated using the Beta method is largely due to this step-up in the threshold for major event day. This effect was also mentioned in IEEE Standard 1366 as an ongoing work "undertaken to develop objective methods for identifying and processing catastrophic events (to eliminate the noted effect on the reliability trend)."

Figure 1 shows a more detailed view of the daily SAIDI data. Part (a) shows the daily SAIDI values of the entire nine-year period (2008–2016), where the surge in 2011 makes the other days look like a flat line. Part (b) excludes the top 20 days with the highest daily SAIDI values. The line plot of the remaining days does not show any obvious seasonal patterns either. Note that the distribution of the log-transformed daily SAIDI values does not pass commonly used normality tests, such as Shapiro–Wilk, Anderson–Darling, and Kolmogorov–Smirnov. Simply speaking, the daily SAIDI values of this LDC do not follow a lognormal distribution. In practice, among the LDCs we work with, we can hardly find any that have the daily SAIDI following a lognormal distribution. In the remaining part of

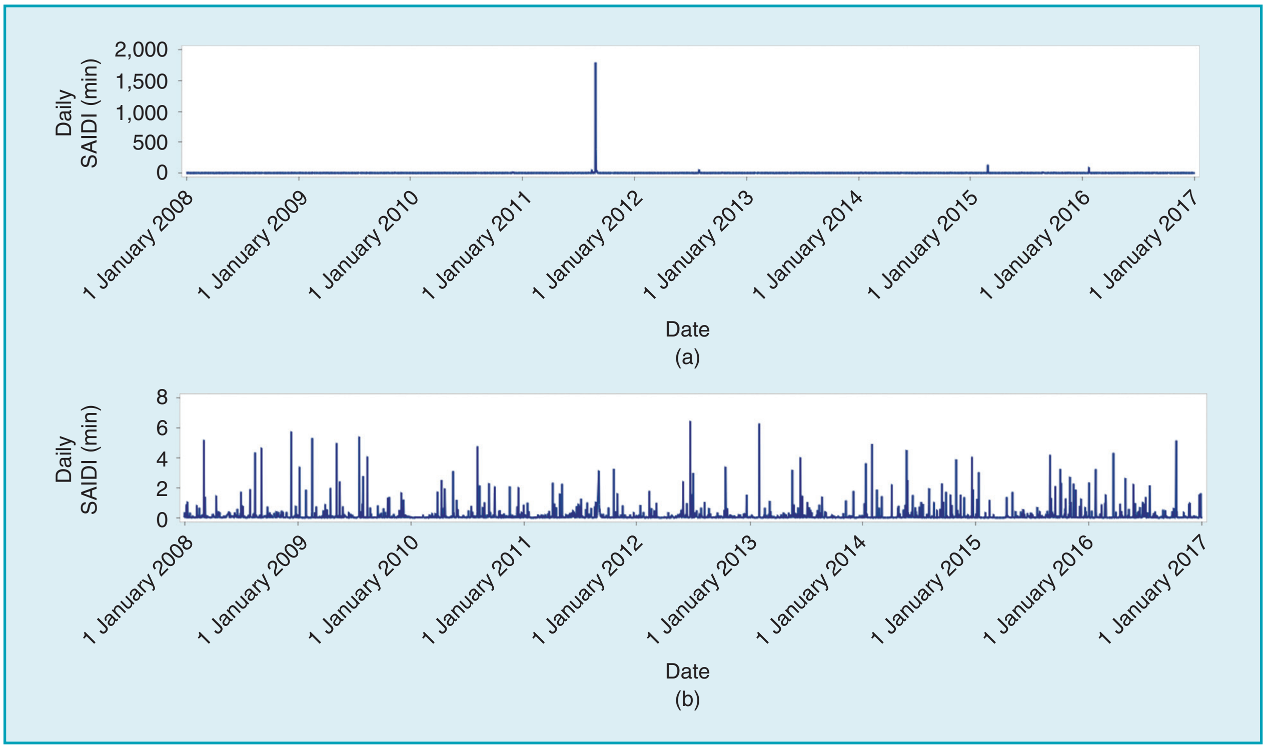

figure 1. Line plots of daily SAIDI values: (a) all SAIDI values and (b) the 20 highest SAIDI values are excluded.

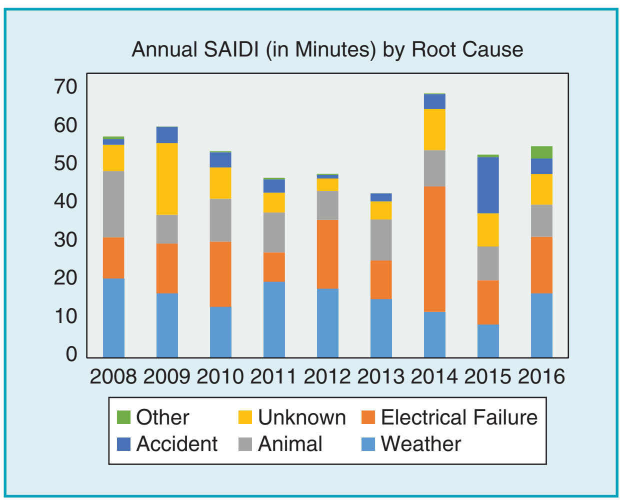

figure 2. The annual SAIDI by root cause. Contributions to the sum of SAIDI values over the nine-year period are weather (29%), electrical failure (28%), animal (19%), unknown (15%), accident (8%), and other (1%).

this case study, we will adopt the simplified Beta method to exclude 20 days from the nine-year period for further analysis because that helps us answer questions about the utility's underlying reliability trend by uniformly removing extreme events across the entire data set.

