Solar Energy 157 (2017) 542–551

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Blending distributed photovoltaic and demand load forecasts

Sue Ellen Haupt[a,*,1], Susan Dettling[a,1], John K. Williams[a,b,1], Julia Pearson[a,1], Tara Jensen[a,1], Thomas Brummet[a,1], Branko Kosovic[a,1], Gerry Wiener[a,1], Tyler McCandless[a,c,1], Crystal Burghardt[a,1]

[a] *National Center for Atmospheric Research, Boulder, CO 80303, United States*
[b] *The Weather Company, IBM Business, Andover, MA, United States*[2]
[c] *Ascend Analytics, Boulder, CO, United States*[2]

ARTICLE INFO

Keywords:
Distributed generation
Grid integration
Load forecasting
Photovoltaic power
Solar energy

ABSTRACT

Utilities consider having accurate electric load forecasts to be critical to their day-to-day operations. But the growing penetration of distributed photovoltaic (DPV) solar power production "behind the meter" makes it more difficult to predict load because it is hard to distinguish between increased solar power generation and decreased power consumption at a site with installed DPV. This paper describes the development of a "top down" solar power forecasting system and how it can be integrated with a load forecasting system that incorporates weather information. Both systems depend on accurate weather forecasts, information regarding the utility variables, and daily and seasonal factors. The load forecasting system provides day-ahead forecasts having roughly 2% error, on average, and the solar power forecast is within 5% error. In general, we find that if the load forecasting system incorporates the same weather and solar predictors as the solar power forecasting system, it implicitly accounts for DPV generation during periods of stationary solar power deployment, and a separate DPV forecast may not be necessary. However, for higher penetrations of solar power during times of rapid deployment of additional solar capacity, it may become important to explicitly incorporate the solar power forecasts to avoid degradation of the load forecasts. This approach could allow utilities and independent system operators to better deal with rapidly increasing penetration of DPV generation.

1. Introduction

Forecasting the electric load is of critical importance to utilities and Independent System Operators (ISOs) for balancing the electric grid (Feinberg and Genethlion, 2005; Hong, 2014). This balancing occurs on several time scales: (1) short term (about one hour to a week) for day-to-day operations, (2) medium range (about one week to a year) for maintenance and longer planning, and (3) longer term (for planning beyond a year). Electric load depends on weather, customers' daily and weekly usage patterns, and the effects of varying weather conditions on these patterns. Some of these usage patterns are commercial or industrial, while others are driven by residential use. The balance between these factors varies greatly with the climate and land usage of the region for which the prediction is being made. Predicting electrical load typically includes examining historical load data along with information about past, current, and predicted future weather conditions, including temperature, solar insolation, humidity, precipitation, and wind speed. These quantities impact buildings' stored heat and their residents' usage patterns, particularly those that affect heating, cooling, and lighting. Accurate load predictions allow utilities to operate more economically and efficiently by avoiding spot market power purchases or unnecessary use of expensive spinning reserves.

There are well known methods to predict load. One is to search for similar days in the past (analogs) and summarize the corresponding observed loads to use as a prediction. Others use statistical learning methods, including regression models, time series methods (Almeshaiei and Soltan, 2011), artificial neural networks (Park et al., 1991; Lee et al., 1992), support vector machines (Hong 2009), and advanced approaches that incorporate multiple artificial intelligence methods (Wang et al., 2012). All of these methods require high quality historical data on usage patterns and the corresponding meteorological data.

The proliferation of "behind the meter" distributed photovoltaic (DPV) solar power production is complicating the prediction of electric load. Because DPV is seldom metered, the production of solar power at

* Corresponding author at: National Center for Atmospheric Research, 3450 Mitchell Lane, Boulder, CO 80303, United States.
E-mail address: haupt@ucar.edu (S.E. Haupt).
[1] This work was performed at the National Center for Atmospheric Research, Boulder, CO, United States.
[2] Present address.

http://dx.doi.org/10.1016/j.solener.2017.08.049
Received 30 December 2016; Received in revised form 14 August 2017; Accepted 16 August 2017

a residence or commercial building appears to the utility as a decrease in its load. It is thus difficult to distinguish between decreased electric usage versus increased solar power production. As the installed capacity of DPV increases, utilities worry about steep decreases in apparent load in the morning when the DPV begins to produce more power, and steep down ramps in the evening as the sun sets, causing solar power production to decrease just as evening residential usage rises. California ISO (CAISO) has widely publicized this problem as a "duck curve" of net load representing underlying conditions that will become more of a challenge as DPV increases (Letendre et al., 2014). Hawaiian Electric Company (HECO) has already reached this challenging level of production and are "underwater" at certain times of the day, with negative net demand, which they refer to as the "Nessie curve." (Nakafugi, 2016). During times in which solar power may account for a large fraction of supply, unanticipated changes in DPV due to changing weather, such as frontal passages or thunderstorms, can have significant impact on apparent load.

