Factoring Behind-the-Meter Solar into Load Forecasting: Case Studies under Extreme Weather

Mucun Sun, Cong Feng, Jie Zhang
The University of Texas at Dallas
Richardson, TX 75080, USA
Email: jiezhang@utdallas.edu

***Abstract*—Distributed energy resources (DERs), especially distributed photovoltaics (PV), have been rising dramatically over the past years. However, behind-the-meter (BTM) PV devices are not monitored, and thus are invisible to utilities and system operators. In addition, electricity demand is likely to increase as a result of extreme hot/cold weather conditions, which stretches the grid to its limits, and thus triggers high electricity price. High electricity prices and extreme temperatures also stimulate the adoption of solar panels, which in turn add difficulties to load forecasting. This paper proposes a data-driven feeder-level load forecasting method by taking account of BTM PV under extreme weather conditions. The BTM PV penetration is first estimated, and in this study the PV penetration is defined as the ratio of total BTM PV capacity to peak load of the feeder. A machine learning model is adopted to quantify the relationship between measured PV power generation and corresponding solar irradiance. The BTM PV generation within the entire feeder can be estimated through the PV penetration and forecasted PV irradiance, which is then integrated in load forecasting. Numerical results of case studies at three distribution feeders show that the performance of load forecasting under extreme weather conditions is significantly enhanced by considering the contribution of BTM PV.**

***Index Terms*—load forecasting, behind-the-meter solar forecasting, extreme weather.**

I. Introduction

Behind-the-meter (BTM) PV mainly refers to small scale roof-top solar for a single building or facility, which is generally invisible to utilities and system operators [1]. In recent years, solar energy penetration has increased dramatically in power systems. According to solar market insight report provided by the Solar Energy Industries Association (SEIA), over 600 MWdc distributed PV was installed in 2019 Q1 [2]. Most unmonitored and invisible PVs are from residential buildings due to privacy concerns or expensive measurement and communication infrastructure. With the ever-increasing distributed PV adoption, it is more challenging to accurately forecast the electric load at the distribution level, e.g., substation, feeder, or transformer. In addition, electricity demand is likely to increase as a result of extreme hot/cold weather conditions, which makes load even more unpredictable. Underestimating the load demand may cause a utility to deliver poor quality of services (even blackouts), and compel the company to buy energy from external sources. Thus, it is interesting to explore and study the impacts of BTM solar on load forecasting, especially under extreme weather conditions.

A number of load forecasting technologies have been developed in the literature, with applications to a variety of power system operation and planning problems. For example, load forecasting has been used for unit commitment and economic dispatch [3], [4], [5], distribution system reconfiguration [6], [7], demand response management [8], [9], and operating reserves [10], [11]. Methods of load forecasting can be classified into time-series models, regression models, and artificial intelligence models [12]. Time-series models rely on explanatory variables but explore the inherent correlation to past events, which have lower data requirements and less computation cost. For example, a time-series load forecasting model with different cycles of seasonality was proposed in [13]. This network separately models both non-seasonal and seasonal cycles of the load data using auto-regressive (AR) and moving-average (MA) components. Different from time-series models, a regression model quantifies the relationship between load profile and other explanatory variables such as meteorological information, day type, and customer activities. For example, Wang *et al.* [14] used a large number of lagged temperature and moving average temperature variables to build multiple linear regression models for load forecasting. Artificial intelligence models have been shown to present better accuracy and robustness than both regression models and time-series models in load forecasting [15]. For example, Jiang *et al.* [6] used support vector regression (SVR) to forecast the aggregated loads of a small section of a distribution system. The California Independent System Operator (CAISO) has adopted a "bottom-up" method to generate load forecasting with BTM solar [16]. The PV information in the state of California is recorded and then combined with high-resolution irradiance values and weather predictions to forecast the BMT PV output for the entire state. While utility-scale solar forecasting has been significantly improved in recent years, BTM solar forecasting is still lagging behind. In addition, the impact of BTM solar on load forecasting accuracy has also not been well studied in the literature.

The most intuitive way of factoring the BTM PV into a load forecasting is to model the distributed PV generation and integrate it into the load forecasting model. To accurately model BTM PV generation, we can utilize: (i) irradiance-to-power models based on measured solar irradiance and PV panel geographical information, or (ii) direct measurement of PV generation. Though all BTM PV generation can be directly measured, one of the chief impediments to the model and control of BTM resources is the lack of communications with these distributed systems. For the irradiance-to-power modeling, accurate locations of PV system and detailed PV panel information may not always be available.