Many factors may lead to a distribution outage. The stacked bar chart in Figure 2 shows the sum of annual SAIDI values for each outage root cause during the past nine years. The legend lists the categories in the ascending order of the SAIDI share during the nine years. The largest category is weather, which includes subcategories of tree, storm, and wind. The other categories, such as electrical failure, animal, unknown, and accident, do not have weather as a main driving factor. In this article, we focus on weather-related outages, which contribute to about one third (29% in this case) of the SAIDI value over the nine-year period.

Predictive Outage and Reliability Analytics

We can further relate the root causes of outages to some weather variables. For instance, we can use wind speed and gust to describe wind; we can use precipitation and wind speed to describe a storm. Although tree faults are not always weather driven, many are due to windblown trees falling on overhead lines. Sometimes a moist tree branch bridges two conductors and gradually dries out due to the small current, which can eventually lead to a short circuit. Weather conditions such as wind, precipitation, humidity, and temperature are often the real root causes of the tree faults. Therefore, we merge tree faults with the other weather-related outages in this article to conduct the analysis. Out of 3,268 days, or nine full years minus the top 20 SAIDI days, this category includes 375 days with weather-related outages.

The major weather service providers in the United States have been reporting weather variables hourly for many major weather stations, including the one close to this LDC. In this case study, we take the following 11 quantitative weather variables: temperature, dew-point temperature, heat index, feels-like temperature (the combination of the heat index and the wind-chill factor), wind chill, wind-chill energy, wet bulb temperature, relative humidity, wind speed, speed gust, and precipitation. To match them up with the daily SAIDI data, we convert the original records of the first ten weather variables to daily values by taking the maximum, mean, and minimum of the hourly values. For precipitation, we sum up the hourly records by day to obtain the daily precipitation. In addition, we include four dummy variables to describe the weather events of the day, such as fog, rain, snow, and thunderstorm. In total, we use 37 candidate-independent variables to build a model that predicts the dependent-variable, weather-related daily SAIDI.

We use the regression procedure of the SAS STAT package to conduct the model selection process. After trying several different options, such as forward selection, backward elimination, stepwise, maximum R^2 improvement, adjusted R^2 selection, and Mallows' C_p selection, we select the following 11 variables: maximum dew-point temperature, minimum dew-point temperature, daily precipitation, average relative humidity, average speed gust, minimum speed gust, thunderstorm, maximum wind chill, average wind chill, average wind speed, and minimum wind speed.

Calibrating this model using the 3,268 days of historical data, we obtain predicted values of the weather-related daily SAIDI for the LDC. Figure 3 shows the annual aggregates of the actual and predicted SAIDI values. Both SAIDI curves show a decreasing trend, with the slope of −0.8 and −0.4 min/year for the actual and predictive SAIDI, respectively. As mentioned earlier, the slope of the actual SAIDI curve indicates the trend of customer experience regarding interruptions rather than the improvement of daily operations. Alternatively, one could assume that the LDC has engaged in relatively consistent daily operational practices over the nine-year period. Under this assumption, the decreasing slope indicates that the weather, rather than improved operational practices, has contributed to the trend of increased reliability over the nine-year period.

However, maintaining constant daily operational practices over a long period (i.e., nine years) is quite a strong assumption. Whether the utility has improved its daily operations is the question we have been trying to answer. To further investigate this issue, we conduct a simulation to offer a probabilistic view of the weather-related SAIDI. We first shrink the training period to three years. We then fit the same 11-variable model using the three-year training data and predict the daily SAIDI values of the nine-year period. By repeating this training window on a rolling basis, we can obtain seven groups of predicted SAIDI profiles, where each group includes nine years of predicted daily SAIDI values representing three years of operational practice under nine different years of weather scenarios from 2008 to 2016. Figure 4 shows the annual aggregates of the predicted SAIDI values that are caused by weather. The nine dashed lines represent the nine weather scenarios. The red line in the middle connects the median (normalized) values of the nine years within each group.

Taking the first three-year training period (2008–2010) for example, each of the corresponding nine dots represents the predicted SAIDI values if those three years were experiencing the same weather as one of the nine-year period (2008–2016). The red dot in the middle is the median value that can be used to indicate the daily operations during this three-year period normalized based on nine years of weather history. Overall, the slopes from the dashed lines range from −0.1 to −1 min/year. In other words, given the same weather from any year of the nine-year period, the weather-related SAIDI of this utility shows a decreasing trend from 2008 to 2016. The normalized curve has a slope of −0.7 min/year, indicating weather-normalized improvement of the weather-related SAIDI at this LDC.

