Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 9 (2023) 115–122
www.elsevier.com/locate/egyr

2022 3rd International Conference on Power, Energy and Electrical Engineering (PEEE 2022)
18–20 November 2022

Short-term net load forecast in distribution networks with PV penetration behind the meter

Alberto Falces, Candido Capellan-Villacian, Montserrat Mendoza-Villena, Pedro J. Zorzano-Santamaria, Pedro M. Lara-Santillan, Eduardo Garcia-Garrido, L. Alfredo Fernandez-Jimenez*, Enrique Zorzano-Alba

Electrical Engineering Department, University of La Rioja, Logroño 26004, Spain

Received 14 December 2022; accepted 27 December 2022
Available online xxxx

Abstract

In recent years there has been a strong expansion of photovoltaic (PV) distributed generation systems. A high PV penetration level can cause uncertainty in the operation and management processes carried out by electric utilities, since most meters register the net load, i.e., the actual load minus the power generated by the PV systems behind the meter. The goal of this paper was to analyze the difference in the net load forecasting error achieved by models using or not using behind-the-meter PV generation data. The PV plant is connected to the lower voltage side of the power substation, representing a penetration level of more than 35% of the total load. The study shows that the best forecasting results are obtained with an indirect approach using two forecasting models, one for the total load and the other for the PV generation. However, the difference with respect to the results obtained with a unique net load forecasting model is almost negligible, which may be of special interest for power system distributors or other agents who do not have access to behind-the-meter generation data.

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Power, Energy and Electrical Engineering, PEEE, 2022.

Keywords: Net load forecast; Behind the meter; PV penetration

1. Introduction

Distributed renewable energy generation, especially local photovoltaic (PV) generation, has expanded in recent years and is expected to increase its expansion in the near future [1]. A high PV penetration level can cause uncertainty in the operation and management processes carried out by electric utilities, since most meters register the net load, i.e., the actual load minus the power generated by the PV systems behind the meter. This situation makes difficult the forecast of the net load [2], necessary to properly operate the power distribution system, due to two fundamental causes: The invisibility of PV generation for distribution system operators (DSOs), which prevents

* Corresponding author.
E-mail address: luisalfredo.fernandez@unirioja.es (L.A. Fernandez-Jimenez).

https://doi.org/10.1016/j.egyr.2022.12.103

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Power, Energy and Electrical Engineering, PEEE, 2022.

it from being quantified when it is produced, and the stochastic nature of PV generation, which makes it difficult to find stable generation patterns (solar irradiance, cloud cover, cell temperature, etc.).

Therefore, it is crucial for DSOs to have tools at their disposal that allow them to forecast with a certain degree of reliability the load demanded at any given time of the day. Many methods for net load forecast with PV penetration behind the meter (BTM) have been proposed in the international literature. These methods can be classified into statistical methods (linear regression, ARMA, exponential smoothing, etc.) and methods based on Machine Learning (ML) techniques such as artificial neural networks (ANNs), deep learning, random forest (RF), etc. There are also hybrid methods that combine or integrate more than one technique. The selection of the method or technique to be applied to obtain the net load forecast depends on the available data and the characteristics of the solution sought (granularity, accuracy, etc.). Based on the available data, short-term net load forecasting (STNLF) can be addressed by two approaches: direct and indirect. The indirect approach determines the net load forecast as the difference of load and PV generation forecasts, while the direct approach directly predicts the net load without any further intermediate steps.

In the international literature related to STNLF, most authors propose models using the techniques described above. Further analysis shows that some authors use estimated data for their models [3–7], other use real load or generation data in their approaches [8–15]. Only some of them compare the results of their models with those obtained with other reference models [3,8,10–14]. While the literature on STNLF models is growing, there are limited works available about the search of the adequate technique to obtain the best forecast with different PV penetration levels.