To study the impacts of distributed PV on load forecasting, in this paper, we seek to develop a feeder-lever load forecasting methodology considering the BTM PV generation under different weather conditions. First, the BTM PV penetration is estimated, and in this study the PV penetration is defined as the ratio of the total BTM PV capacity to the feeder peak load. Then, a machine learning model is adopted to quantify the relationship between the measured PV generation and corresponding PV irradiance. The BTM PV generation within the whole feeder can be estimated through the PV penetration and forecasted solar irradiance. The estimated BTM PV generation is then used in conjunction with weather forecasts and feeder load to train the load forecasting model, and specially extreme hot/cold weather is explored in this paper. Finally, the trained model is used to generate load forecasts under extreme weather conditions.

The rest of the paper is organized as follows. Section II describes the proposed load forecasting method, which consists of feeder-level BTM PV estimation, extreme weather extraction, and load modeling with BTM solar. Section III applies the developed load forecasting method to three distribution feeders with high PV penetration under extreme weather conditions. Concluding remarks and future work are discussed in Section IV.

II. Methodology

The overall framework of the proposed load forecasting methodology by considering BTM PV is illustrated in Fig. 1, which consists of three major steps: feeder-level BTM PV estimation, load modeling, and load forecasting considering BTM PV under extreme weather conditions. The three major steps are briefly described as follows:

1) Step 1: The relationship between measured PV generation and corresponding solar irradiance is quantified through a support vector regression (SVR) model. With the trained SVR model, for a specific feeder, the BTM PV generation can be estimated based on the PV penetration and forecasted solar irradiance.
2) Step 2: The estimated BTM PV output of the distribution feeder is used in conjunction with meteorological variables (e.g., humidity, cloud coverage, heat index, temperature, and wind speed) to train load forecasting models under extreme hot/cold weather conditions.
3) Step 3: Forecasted weather is fed to the trained extreme hot/cold load forecasting models to generate load forecasts under extreme weather conditions.

A. Feeder-level Distributed PV Generation Modeling

It is generally challenging for a grid operator to collect all necessary factors to model each individual PV system and then aggregate them together. A reduced-form approach to model the feeder-level distributed PV generation requires no specific information of each distributed PV system, which can be formulated using historical partially measured PV power output, corresponding solar irradiance, and feeder PV penetration. Note here the solar irradiance consists of Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI), and Diffuse Horizontal Irradiance (DHI).

For a distribution feeder, the PV penetration in this paper is defined as the ratio of the total BTM PV capacity to the peak load of the feeder, which is expressed as:

$$\eta = \frac{\sum C_{i,j}}{y_j^{max}} \tag{1}$$

where η is the PV penetration, $C_{i,j}$ is the capacity of the ith distributed PV (belongs to the jth feeder), and y_j^{max} is the peak load of the jth feeder.

Let p_i denote the measured PV time-series power of the ith distributed PV, and $x_{irr,i}$ denotes the corresponding solar irradiance time-series, the relationship between p_i and $x_{irr,i}$ can be quantified through a regression model:

$$p_i = f(x_{irr,i}) \tag{2}$$

Several possible regression models can be used, such as SVR, artificial neural network (ANN), and random forest (RF). SVR is adopted in this paper due to its high accuracy in case studies. More details about the SVR model can be found in [17].

Given the relationship between solar irradiance and distributed PV power generation, the total BTM PV power output of the entire feeder can be estimated based on the PV penetration and forecasted irradiance.

B. Extreme Weather Extraction

This paper focuses on the load forecasting under extreme weather conditions, which is more critical for distribution system operators. Thus, in the training stage, the first step of extreme load forecasting is to cluster data that conforms the definition of extreme hot/cold weather. A day with surface temperature at any hour greater than a threshold (e.g., $95°F$) is defined as an extreme hot day, while a day with surface temperature at any hour lower than a threshold (e.g., $36°F$) is defined as an extreme cold day.