Have utility operations been improving over the last nine years? At this stage, we have answered the question for the weather-related outages, which assemble the largest share of the total SAIDI for the small LDC in this case study. Since weather

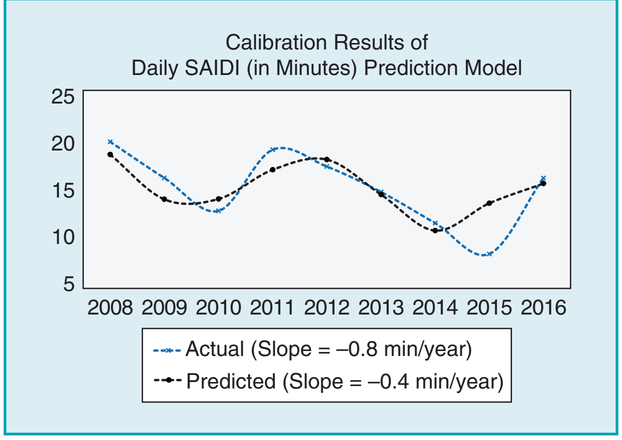

figure 3. The calibration results of the model that predicts the weather-related daily SAIDI.

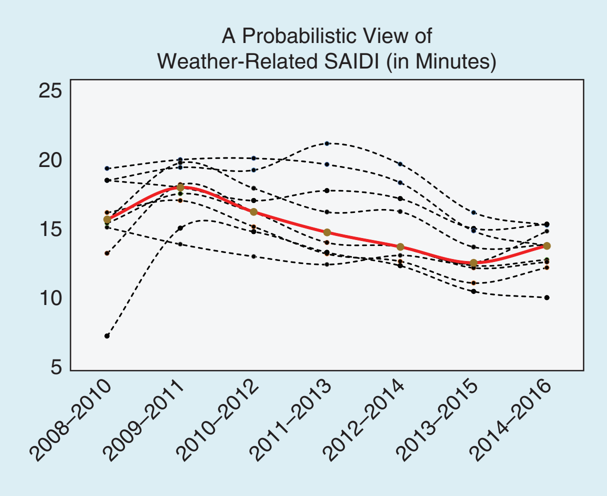

figure 4. A probabilistic view of the weather-related SAIDI over a nine-year period.

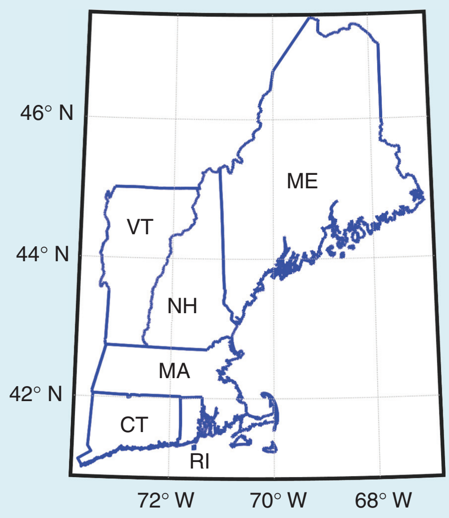

figure 5. A map of the New England states. CT: Connecticut; MA: Massachusetts; ME: Maine; NH: New Hampshire; RI: Rhode Island; VT: Vermont.

table 2. A summary of total installed PV nameplate capacity and number of installations, by state, by 31 December 2017. (Data provided to ISO New England by New England distribution utilities.)

State	Installed Nameplate Capacity (MW_{AC})	Number of Installations
Connecticut	366	29,512
Massachusetts	1,602	78,047
Maine	33	3,598
New Hampshire	70	7,330
Rhode Island	62	4,148
Vermont	257	9,773
New England	**2,390**	**132,408**

is considered a major root cause of distribution outages in many other large and small utilities, this type of analysis can offer organizations an improved understanding of their underlying reliability trend and a clearer direction in reliability planning.

Behind-the-Meter Solar PVs

Solar PV deployments have grown substantially in recent years and are making an increasing but still relatively new contribution to the resource mix within modern power systems. A large share of PVs is distributed, or small scale (i.e., having a nameplate of several megawatts or lower) and installed behind the meter (BTM), embedded with load on the distribution system. This aspect of PVs means that it is not observable to independent system operators/regional transmission organizations (ISOs/RTOs) such as ISO New England (Figure 5). Overall, the proliferation of PVs has led to a growing industry emphasis on the need for reliable and efficient integration of these newer resources into power grids.

In areas anticipating high penetrations of BTM PVs, ISOs/RTOs need to prepare for the resulting changes to their system, such as the possible need for greater system flexibility and the related impacts to generation unit commitment and dispatch. In such cases, system planning studies must consider the anticipated load impacts from BTM PVs to determine the timing and magnitude of concerns related to these types of integration challenges. We discuss how it is now possible to develop realistic simulations of anticipated BTM PV fleets using publicly available data and tools to be able to perform detailed analysis and plan for their arrival. To demonstrate these capabilities, we describe a simulation of a BTM PV fleet in New England over the historical period of 1998–2014 using high-resolution weather data developed by the National Renewable Energy Laboratory (NREL). Results highlight the value in using these emerging simulation capabilities to discover the effects of localized weather on BTM PVs over short timescales as well as the long-term resource characteristics of BTM PVs for a region, which are both important considerations for power system planning and operation.