Among the works that describe STNLF models by the direct approach, the following can be highlighted: Zhang et al. [8] establish a comparison between 3 different hourly STNLF models with a horizon of one day, specifically models based on Linear Regression (LR), RF and Gradient Boosting. The main handicap of their method is that they use as explanatory variable the PV installed capacity and not an estimation of the PV generation; this affects the accuracy of the forecast. Shaker et al. [9] propose a method based on Fuzzy Arithmetic Wavelet Neural Networks (FAWNN) that enables to forecast the net load in a region with high PV penetration level. The proposed model uses historical PV power generation data from a limited number of representative locations in the region and forecasts of weather variables obtained by a Numerical Weather Prediction (NWP) model. Although the forecasting error is low, they do not compare their results with those obtained by any other method. Razavi et al. [12] propose an *ad-hoc* model based on a recurrent neural network that allows to obtain accurate net load forecasts from various load profiles. They do not carry out any comparison of the forecasting results achieved by the proposed model.

There are several published works that describe STNLF models developed with the indirect approach, among them the following can be highlighted: Saeedi et al. [10] use weather data, minimal PV generation data, and PV plant location to estimate the PV power generation. Then, they use the estimated PV generation and measured net load to estimate the total actual load at each instant of time. They apply several techniques, obtaining RF the best results. Sun et al. [6] determine the impact of PV penetration level on the performance of STNLF models for a distribution system. They estimate PV power generation using weather data and calculate the total load; afterwards they decompose load and weather data using wavelets and forecast both individually; finally, they add both partial forecasts to get the final forecast. Aponte et al. [11] address a methodology based on ML to forecast peak loads with and without renewable energy generation BTM. They separately predict load and PV generation calculating afterwards net load. Landelius et al. [13] evaluate direct and indirect approaches to predict net load. In the direct approach, they apply ML techniques (LR and ANNs) to forecast the net load without any data on the total load. In the indirect approach, they use a load forecasting model based on the open-source software PVLIB to subsequently forecast the net load. Mejía et al. [14] introduce an interesting novelty in the forecast of net load: they use the solar distributed generation as a regressor variable, that is, as another explanatory variable to be included in the STNLF models. In [15] the authors propose a method based on online load and PV generation forecasts to predict the residual load in a building. They choose the persistent model to evaluate the accuracy of their proposed model.

As mentioned above, in published works related to STNLF models we can find models based on the direct approach [8,9,12,13] and models based on the indirect approach [6,10,11,13–15]. However, the results of these models, in general, are compared with the ones obtained with other models developed with the same approach, but not with those obtained with models developed with the opposite approach. It should be noted that, in most cases, the authors do not have the data to develop models using both approaches (PV generation is unknown), but the question remains: which approach would provide the best STNLF results?

This paper presents a comparative study of the forecasting results obtained by STNLF models developed with 4 popular ML techniques using the two approaches. For the development of the models, we used data from a real power substation that include all the hourly electric load of a small town and the hourly generation data of a PV plant connected to the substation and whose generation represents a penetration level of 35.6%. Three possible scenarios have been considered in the comparisons of the obtained results, corresponding to penetration levels of 0, 17.8 and 35.6%. The forecasts provided by the models are the 24 hourly expected values for the output variable, which can be total load, PV generation or net load, for the following day (the prediction is carried out in the first hours of the previous day), which means that the forecasting horizon ranges from 24 to 48 h. A wide set of explanatory variables have been used for the development of the models. The tests carried out show that the best forecasting results are obtained with the indirect approach. However, the difference with respect to the results obtained by applying the direct approach are relatively small, which may be of particular interest for agents in the electric sector interested in net load forecasting, but who do not have access to data of PV generation behind the meter.

The paper is structured as follows: Section 2 presents the four techniques selected to develop the STNLF models. Section 3 presents a case study with the results obtained in the forecast of the net load corresponding to a power substation with PV generation behind the meter. Finally, Section 4 presents the conclusions.