Algorithm 1 Gradient boosting machine (GBM)

1: $F_0(x) = argmin_\rho \sum_{i=1}^{N} \Psi(y_i, \rho)$
2: **for** $m = 1$ to M **do**
3: Compute the negative gradient of the loss function:
$\bar{y}_i = -\left[\frac{\partial \Psi(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}, i = 1, ..., N$
4: Fit a model to $\bar{y}$ by least-square to calculate a_m:
$a_m = argmin_{\alpha,\beta} \sum_{i=1}^{N} [\bar{y}_i - \beta h(x_i, a)]^2$
5: Calculate ρ_m by:
$\rho_m = argmin_\rho \sum_{i=1}^{N} \Psi(y_i, F_{m-1}(x_i) + \rho h(x_i, a_m))$
6: Update the model by:
$F_m(x) = F_{m-1}(x) + \rho_m h(x, a_m)$
7: **end for**

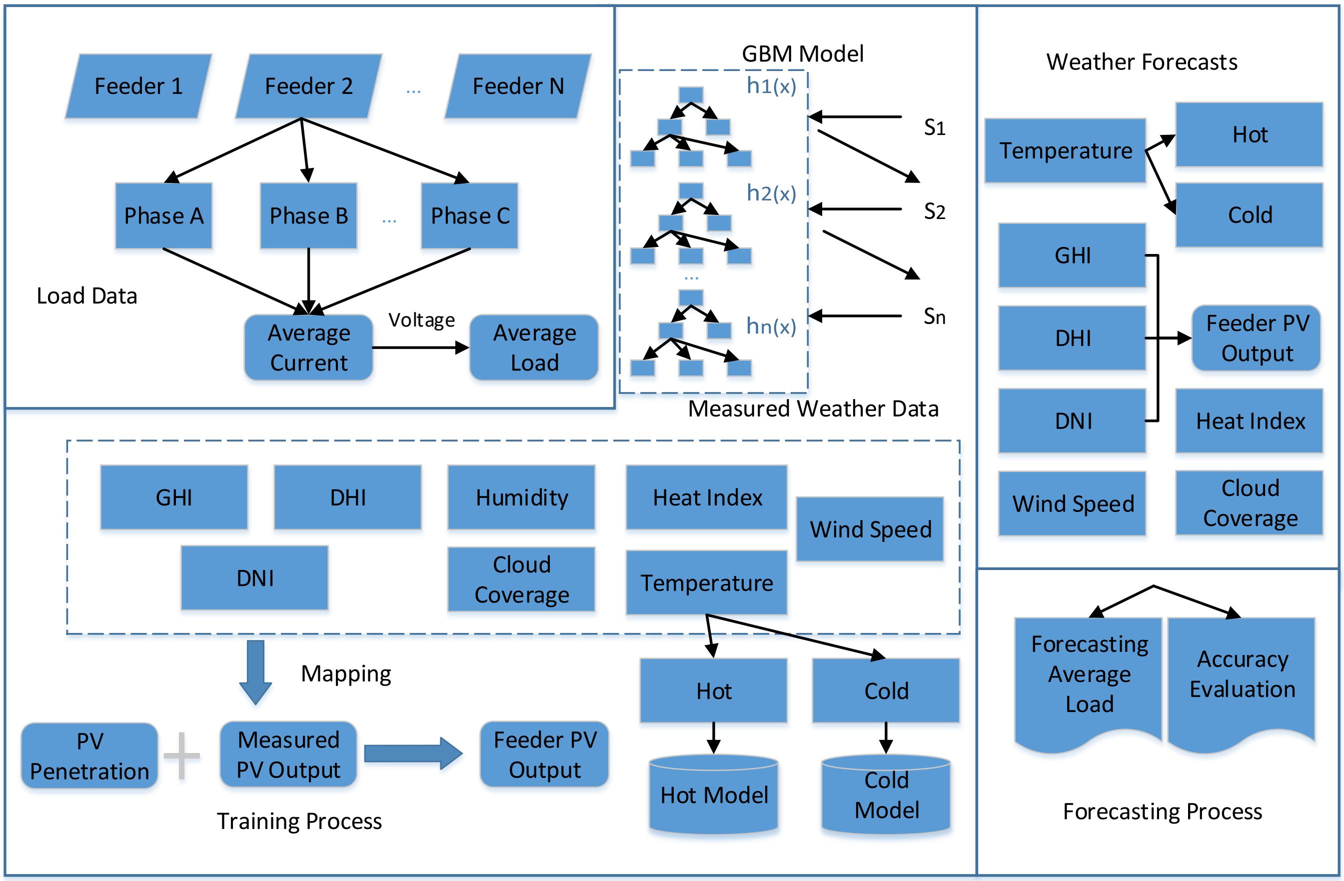

Fig. 1: The Overall framework of the developed load forecasting model considering BTM PV under extreme weather conditions