Significant PV-installed cost reductions, federal incentives, and state policies together have stimulated significant growth of distributed PVs in the United States. In New England, these factors have led to the growth in aggregate PV nameplate capacity from lower than 50 MW_{ac} at the beginning of 2010, to 2,390 MW_{ac} at the end of 2017, the vast majority of which is BTM. Table 2 lists a summary of distribution utility data provided to ISO New England, including the state-by-state PV nameplate capacity and number of installations in service by the end of 2017. Figure 6 is a heat map illustrating the spatial distribution of PVs in New England at the end of 2017 and at the municipal level.

Long-Term Load Forecasting with BTM PVs

Long-term, probabilistic load forecasts are a fundamental input to system planning studies, where the objective is often to guide infrastructure investment to yield reliability and market efficiency. However, while the general effects of weather on the load in a given area are well understood, the conjoined influence of weather on both electricity demand and solar PV performance is much less understood. Furthermore, rare or otherwise extreme weather conditions such as heat waves occur infrequently, making the accurate prediction of the future load effects of these important conditions a challenging but necessary exercise. To ensure that their models capture these extreme conditions, many long-term load forecasters use a decade or longer of historical weather and load data in the development of their forecasts.

An in-depth understanding of the load impacts of BTM PV proliferation requires relatively long-term, contemporaneous records of load, weather, and solar data. Historically, there were often insufficient or incomplete solar data to perform forecasting and analysis of this caliber. However, with the recent advent of comprehensive, high-quality meteorological data sets such as NREL's National Solar Radiation Database (NSRDB), the electric power industry now has considerably more information to better prepare for future integration challenges.

Important considerations regarding BTM PV's prospective load reduction impacts include changes in the magnitude and timing of forecasted peak loads; changes in load ramping and volatility; the frequency, timing, and duration of light load conditions when load is low and BTM PV output is high; and the overall impacts to typical load shapes. Since all of these considerations depend on the penetration of BTM PVs, they require scenario-building and in-depth analysis of resulting net loads (i.e., load minus BTM PV output).

Characteristics of BTM PV Fleets

BTM PV fleets include a large population of individual systems, perhaps totaling in the hundreds of thousands, installed heterogeneously throughout a broad geographical region. For many regions such as New England, measured performance data are unavailable for much of the BTM PV fleet. Metadata concerning the exact design and location of all BTM PV systems are also often relatively limited. This lack of information is one of the challenges with respect to understanding the seasonally and diurnally varying influence of BTM PVs on load given a variety of weather conditions and as BTM PV fleets become larger.

The realistic simulation of a decentralized fleet of small-scale, BTM PV systems should include all of the following:

- ✔ comprehensive, high-resolution weather data that includes the main drivers of PV performance
- ✔ some level of detailed information concerning the location and design of individual systems
- ✔ simulation tools capable of modeling PV system performance given a variety of weather conditions and a diversity of individual system design characteristics.

In cases where detailed design characteristics of PV systems are unknown, it is possible to make informed assumptions regarding these inputs and, ideally, to then validate the simulation results against measured data from installed BTM PV systems. We will discuss the development of such assumptions and validation of results in the New England test case next.

The National Solar Radiation Database

Ground-based solar irradiance measurements needed for PV modeling are sparsely available due to the expense and technical difficulty of installing and maintaining the requisite solar radiation measurement network. To fill in this gap, the U.S. Department of Energy has funded the ongoing development of the NSRDB by NREL. In support of the Department of Energy's SunShot project goals, NREL developed and released the third generation of NSRDB, a comprehensive gridded solar irradiance data set covering the continental United States and other parts of North, Central, and South America. This NSRDB release has significantly advanced the capability of users to perform in-depth solar resource assessments within the covered areas. NREL, in collaboration with the University of Wisconsin and the U.S. National Oceanic and Atmospheric Administration, made the most recent release of the NSRDB available in late 2015. The data set covers the years 1998–2014 at a 30-min temporal resolution and is gridded at 4-km spatial resolution. This means that there are almost 12,000 weather grid points within New England.

NREL used a physical approach to model surface radiation in detail by retrieving cloud and aerosol information from Geostationary Operational Environmental Satellite imagery and subsequently used this information in a radiative transfer model. Resulting NSRDB data includes the three components of irradiance: direct normal, diffuse horizontal, and global horizontal. Other ancillary weather data, including dry-bulb and dew-point temperature, wind speed, wind direction, relative humidity, and atmospheric pressure, are downscaled from the National Aeronautics and Space Administration's Modern Era Retrospective-Analysis for Research and Applications reanalysis data set.