2. Selected techniques

Four ML techniques were selected to develop all the short-term forecasting models needed in our study. These techniques included: RF, Extreme Gradient Boosting (XGB), Support Vector Regression with Radial Basis function Kernel (SVR), and Cubist (CUB).

RF is a ML algorithm based on an ensemble method, which creates a series of small decision trees from randomly selected data samples. The model combines the forecast obtained by each tree to obtain an accurate final forecast [16].

XGB is a fast implementation of the gradient boosting technique. The basis is to generate multiple "simple" decision trees sequentially to achieve a final model. This is done by iteratively adding the new trees and fitting the residuals of the prior model so that more accurate results are obtained in each iteration [17].

SVR is the support vector machine adaptation for regression tasks [18]. The Radial Basis Function Kernel [19] is used to transform the original input space (n-dimensional) into an m-dimensional input space, where m is much higher than n, and then applies the dot product efficiently. The goal is to obtain a linear regression curve in the higher dimensional space that converts to a nonlinear regression curve in the lower dimensional space.

Cubist [20] is a rule-based model with added instance-based corrections, as an extension of the M5 model tree. The tree is reduced to a set of rules. The rules are simplified by pruning or combining. CUB is a suitable algorithm for creating rule-based models that provide accurate forecasts with clarity requirements.

In order to establish comparisons according to different PV penetration levels, three scenarios were considered:

- Scenario 1. This scenario corresponds to a PV penetration level of 0%. It is considered that there is no PV generation, so the net load coincides with the total load.
- Scenario 2. This scenario corresponds to a PV penetration level of 17.8%. Half of the PV generation is considered. With each of the selected techniques, two other short-term forecasting models need to be developed: one for PV generation and another for net load.
- Scenario 3: This scenario corresponds to a PV penetration level of 35.6%. The total PV generation is considered. As in scenario 3, two short-term forecasting models must be developed with each of the 4 selected techniques.

The forecasting models for the total load in scenarios 2 and 3 are the same that the ones developed for scenario 1 since the total load does not change in the three scenarios. A total of 20 models has to be developed: 4 short-term total load forecasting models, one with each technique; 4 STNLF models for scenario 2 and other 4 for scenario 3; 4 short-term PV generation forecasting models for scenario 2 and other 4 for scenario 3. All models were developed following an optimization process by choosing the most appropriate values of their parameters (hyperparameters) to avoid overfitting and to increase the generalization of the models. This optimization was performed by a grid search procedure on a set of possible candidate values for the parameters, selecting those that offered a lower average forecasting error (mean square error) with the 5-fold cross-validation procedure.

The objective of the model development is the forecast of the net load in the three scenarios. This value is directly provided by 4 models in each scenario, one for each of the techniques used. The outputs of these 4 models correspond to the forecasts of net load using the direct approach. A second possibility is the forecast of net load using the indirect approach. In this case, the net load forecast for hour t, $\hat{L}_{net-ind}(t)$, can be determined as the difference between the values provided by a total load forecasting model for hour t, $\hat{L}_{total}(t)$, and a PV generation forecasting model for hour t, $\hat{P}_{PV}(t)$, as shown in (1).

$$\hat{L}_{net-ind}(t) = \hat{L}_{total}(t) - \hat{P}_{PV}(t) \tag{1}$$

Thus, for each of the three scenarios and for each of the 4 selected techniques, two STNLF models are available, one using the direct approach and the other using the indirect approach. The comparison of the results obtained by models of one approach or another, can give us an answer to the question raised in the introduction and related to which approach would provide the best forecasting results for the net load.

