Solar Energy 171 (2018) 397–413

Contents lists available at ScienceDirect

Solar Energy

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

Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals

D.W. van der Meer*, J. Munkhammar, J. Widén

Built Environment Energy Systems Group (BEESG), Division of Solid State Physics, Department of Engineering Sciences, Uppsala University, P.O. Box 534, SE-751 21 Uppsala, Sweden

ARTICLE INFO

Keywords:
Probabilistic forecasting
Quantile regression
Gaussian process
Solar power
Electric load
Net load

ABSTRACT

This paper presents a study into the effect of aggregation of customers and an increasing share of photovoltaic (PV) power in the net load on prediction intervals (PIs) of probabilistic forecasting methods applied to distribution grid customers during winter and spring. These seasons are shown to represent challenging cases due to the increased variability of electricity consumption during winter and the increased variability in PV power production during spring. We employ a dynamic Gaussian process (GP) and quantile regression (QR) to produce probabilistic forecasts on data from 300 de-identified customers in the metropolitan area of Sydney, Australia. In case of the dynamic GP, we also optimize the training window width and show that it produces sharp and reliable PIs with a training set of up to 3 weeks. In case of aggregation, the results indicate that the aggregation of a modest number of PV systems improves both the sharpness and the reliability of PIs due to the smoothing effect, and that this positive effect propagates into the net load forecasts, especially for low levels of aggregation. Finally, we show that increasing the share of PV power in the net load actually increases the sharpness and reliability of PIs for aggregations of 30 and 210 customers, most likely due to the added benefit of the smoothing effect.

1. Introduction

The increasing penetration of photovoltaic (PV) power into the electricity generation mix has at least two consequences that affect one another in an interesting way. Firstly, an increase in the number of installed PV systems in a wider area is accompanied by the smoothing effect, as described by Perez and Hoff (2013). If the data are available, this could allow for more accurate city scale or regional PV power forecasts. In case of insufficient data, one could estimate the output of these "invisible" sites using a representative set of sites, see, e.g., Shaker et al. (2015) for a thorough study. Secondly, however, the increase in the number of installed PV systems or their sizes simultaneously reduces the accuracy of net load forecasts at the distribution feeder or even finer spatial resolution because of the increased variability and ramps, see e.g., Denholm and Margolis (2007), Nguyen et al. (2016). It should be noted that Kaur et al. (2013) showed that penetration did not affect the net load forecast error to the same extent as variability did. However, the reason for their conclusion can mainly be ascribed to the fact that they considered a single PV power farm and looked at the influence of the cloud conditions and the subsequent penetration levels due to these conditions. A logical conclusion would therefore be to ascribe the forecast error to variability, since very little smoothing occurs and since in this case, the penetration is a direct function of variability.

The installed capacity of PV power is still increasing, around 75 GW of capacity was installed during 2016 (IEA, 2017). In Germany, PV power covered 7.4% of electricity consumption during 2016, which increased to 35% on sunny weekdays (Wirth, 2017). At such levels, challenges such as feeder loading, grid losses and voltage fluctuations occur in the distribution grid (Aguero and Steffel, 2011; Mohammadi and Mehraeen, 2017; Walling et al., 2008). Accurate forecasts are generally seen as a cost-efficient solution because they allow for e.g., unit commitment, curtailment and demand response, which in turn can be used to reduce operational risks (Kaur et al., 2016b). However, deterministic, or point forecasts, do not express uncertainty, which is important for stakeholders to make informed decisions. Probabilistic forecasting has gained more attention recently, as can be inferred from recent review studies (Hong and Fan, 2016; van der Meer et al., 2018b). The advantage of probabilistic forecasting is that the level of uncertainty can be expressed that accompanies the prediction by means of

* Corresponding author.
E-mail address: dennis.vandermeer@angstrom.uu.se (D.W. van der Meer).

https://doi.org/10.1016/j.solener.2018.06.103
Received 28 March 2018; Received in revised form 19 June 2018; Accepted 25 June 2018

a probability density function (PDF) or prediction interval (PI), which enables stakeholders to make better informed decisions. The smoothing effect has been shown to improve the accuracy of deterministic forecasts, see e.g., Perez and Hoff (2013). In this paper, we investigate whether this holds in case of probabilistic forecasting.

1.1. Previous work

1.1.1. The effect of aggregation

Variability in solar irradiance, and PV power production by extension, and electricity consumption can be reduced by the well-known smoothing effect. An extensive review by Widén et al. (2015) revealed a series of notable studies into the variability of irradiance and the smoothing effect. For instance, Lave et al. (2012) studied aggregated irradiance measurements of six sites within 3 km of each other on timescales varying from 1 s to 1 h. They found that on that particular geographical scale and at a time span below 5 min, the ramp rates became uncorrelated and the smoothing effect was substantial. Furthermore, the authors noted that for timescales shorter than 256s, the variance was reduced by a factor of 6 when compared to a single system. On a larger geographical scale where 7 PV plants were separated between 6 and 360 km, Marcos et al. (2012) found that, at a temporal resolution of 1 s and distance of 6 km, the power output was already uncorrelated, thus substantially increasing the smoothing effect. However, at lower temporal resolutions, e.g., 30 min as used in this work, the smoothing effect is less pronounced when the same plants are considered, because such a resolution allows similar weather patterns to occur over a wider area (Widén et al., 2015).