BTM PV Fleet Simulation

The BTM PV fleet was simulated using NREL's System Advisor Model (SAM), a technoeconomic performance model used as a desktop application and in the SAM software development kit (SDK). SAM combines PV module and inverter submodels to calculate a PV system's performance given a weather

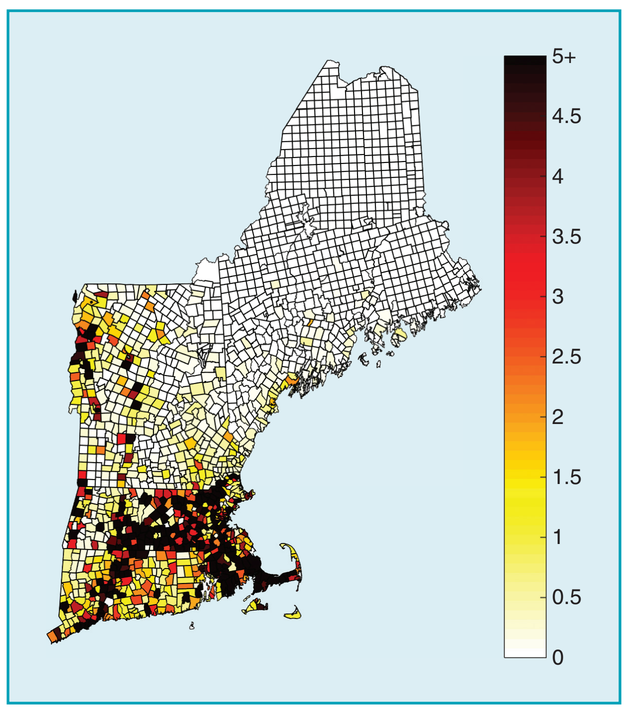

figure 6. A heat map showing the spatial distribution of installed nameplate capacity (in MW_{AC}) as of 31 December 2017.

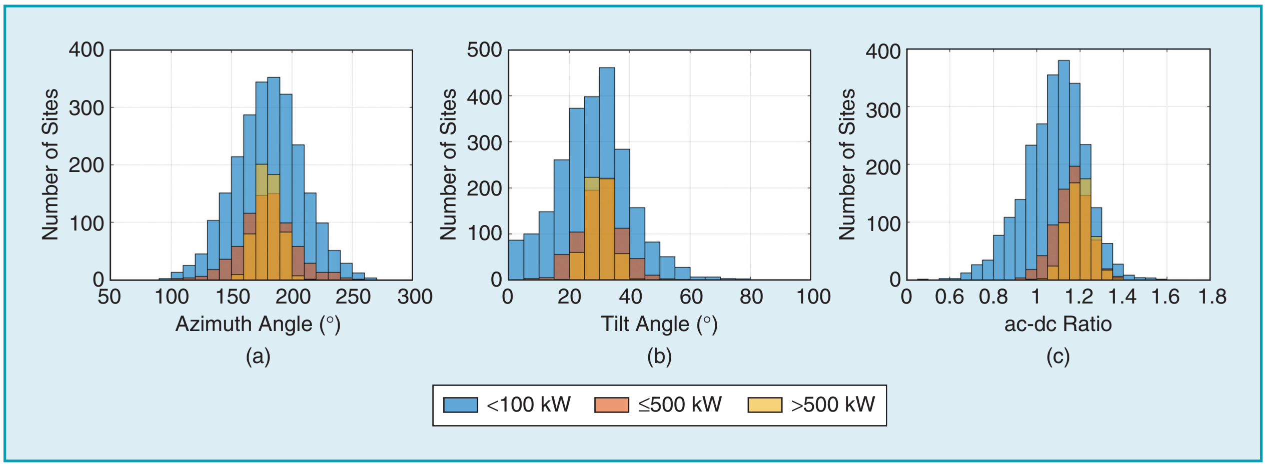

figure 7. Distributions of simulated BTM PV system design parameters for Connecticut: (a) azimuth angle, (b) tilt angle, and (c) dc-to-ac ration.

file and input data describing the system's characteristics. The SAM Simulation Core SDK enables the user to access the SAM application programmatically. SAM simulations used the most recent version of PVWatts (version 5). MATLAB was used for all simulation work; however, the SDK also contains language wrappers for C#, Java, and Python.

For the New England test case, simulations include more than 41,000 unique PV systems located throughout the region to capture diversity in the influence of both weather and PV system design. Designs include project sizes ranging from 4 KW_{dc} to 2,500 KW_{dc} and a variety of azimuth angles (also known as array orientation), tilt angles, and ac-to-dc ratios. Since tracking PV systems are rare in New England, all simulated systems were assumed to be fixed axis. In the absence of known values for these parameters, a few high-level assumptions that reflect general observations of BTM PV installations guided parameter selection. First, the total population of each parameter's values reflects a normal distribution centered on approximated expected values. These expected values are 180° for azimuth angle (i.e., solar south), 1.2 for the dc-to-ac ratio, and 30° for system tilt. Second, parameter distributions reflect a tendency for their values associated with larger PV systems to be closer to expected values (i.e., their values exhibit less variance), and those of smaller systems reflect a tendency for greater diversity (i.e., to exhibit more variance). These guiding principles reflect the observation that smaller rooftop systems are often installed in less-optimal conditions, whereas larger greenfield PV developments are more likely to be optimally designed to maximize revenue, which reflects performance-based incentives. In addition to their use for addressing grid integration issues, it is worth noting that BTM PV simulation capabilities are also valuable for policy, investment, and system design decision-making considerations. Figure 7 shows resulting distributions of simulated parameter values for azimuth angle, tilt angle, and ac-dc ratio for different system size classes in Connecticut. Table 3 lists the total number of systems simulated and selected across the entire New England region for the development of the finalized town-level profiles.

table 3. The simulated BTM PV systems by size class.