One way to combat this problem is to provide better situational awareness in the form of forecasts of DPV and its impact on apparent load. Recent studies have provided evidence of real savings to rate payers due to improved solar power forecasting (Brancucci Martinez-Anido 2016; Haupt et al., 2016a, 2017). There are several ways that this can be accomplished, which are reviewed by Tuohy et al. (2015) and Orwig et al. (2014). The first is the "bottom up" approach, which uses details of the individual installations, including their location, capacity, tilt, azimuth, etc., to directly predict the production at each DPV installation and sum those over the region of interest. This methodology has been used to integrate DPV with load forecasts in California (Hoff, 2016). The second approach is "top down," in which such detailed installation information may not be available, but a forecast of irradiance is provided for a particular region, then transformed to DPV production based on known installed capacity (Lorenz et al., 2014). This current study takes the latter approach due to the lack of detailed information regarding characteristics of individual sites. Thus, we leverage historical, current, and forecasted weather information; prior and current solar energy production; and electrical load to provide forecasts of solar power and determine how it modulates the load forecast. Kaur et al., (2016a) discuss the benefits of providing solar power forecasts, allowing a substantially lower requirement for flexibility reserves than with baseline forecasts of either persistence or smart persistence (allowing changes in solar angle, but persistence in sky condition).

This research is performed in collaboration with Xcel Energy's Public Service Company of Colorado (PSCo), which has a rapidly growing DPV capacity, about 196 MW at the beginning of the study (toward the end of 2013) and rapidly growing with 244 MW installed by year-end 2014. With a peak load of 6401 MW in 2014, this capacity represents only 3.8% of the average load over the service area in 2014. This is in addition to the 84 MW of utility-scale solar installed at that time (also rapidly growing), which is forecast explicitly. The challenge to DPV prediction is the lack of detailed information about each DPV installation or any detailed data on its historical solar power production. Thus, we are constrained to use the "top down" approach to solar power forecasting. We also supply an electric load forecast and assess the impact of solar power production on the apparent load. To accomplish this, we address a series of questions:

1. Is the DPV discernible in the Colorado load under current installed capacity?
2. Can we build an accurate forecast for DPV power generation without access to installation metadata or historical production information?
3. How do we assess accuracy of our DPV forecast?
4. If we include the DPV forecast as a component, does it substantially improve accuracy of the net load forecast?
5. During times of rapid growth in DPV capacity, is it necessary to include the DPV forecast in the net load forecast to improve its accuracy?

To address these issues, the National Center for Atmospheric Research (NCAR) developed and tested both load and DPV forecasting systems. These systems were designed so that they can be coupled to produce a net load (demand load minus DPV) forecast well suited to adapt as the penetration of DPV in Colorado grows (Haupt et al., 2016b; Williams et al., 2014b).

The remainder of the paper describes our data, methodology, construction of the forecast systems, and analysis of results. Section 2 provides a system overview and examines whether Colorado has reached a capacity of DPV that merits its explicit inclusion in the load forecast. Multiple types of data are input to the load and DPV forecasting systems as detailed in Section 2. Section 3 describes the model to predict DPV and section 4 describes the load forecasting model. We examine the impact of the DPV forecast on the net load forecast as the capacity of DPV grows over time in Section 5. Section 6 summarizes the work and provides conclusions and recommendations for further research and applications.

2. Load analysis and system overview

As the capacity of DPV grows, it increasingly impacts net load at all timescales. To address our first question for Xcel Energy's Colorado service area, we wish to evaluate whether the current level of DPV generation is evident as a load cutout. Then, in Section 2.2, we review the overall architecture of the combined load/DPV prediction system before describing it in more detail in subsequent sections.

2.1. Load cutout statistical analysis

In order to identify and define the load cutout for PSCo, a statistical analysis was performed to compare load data between clear and cloudy days during different seasons. We were looking to see whether the load is discernably reduced during the sunny periods. Note, however, that when sunny periods coincide with hot days, the analysis will be confounded given the impact of cooling on actual demand.

Using Boulder Airport climate data available from the National Weather Service (NWS) website, days during 2013 were classified as either clear or cloudy. The NWS dataset records the daily average sky cover between sunrise and sunset on a 0–10 scale, where "0" indicates that no clouds were observed and "10″ means clouds covered the entire sky for that day. Days in 2013 were selected for which the sky cover was either "0" (clear) or "5–10" (cloudy); days in the 1–4 range (partly cloudy) were discarded and not analyzed further for this purpose. Any day falling in December, January, or February was labeled as "winter"; in March, April, or May as "spring"; in June, July, or August as "summer"; and in September, October, or November as "fall". For each of these days, the average load during daylight hours was computed. Each day was also classified as a weekday, a weekend day, or a holiday.

For each season, the distributions of average daily load on clear and cloudy days were compared. The null hypothesis was that there is no difference between the distributions of load on clear vs. cloudy days. This hypothesis was tested using the Mann-Whitney test, a nonparametric test that allows two sets of data to be compared without assuming that their values are normally distributed. The test is performed on ranked data; thus, the average loads in each set were ranked. Specifically, the smallest value was assigned a rank of 1, the next smallest a rank of 2, and so on. The Mann-Whitney test compares the magnitude of the elements in the two sets; if most of the Y's are greater than most of the X's, or vice versa, that would provide evidence against random mixing, discrediting the null hypothesis. The lower the p-value, the more statistically significant are the results.

Table 1 summarizes the results of this analysis. The first column provides the seasons into which the data were stratified; the second lists whether clear or cloudy days had higher loads; the third provides the Mann-Whitney test p-values for this ordering of clear and cloudy, where p-values indicating statistical significance at 5% (two-tailed test) are