Three indexes were used to evaluate the forecasting performance of the models. They are the Root Mean Square Error (RMSE), defined in (2), the Mean Absolute Error (MAE), defined in (3) and the Mean Absolute Percentage Error (MAPE), defined in (4),

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=h_1}^{h_2}\left(\hat{y}(t) - y(t)\right)^2} \tag{2}$$

$$MAE = \frac{1}{N}\sum_{t=h_1}^{h_2}\left|\hat{y}(t) - y(t)\right| \tag{3}$$

$$MAPE = \frac{1}{N}\sum_{t=h_1}^{h_2}\frac{\left|y(t) - \hat{y}(t)\right|}{y(t)} \tag{4}$$

where $\hat{y}(t)$ represents the output of the forecasting model for hour t, y(t) represents the actual value for hour t, h_1 and h_2 the first and the last hour, and N the total number of hours in the testing period.

Since each of the techniques selected for the development of the prediction models had parameters whose values could be tuned or optimized to avoid overfitting to the training data and to increase the generalization of the models, the parameter tuning with repeated grid-search cross-validation technique was used [21]. This technique consists of evaluating the forecasting results achieved by a set of models defined by the values of the parameters contained in a search grid applying k-fold cross-validation procedure a specified number of times. The technique tests all combinations of parameters and selects the set that achieves the lowest average RMSE with all the repeated k-folds.

3. Case study

The load and PV generation data needed for this study were obtained from the metering records of a 66/13.2 kV power substation located in the north of Spain, which feeds a small town of about 5000 inhabitants. The entire town electric grid is connected to the low voltage side of the substation, so the 66 kV grid operator only records data corresponding to the substation's net load. The 13.2 kV network, operated by another distribution operator, supports around 3000 consumers, mainly residential ones, although the largest share of consumption corresponds to industrial consumers. A PV plant with an installed capacity of 2 MW is connected to the 13.2 kV grid and is located just 4 km from the town, and close to the substation. The data available from the power substation correspond to hourly records of the total load supplied to the 13.2 kV grid and the hourly generation records of the PV plant measured at the connection point in the substation. The data covers the period between October 1, 2008 and March 31, 2011 (30 months). For the development of the forecasting models, the data set was completed with the following variables:

- Forecasts of weather variables. The weather forecasts for the period considered corresponded to the values predicted in the early hours of the day for the 24 h of the next day. The forecasts were downloaded from the web server of a regional weather forecast service that provides hourly forecasts, from a NWP model, of a set of weather variables for a grid of points on the Earth's surface located about 12 km apart from each other. The forecast of each weather variable for the town location was calculated as the weighted average of the values for the four nearest points of the grid, using as weighting factor the inverse of the square distance.

- Dummy variables. They are related to the hour for which the forecast is carried out: hour of the day, day of the week, month of the year, holidays and European summer time.
- Load values in the previous days. It is well known that one of the most relevant explanatory variables in short-term load forecasting is the hourly load in the previous days, so the database was extended with the values of total and net load values 48, 72, 96, 120, 144 and 168 h before the forecasting horizon. The forecast is carried out the previous day, so the last known value of the load for any hour of the day is the one 48 h before.
- Finally, the data set was completed with three more variables related to the position of the sun with respect to the PV plant (practically the same as that of the town).

The variables contained in the database are listed in Table 1. The most appropriate variables for the development of each of the three types of models were selected from the database. Thus, for the development of the total load forecasting models, the explanatory variables were seven of the weather forecasts, all the dummy variables and the hourly total load values lagged 48 to 168 h. For the development of the forecasting model of the generation in the PV plant, all the weather forecasts and the three variables related to the sun position were used. For the STNLF model, all the explanatory variables used in the forecast of the PV generation were used together with those used for the total load forecast, although instead of using the lagged values of this total load, the lagged values of the net load from 48 to 168 h were included as explanatory variables. Thus, the total load forecasting models used 55 explanatory variables, the PV generation forecasting models used 16 variables and the net load forecasting models (direct approach) used 64 variables. The complete dataset was divided in training dataset and testing dataset. The training dataset comprised the period from October 2008 to September 2010 (24 months), and the testing dataset comprised the period from October 2010 to March 2011 (6 months). All the hourly records were used to develop the short-term forecasting models of the total load and net load (direct approach). In the case of the short-term PV generation forecasting models, only the records corresponding to hours with sunlight were used, since for the rest of the hours the generation is null.