Capacity (kW)	Total Simulations	Total Selected
<100	26,676	26,676
≥100 ≤500	8,208	554
>500 ≤2,500	6,156	126
Total	**41,040**	**27,356**

Benchmarking Simulations to Measured BTM PV Production

Given the lack of both detailed BTM PV design and a complete set of measured BTM PV performance data, ISO New England uses an approach analogous to one used in Germany, for which a subset of information concerning installations and available production data from a sample of representative BTM PV sites are the basis of accurate estimates of BTM PV production. In ISO New England's case, the two primary inputs to the estimation are BTM PV installation data and vendor-provided BTM PV performance data, both at the town/municipal level of granularity. The performance data at any given time step represent relative BTM PV performance on a per-unit nameplate capacity basis, as illustrated in the heat maps in Figure 8. For example, the colors depicted within each map represent the BTM PV power output located in each town as a relative share of its total nameplate capacity. For both the simulated and measured BTM PV production data, there are different numbers of individual installations in each town with differing design characteristics that are the basis of the per-unit profile. Towns without any PV systems are colored gray or black in the measured BTM PV data maps. With the data structured in this manner, the user can simply scale up a town-level profile by the amount of capacity installed in the town to yield the estimated BTM PV output in that town and aggregate the resulting town-level estimates to yield a total estimated BTM PV power profile for a larger region of interest.

The primary rationale behind this approach is that once BTM PV fleets mature and reach the tens or hundreds of thousands of individual installations covering a large area, a statistical approach to modeling their aggregate performance characteristics is possible. In a sense, this type of approach is a form of data compression, with the goal being the use of a limited amount of information to infer the total BTM PV performance and, in turn, eliminate the need for any information that is redundant. This kind of approach will work as long as the most salient features that govern the BTM PV fleet's performance are established and a sufficient amount of data is available to represent these features. For the BTM PV simulations, the most salient factors are the localized weather conditions represented by the 4-km NSRDB weather data and the assumed system design characteristics of the BTM PV fleet used as inputs to the simulations. Note that this approach may not be well suited for all applications. For example, in applications limited to localized, distribution-level analysis and forecasting, the BTM PV fleet is much smaller, which may warrant greater consideration of, for example, more detailed PV system design characteristics.

This statistical approach is similar to top-down load forecasting methods typically used by ISOs/RTOs. For these