C. Load Forecasting Model Training

Gradient boosting machine (GBM) is one of the most widely used machine learning techniques for regression and classification problems, which is adopted in this paper for load forecasting modeling. GBM is a type of ensemble learning method that implements a sequential boosting algorithm. The main idea of GBM is to iteratively improve the imperfect model $F_m(x)$ at the mth stage by adding a weighted function $\rho_m h(x, a_m)$, where $h(x, a_m)$ is invariably a tree based learner and ρ_m is a weight coefficient. Basically, the objective of GBM is to minimize the expectation of the loss function [18]. Finally, the target model is obtained by iteratively conducting the previous steps. Squared, Laplace, and T-distribution loss functions are the most popular used loss functions in GBM [19], and the Laplace distribution [20] is selected as the loss function in this study. The pseudo-code of the GBM is illustrated in Algorithm 1 [21]. In this study, in addition to the estimated feeder PV power generation, the explanatory variables include humidity, cloud coverage, head index, temperature, and wind speed. Based on the extracted extreme weather days, two set of GBM models are trained: one for extreme hot weather days and the other one for extreme cold weather days.

III. Case Study and Results

A. Data Summary

The developed load forecasting approach is applied to three real distribution feeders (all with high distributed PV penetration) in Texas. The provided load and weather forecasts data for those three feeders are at a 1-hour resolution ranging from August 13rd 2017 to May 31st 2019. Approximately 3% of the historical load data is missing or abnormal and is interpolated using neighboring observations. The duration of the collected data and feeder information are summarized in Table I. In this study, days with temperature higher than $95°F$ denote extreme hot days, while days with temperature lower than $36°F$ denote extreme cold days. Based on these thresholds, after data extraction, the length of extreme hot days ranges from 78 days to 144 days, and the length of extreme cold days ranges from 36 to 61 days. For all the three feeders, the first 3/4 of the extracted data is used as training data. The accuracy of the forecasts is evaluated by the remaining 1/4 of data. In this study, the forecasting lead time is the same as the weather forecasts.

TABLE I: Data summary of the selected feeders

Feeder	Data duration	η (%)	No. H	No. C
C1	2017-08-13 to 2019-05-31	6.48	144	36
C2	2017-08-13 to 2019-05-31	5.36	102	55
C3	2017-08-13 to 2019-05-31	5.08	78	61

Note: No. H and No. C are the abbreviations for the number of extreme hot and cold days, respectively.

B. Load Forecasting Results

Three evaluation metrics are used to assess the forecasting accuracy, which are the normalized root mean squared error (nRMSE), normalized mean absolute error (nMAE), and mean absolute percentage error (MAPE). The mathematical expressions of the three metrics are expressed as:

$$nRMSE = \frac{1}{y_{max}} \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \quad (3)$$

$$nMAE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_{max}} \right| \quad (4)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

where $\hat{y}$, y, and y_{max} are the forecast load, actual load, and maximum actual load, respectively.

For these metrics, a smaller value indicates better forecasting performance. Deterministic forecasting errors of models with and without BTM PV contributions at the selected feeders are summarized in Table II and Table III. For hot weather days, the forecasting NMAE, NRMSE, and MAPE are in the ranges of 2%-7%, 8%-12%, and 3%-10% respectively. Similarly, for cold weather days, it is shown that the forecasting NMAE, NRMSE, and MAPE are in the ranges of 6%-8%, 14%-23%, and 11%-17%, respectively. The model without considering BTM PV contributions (GBM-w/) is used as a baseline. Overall, the load forecasting by considering BTM PV performs better than the benchmark GBM-w/ model, which shows the importance of considering BTM PV in load forecasting. Figure 2 shows a linear relationship between the MAPE improvement (by considering BTM PV) and PV penetration under both extreme hot weather days and extreme cold weather days. It is interesting to observe that a feeder with higher PV penetration has a higher chance to be improved by considering the contribution of BTM PV. It is seen that the MAPE improvement in extreme hot weather days could be more than 200%. However, the improvement in extreme cold days by considering BTM PV are not significant (i.e., 1-3%). This is mainly due to that PV power generation is normally higher in extreme hot days than in extreme cold days. In addition, the forecasting error of extreme cold weather days are higher than that of the extreme hot weather days. These are mainly because the number of extreme cold days are limited in Texas, which results in a shorter length of training data for the forecasting model under extreme cold days.