Table 1. Explanatory variables.

Denomination	Variables	Direct approach	Indirect approach	
			Load	PV generation
V1	Temperature (K)	×	×	×
V2	Global horizontal irradiance (W/m^2)	×	×	×
V3	Wind speed (m/s)	×	×	×
V4	Wind direction (°)	×		×
V5	Pressure (hPa)	×		×
V6	Relative humidity (per unit)	×	×	×
V7	Total cloud cover (per unit)	×	×	×
V8	Cloud cover at low levels (per unit)	×		×
V9	Cloud cover at mid levels (per unit)	×		×
V10	Cloud cover at high levels (per unit)	×		×
V11	Visibility (m)	×		×
V12	Rainfall (kg/m^2)	×	×	×
V13	Snow (kg/m^2)	×	×	×
V14	European summer time (logical)	×	×	
V15–V20	Net load (kW) lag_i hourly net demand lagged "i" hours, i = 48,...,168	×		
V21–V26	Load (kW) lag_i hourly load lagged "i" hours, i = 48,...,168		×	
V27	Solar altitude (rad)	×		×
V28	Solar azimuth (rad)	×		×
V29	Extra-terrestrial solar irradiance (W/m^2)	×		×
V30–V52	H2, H3,..., H24 Hourly dummy variables for the hour of the day	×	×	
V53–V58	D2, D3,..., D7 Hourly dummy variables for day of the week	×	×	
V59–V69	M2, M3,..., M12 Hourly dummy variables for the month of the year	×	×	
V70	F1 Hourly dummy variable for national, regional or local holiday	×	×	

4. Results and discussions

First, the total load forecasting models were developed, which are the same for the three scenarios. The optimal values of the parameters (or hyperparameters) that define each of the models were selected by the repeated grid-search cross-validation technique using 5 repetitions with 5-fold. After the selection of the optimal values of the parameters, each model was trained using all the data of the training dataset and used to provide the forecasts for the data of the testing dataset. Table 2 shows the results obtained by the 4 models in the short-term forecasting of the total load for the testing period. The model that obtained the best results was the CUB model, with the best error indexes values, and achieving a MAPE of 4.233%. The second one with the best error indexes was the RF model, with values relatively close to those obtained by the CUB model.

Table 2. Total load forecasting results for the testing period.

Model	RMSE (kW)	MAE (kW)	MAPE (%)
RF	216.748	157.766	4.250
XGB	219.918	162.311	4.370
CUB	**211.530**	**156.244**	**4.233**
SVR	227.295	169.132	4.599

The next models to be developed were those corresponding to scenario 2 (PV penetration level of 17.8%). Net load (direct approach) and PV generation forecasting models were developed. From the forecasts of the latter, the net load (indirect approach) was calculated using (1) for every hour, considering that in the hours without sunlight the PV generation was null. Note that the total load forecasting models needed to apply Eq. (1) are those developed for scenario 1 (results in Table 2). Table 3 shows the results obtained by the 4 PV generation forecasting models, by the 4 STNLF models with the direct approach, and the results obtained with the indirect approach. As is shown in Table 3, the model that achieved the best forecasting results for the net load with the direct approach was the CUB, with a MAPE value of 5.407%. However, with the indirect approach the RF model achieved better results, with a MAPE of 5.104%, as consequence of its better performance in PV generation forecast. The SVR model, which presented good results in the forecast of PV generation, achieved worse results in the net load forecast with the indirect approach than the obtained by RF model as a consequence of the better results obtained by the latter in forecasting the total load.

Table 3. Forecasting results for scenario 2 (PV penetration level of 17.8%).