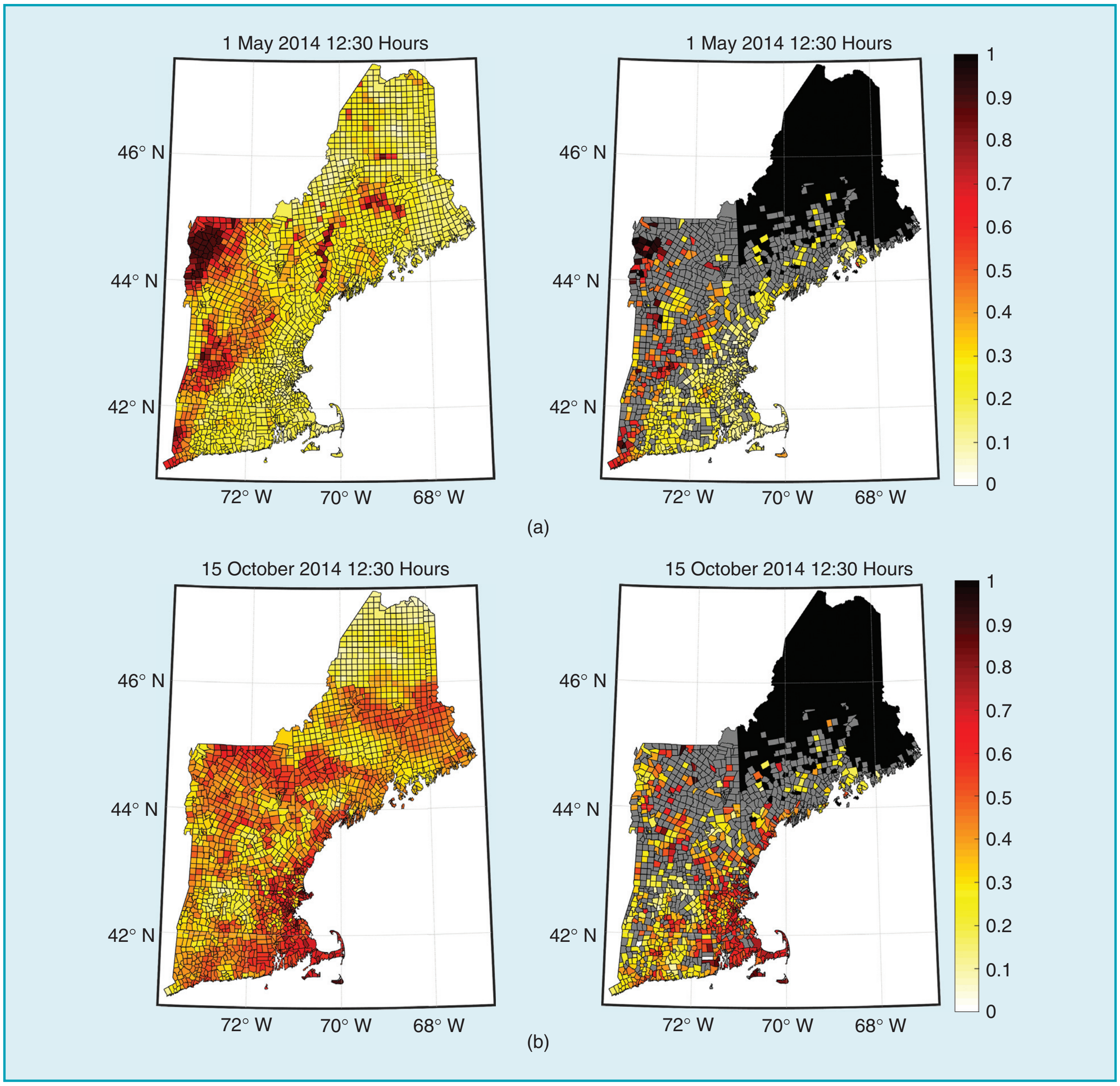

figure 8. (Left) simulation and (right) measured BTM PV performance on (a) 1 May 2014 and (b) 15 October 2014. (Graphic courtesy of ISO New England; Quantitative Business Analytics, Inc. is the data source for measured data.)

methods, rather than forecasting demand for each light fixture or appliance at the house or business level, a diversity of coincident, individual electricity end uses are aggregated into electricity consumption for an area, and statistical models trained on historical data leverage the implicit load patterns. In turn, these models can then accurately predict load based solely on calendar and weather inputs.

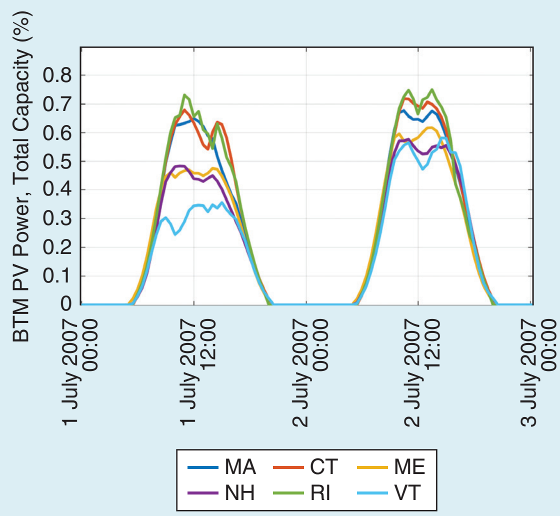

figure 9. The average state-level BTM PV power profiles, 1–2 July 2007.

High-Quality, High-Fidelity Results

Figure 8 illustrates a comparison of the town-level simulation results (left) and the town-level performance of more than 2,000 actual PV installations (right) on (a) 1 May 2014 at 12:30 Eastern Prevailing Time (EPT) and (b) 15 October 2014 at 12:30 EPT. Both data sets represent the relative amount of solar production at the municipal level, as described previously. Despite the two sets being independently sourced, there is broad agreement between the simulated versus measured PV data. The high degree of spatial color coherence in the representation of both sets demonstrates the value of using the town-level, per-unit production data in the estimation process, as the heat maps clearly depict their ability to encapsulate the effects of local weather (mainly cloud cover) on the BTM PV fleet. For example, on (a) the 1 May plots, both maps illustrate low PV output due to mostly cloudy conditions throughout much of the southeastern portion of the region and higher PV output in northwestern parts of the region due to less cloud cover. Similarly, (b) the 15 October plots illustrate mostly sunny conditions in the southeast, with the effects of greater cloud cover reflected in various pockets of lower PV output in the remainder of the region.

table 4. The average annual capacity factors for simulated BTM PV in New England states. (Note: the NSRDB data for Maine is missing for 2010.)

Year	Massachusetts	Connecticut	Maine	New Hampshire	Rhode Island	Vermont
1998	15.5%	15.6%	14.3%	14.8%	15.8%	14.6%
1999	16.5%	16.4%	15.3%	15.7%	16.6%	15.7%
2000	15.8%	16.0%	14.6%	14.9%	16.2%	14.6%
2001	16.6%	16.5%	15.5%	15.8%	16.9%	15.5%
2002	16.1%	16.1%	15.0%	15.3%	16.4%	15.1%
2003	14.9%	14.7%	14.4%	14.5%	15.1%	14.6%
2004	15.2%	15.2%	14.7%	14.7%	15.4%	14.6%
2005	15.3%	15.4%	14.2%	14.6%	15.7%	14.8%
2006	15.1%	15.2%	13.7%	14.2%	15.2%	13.7%
2007	15.6%	15.6%	14.5%	14.8%	15.9%	14.7%
2008	15.0%	15.2%	13.7%	14.2%	15.3%	14.0%
2009	14.8%	14.7%	13.8%	14.2%	15.0%	14.1%
2010	15.4%	15.4%	—	14.4%	15.7%	13.9%
2011	15.2%	15.2%	14.1%	14.4%	15.6%	14.3%
2012	15.8%	15.9%	14.8%	15.1%	16.1%	15.2%
2013	15.8%	15.8%	14.4%	14.9%	16.0%	14.5%
2014	15.4%	15.5%	14.3%	14.6%	15.8%	14.4%
Average	**15.5%**	**15.5%**	**14.5%**	**14.8%**	**15.8%**	**14.6%**