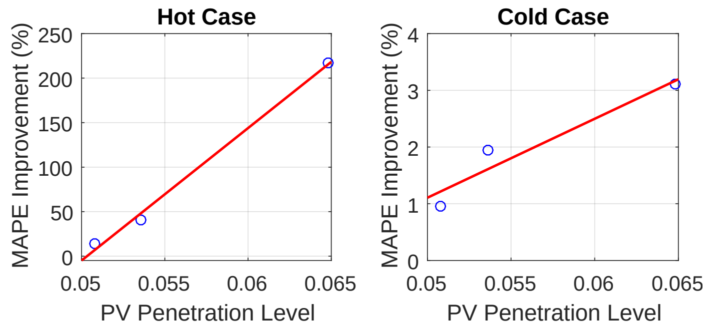

Fig. 2: Relationship between MAPE improvement and PV penetration

TABLE II: Load forecasting performance under extreme hot weather

Model	Metric	Feeder C1	Feeder C2	Feeder C3
GBM	NMAE(%)	2.24	5.20	6.33
	NRMSE(%)	8.01	11.66	12.70
	MAPE(%)	3.47	7.57	9.38
GBM-w/	NMAE(%)	6.93	7.03	7.11
	NRMSE(%)	11.76	11.43	11.39
	MAPE(%)	11.00	10.66	10.72

TABLE III: Load forecasting performance under extreme cold weather

Model	Metric	Feeder C1	Feeder C2	Feeder C3
GBM	NMAE(%)	6.57	7.30	7.38
	NRMSE(%)	14.82	18.33	22.90
	MAPE(%)	11.93	15.12	16.80
GBM-w/	NMAE(%)	8.33	10.21	10.98
	NRMSE(%)	16.95	20.30	28.43
	MAPE(%)	12.30	15.42	16.96

C. Sensitivity Analysis

To further explore the sensitivity of the forecasting accuracy to the temperature threshold, we have tested different threshold values for choosing extreme hot and cold training days. The forecasting days are still selected based on the $95°F$ and $36°F$ values for hot and cold days, respectively. For example, for the extreme hot weather case, if now the temperature threshold is $94°F$, only those days where the temperature is higher than $94°F$ are selected for training, while in the forecasting process, the threshold is maintained as $95°F$. For the extreme hot weather case, we have explored 10 temperature thresholds ranging from $90°F$ to $99°F$. For the extreme cold weather case, we have also explored 10 temperature thresholds, ranging from $32°F$ to $41°F$. The results of the three feeders are shown in Fig. 3 and Fig. 4. Each sub-figure shows the relationship between the temperature threshold and the forecasting MAPE. It is seen from Fig. 3 that the forecasting accuracy of all the 3 feeders under extreme hot weather days are first increased by increasing the temperature threshold and then decreased.

Similar findings are also observed in the case of extreme cold weather days. Overall, there exists an optimal temperature threshold for selecting days for the the optimal training model, and this optimal threshold varies with the feeder.

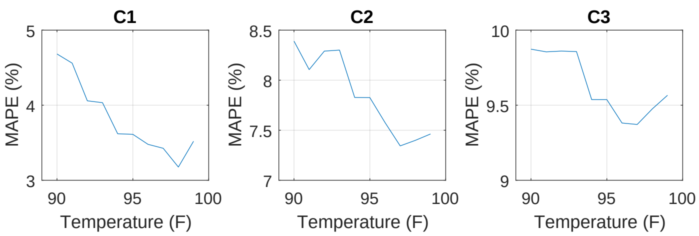

Fig. 3: MAPE of selected feeders for different temperature thresholds in the hot weather case

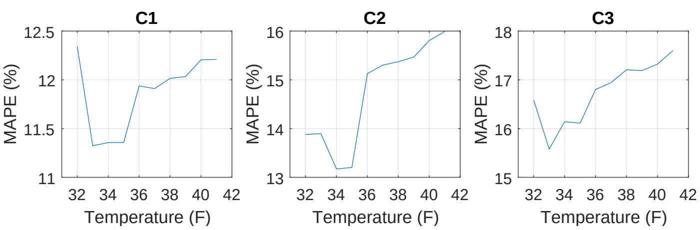

Fig. 4: MAPE of selected feeders for different temperature thresholds in the cold weather case

IV. Conclusion

This paper developed a data-driven feeder-level load forecasting method by taking account of BTM PV contribution. Results of the case study under extreme weather conditions at three selected feeders showed that: (i) the developed load forecasting method has shown improved forecast accuracy by considering BTM PV contribution under both extreme hot and cold weather conditions; (ii) there exists an optimal temperature threshold for selecting days for the training model; (iii) there exists a linear relationship between the forecast improvement by considering BTM PV and the PV penetration. Future work will explore: (i) calendar effects on feeder level load forecasting with BTM PV, (ii) spatio-temporal correlations among different feeders, and (iii) probabilistic load forecasting with BTM PV.