In case of electricity consumption forecasting, Hayes et al. (2015) studied what effect aggregation has on the mean absolute percentage error (MAPE) and examined the correlations between the dependent and independent variables. The authors noted that at the low voltage (LV) feeder, i.e., a low level of aggregation, these correlations were significantly lower than at the primary or secondary substations. Consequently, the MAPE decreased from 20%-30% at the end user level to 5% at the primary substation. Hayes et al. did not continue to investigate the probabilistic aspect of aggregation. Taieb et al. (2016) used quantile regression (QR) in combination with gradient boosting (GB) to forecast individual and aggregated electricity consumption, although forecasts at intermediate aggregation levels were not performed. The results showed that the prediction intervals (PIs) were substantially wide in case of end user electricity consumption forecasting, due to the high variability of this time series. In contrast, PIs for the aggregated demand of 3639 customers were highly informative and, moreover, accurate. The authors of both aforementioned studies noted that theirs were the first research papers that attempted to quantify the effect of aggregation on forecast accuracy (Hayes et al., 2015) and the second paper to utilize data from individual smart meters to produce probabilistic forecasts (Taieb et al., 2016).

In van der Meer et al. (2017) we showed that aggregation can drastically improve the quality of PIs with aggregations of modest numbers of customers. In this study, we continue that research by, among other things, using more subsets of the data in order to filter away noise and improve accuracy.

1.1.2. The seasonal effect

Few studies have investigated the effect of seasons on forecast accuracy, and fewer on probabilistic forecast accuracy. In general, such studies have been performed as a side step when, e.g., proposing a new model. For example, Vaz et al. (2016) proposed an artificial neural network (ANN) that utilized spatial and temporal information to produce deterministic forecasts of the PV power output in Utrecht, the Netherlands. The results showed that the forecasts during winter were more accurate in terms of the normalized root mean squared error (nRMSE) than those in summer, and the authors ascribed this to the fact that overcast days are more straightforward to predict. An example of a paper dedicated to study the impact of seasonal variations on deterministic load forecasting was performed by Lusis et al. (2017) in New South Wales, Australia. Here, the authors looked at a wide variety of scenarios where some included dummy variables to indicate the season, and other scenarios included subsets of the data that included only the season of interest. The results showed that, for a single residential load pattern, the scenarios including the dummy variables yielded marginally improved results, whereas the scenarios that included dedicated subsets of the seasons yielded worse forecast accuracy. The authors ascribed the latter result to a lack of robustness to outliers because of a reduced amount of subsets. However, in case of aggregated residential load patterns these negative effects were less pronounced.

As for probabilistic forecasting, there are only a few publications that go into detail as regards the effect of seasons on PIs. Firstly, Scolari et al. (2016) proposed a novel method to predict solar irradiance for very short time horizons, in which the PIs were produced based on quantiles obtained from *k*-means clusters. The location for their study was Lausanne, Switzerland. They assessed the performance for three seasons, i.e., summer, fall and winter, and noted that the season with highest variability was summer, followed by winter and fall. Consequently, the results showed that their method produced the least sharp PIs during summer, followed by winter and fall, as could be expected by the variability mentioned earlier. In addition, the authors showed that the coverage probability of the PIs remained above the selected nominal coverage levels (85%, 95% and 99%), although it exceeded the nominal coverage level by a modest margin during winter. Secondly, Davò et al. (2016) used an analog ensemble (AnEn) to predict regional wind power and solar irradiance in Sicily, Italy and Oklahoma, USA, respectively. The AnEn, originally proposed in Delle Monache et al. (2013), combines multiple deterministic forecasts from the past conditional on observations of the current state. The authors showed that the continuous ranked probability score (CRPS, negatively oriented score and formally introduced in Section 2.4) fluctuated substantially as a function of season, with the highest score achieved during summer and the lowest score during fall. Finally, Wang et al. (2017a) utilized a deep convolutional neural network (DCNN) in combination with QR to produce probabilistic forecasts of PV power production and reported that CRPS was lowest during winter, and generally highest in summer, where the latter was attributed to the local volatile weather conditions. The location in the study was Flanders, Belgium.

From this literature review, we can conclude that it is reasonable to expect deteriorating performance of the PIs when the variability of the data set increases. Although the papers mentioned here concerned solar power or solar irradiance, it is reasonable to expect similar findings in load forecasting due to the dependence of electricity demand on weather variables, see e.g., Drezga and Rahman (1998), Raza and Khosravi (2015). Moreover, due to climatological differences, it is difficult to predict what season will be least straightforward to predict. Studies into the variability of the data set per season are therefore required, which will be done in Section 2.1.

1.1.3. The effect of penetration

The effect of large scale integration of renewable energy sources (RESs) such as PV power on the power system is an important topic of research because it directly affects the net load, which is the amount of power the utility has to provide or absorb. Several studies exist that study this effect, see e.g., Shaker et al. (2016) for a study into the impact of large scale wind and solar power penetration, who noted that the volatility of the net load could increase up to eight times, making accurate forecasts ever more challenging. Another notable example is the study by Olauson et al. (2016), who investigated the net load variability in Scandinavia with high or full integration of RESs in the power system, and the optimal mix of RESs. They concluded that the optimal mix depends on the time span under consideration, e.g., diurnal or seasonal, and, more importantly, that a fully renewable power system is possible with diverse RESs backed up by hydropower.