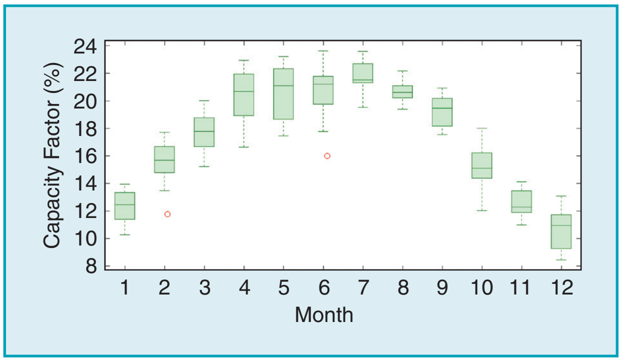

figure 10. The monthly average capacity factors in Massachusetts, 1998–2014. The upper and lower ends of each box represent the first and third quartiles, respectively. The band inside the box is the median value, and the ends of the lines extending above and below the boxes ("whiskers") reflect the degree of variability outside of the upper and lower quartiles. The red circles reflect statistical outliers.

Figure 9 illustrates the simulated per-unit BTM PV profile for the six New England states on 1–2 July 2007. These profiles reflect an arithmetic mean of all town profiles within each state and show the overall effect of weather in each state on the simulated BTM PV. Based on the same set of results, it is also possible to develop aggregate BTM PV profiles that reflect a specified town-level geographical distribution of installed BTM PVs by simply weighting the town profiles by each town's installed capacity.

Table 4 summarizes the average annual capacity factors for the six states over the entire 17-year NSRDB-based simulation period, and Figure 10 is a box plot showing the variability of BTM PV's monthly capacity factors in Massachusetts over the entire period. Since New England typically experiences snow events during winter and the effects of snow cover were not explicitly modeled in the simulations, the annual performance metrics are somewhat optimistic for the winter months (and annually as a result). However, the winter results are representative of the solar resource availability in winter, absent snow.

Overall, the broad agreement between the granular, town-level simulation results and measured data demonstrates that the BTM PV fleet simulation yielded reasonable, realistic results and highlights that NREL's new NSRDB data set is not only spatially and temporally comprehensive but also of a very high quality. Given that the simulation period is 17 years, a relatively robust understanding of the patterns of solar resource availability is now possible in the area covered by NSRDB. Perhaps, more importantly, coupling the results with long-term forecasts of BTM PV and historical loads enables a thorough investigation of net load patterns as BTM PV penetrations increase. Using this information, greater clarity is now possible concerning the timing and magnitude of mitigation measures that may be required to integrate increasingly larger BTM PV penetrations in the future. Lastly, the NSRDB now makes it possible for long-term load forecasters to factor BTM PVs into their forecasts, where and when necessary.

Concluding Remarks

More than 70 years ago, researchers recognized temperature as a driving factor of electricity demand. Since then, temperature data has been widely used in load forecasting models. Due to the increasing need of accurate energy forecasts, more weather data are being adopted by the power industry. In this article we presented two utility applications that rely on weather data. We used some conventional and easily accessible weather data to offer a novel and probabilistic view of SAIDI, which helps reveal the utility's reliability trend. In the BTM PV simulation study, we used some recently developed comprehensive weather data to tackle an emerging problem in renewable integration. Researchers and practitioners have created a wealth of knowledge in the science of meteorology and atmosphere science in general. We hope that this article will inspire you to try out more weather data and expand the footprints of meteorology in the energy analytics arena.

For Further Reading

IEEE Guide for Electric Power Distribution Reliability Indices, IEEE Standard 1366-2012 (revision of IEEE Standard 1366-2003), May 2012.

T. Hong, J. Wilson, and J. Xie, "Long term probabilistic load forecasting and normalization with hourly information," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 456–462, Jan. 2014.

T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecasting*, vol. 32, no. 3, pp. 914–938, July–Sept. 2016.

National solar radiation database. (2017). National Renewable Energy Laboratory. [Online]. Available: https://nsrdb.nrel.gov/current-version#psm

National Renewable Energy Laboratory. System advisor model. [Online]. Available: https://sam.nrel.gov/

E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, "Regional PV power prediction for improved grid integration," *Progress Photovoltaics Res. Applicat.*, vol. 19, no. 7, pp. 757–771, Nov. 2011.

Biographies

Jonathan Black is with ISO New England, Holyoke, Massachusetts.

Alex Hofmann is with the American Public Power Authority, Arlington, Virginia.

Tao Hong is with the University of North Carolina at Charlotte.

Joseph Roberts is with ISO New England, Holyoke, Massachusetts.

Pu Wang is with the SAS Institute Inc., Cary, North Carolina.