V. Acknowledgement

This work was supported by the Oncor Electric Delivery Company LLC.

References

[1] H. Shaker, H. Zareipour, and D. Wood, "Estimating power generation of invisible solar sites using publicly available data," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2456–2465, 2016.

[2] D. J. Feldman and R. M. Margolis, "Q4 2018/q1 2019 solar industry update," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2019.

[3] A. A. Girgis and S. Varadan, "Unit commitment using load forecasting based on artificial neural networks," *Electric power systems research*, vol. 32, no. 3, pp. 213–217, 1995.

[4] Y. Wang, Q. Xia, and C. Kang, "Unit commitment with volatile node injections by using interval optimization," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1705–1713, 2011.

[5] A. S. Nair, M. Campion, D. Hollingworth, and P. Ranganathan, "Two-stage load forecasting for residual reduction and economic dispatch using pjm datasets," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*. IEEE, 2018, pp. 0691–0695.

[6] H. Jiang, F. Ding, and Y. Zhang, "Short-term load forecasting based automatic distribution network reconfiguration," in *2017 IEEE Power & Energy Society General Meeting*. IEEE, 2017, pp. 1–5.

[7] F. Ding and K. A. Loparo, "Feeder reconfiguration for unbalanced distribution systems with distributed generation: A hierarchical decentralized approach," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1633–1642, 2015.

[8] F. He, Y. Gu, J. Hao, J. J. Zhang, J. Wei, and Y. Zhang, "Joint real-time energy and demand-response management using a hybrid coalitional-noncooperative game," in *2015 49th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2015, pp. 895–899.

[9] M. Sun, C.-L. Chang, J. Zhang, A. Mehmani, and P. Culligan, "Break-even analysis of battery energy storage in buildings considering time-of-use rates," in *2018 IEEE Green Technologies Conference (GreenTech)*. IEEE, 2018, pp. 95–99.

[10] K. Metaxiotis, A. Kagiannas, D. Askounis, and J. Psarras, "Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher," *Energy conversion and Management*, vol. 44, no. 9, pp. 1525–1534, 2003.

[11] S. S. Reddy, P. Bijwe, and A. R. Abhyankar, "Joint energy and spinning reserve market clearing incorporating wind power and load forecast uncertainties," *IEEE Systems Journal*, vol. 9, no. 1, pp. 152–164, 2015.

[12] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, no. 3, pp. 896–913, 2016.

[13] K. G. Boroojeni, M. H. Amini, S. Bahrami, S. Iyengar, A. I. Sarwat, and O. Karabasoglu, "A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon," *Electric Power Systems Research*, vol. 142, pp. 58–73, 2017.

[14] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: A big data approach," *International Journal of Forecasting*, vol. 32, no. 3, pp. 585–597, 2016.

[15] Y. Liang, D. Niu, and W.-C. Hong, "Short term load forecasting based on feature extraction and improved general regression neural network model," *Energy*, vol. 166, pp. 653–663, 2019.

[16] A. Kankiewicz and E. Wu, "Integration of behind-the-meter pv fleet forecasts into utility grid system operations," *3rd ICEM*, 2015.

[17] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[18] C. Feng, M. Cui, B.-M. Hodge, and J. Zhang, "A data-driven multi-model methodology with deep feature selection for short-term wind forecasting," *Applied Energy*, vol. 190, pp. 1245–1257, 2017.

[19] C. Feng, M. Sun, and J. Zhang, "Reinforced deterministic and probabilistic load forecasting via q-learning dynamic model selection," *IEEE Transactions on Smart Grid*, 2019.

[20] G. Ridgeway, "Generalized boosted models: A guide to the gbm package," *Update*, vol. 1, no. 1, p. 2007, 2007.

[21] L. Silva, "A feature engineering approach to wind power forecasting: Gefcom 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 395–401, 2014.