Model	PV forecasting		Net load forecasting (direct)			Net load forecasting (indirect)		
	RMSE (kW)	MAE (kW)	RMSE (kW)	MAE (kW)	MAPE (%)	RMSE (kW)	MAE (kW)	MAPE (%)
RF	**195.719**	145.518	269.807	193.367	5.490	**244.389**	**178.420**	**5.104**
XGB	200.675	154.137	270.327	195.408	5.600	249.503	183.635	5.246
CUB	207.247	150.270	**260.659**	**188.687**	**5.407**	246.888	179.883	5.167
SVR	197.081	**139.544**	326.411	236.148	6.807	258.494	188.404	5.427

Table 4. Forecasting results for scenario 3 (PV penetration level of 35.6%).

Model	PV forecasting		Net load forecasting (direct)			Net load forecasting (indirect)		
	RMSE (kW)	MAE (kW)	RMSE (kW)	MAE (kW)	MAPE (%)	RMSE (kW)	MAE (kW)	MAPE (%)
RF	**392.040**	**289.600**	353.675	242.766	7.781	**328.944**	**227.931**	**7.331**
XGB	403.739	296.477	346.149	239.843	7.685	339.105	235.147	7.582
CUB	405.580	294.002	**342.583**	**236.053**	**7.637**	337.618	231.076	7.427
SVR	402.428	290.934	363.305	251.513	8.077	347.898	239.235	7.652

Afterwards, the forecasting models corresponding to scenario 3 (PV penetration level of 35.6%) were developed following the same methodology used for scenario 2. Table 4 shows the results obtained by the short-term forecasting models, and the results obtained with the direct and indirect approaches. As is shown in Table 4, the model that achieved the best forecasting results for the net load with the direct approach was again the CUB model, with a MAPE value of 7.637%. Also again, with the indirect approach the RF model achieved better results, with a MAPE of 7.331%.

Comparing the results in Tables 2–4, interesting conclusions can be drawn. The error in the forecast of net load increases with the level of PV penetration. This occurs for both approaches and for the models of the 4 techniques, although the increases in error indexes with the PV penetration level are slightly smaller in the indirect approach. For example, for the technique that achieves the best results with the direct approach (CUB), the RMSE error increases from 211.53 kW in scenario 1 (without PV generation) to 260.659 kW in scenario 2 and to 342.583 kW in scenario 3, which represents an increase near to 23% and 62%, respectively. This same technique, in the indirect approach, presents increases in RMSE in scenarios 2 and 3 close to 17% and 60%, respectively. This increase in error indexes is a logical consequence of the stochastic nature of PV generation. Similar findings were obtained in [3] on the increase of the prediction error with increasing PV penetration level, although the authors used synthetic data instead real data as in our study. The results shown in the previous tables can help to answer the question posed in the introduction: The indirect approach provides better forecasting results than the direct approach. This occurs in the two scenarios with PV generation (scenarios 2 and 3) and for the models of the 4 techniques.

The CUB model provides the best forecasting results for the net load with the direct approach. The CUB model is the best in scenario 1 (net load coincides with total load as there is no PV generation), and it is the best for scenarios 2 and 3. However, in the indirect approach, the RF model improves the forecasting results of the CUB model. This better performance of the RF model is due to its superiority in forecasting PV generation. For example, in scenario 2, the RMSE value for PV generation forecast of the CUB model is 3.45% higher than that of the RF model. Overall, as expected, the errors in the prediction of PV generation in scenario 2 are about half those of scenario 3 for all models.

Finally, if the results obtained by the best model with the direct approach are compared with those of the best model with the indirect approach, the differences are relatively small. For example, for scenario 3, the difference in the RMSE between the CUB model with the direct approach and the RF model with the indirect approach is 4.14% (342.583 kW vs. 328.944 kW). The same is true if we compare the RMSE and MAPE indexes for the best models of both approaches in scenarios 2 and 3. This relative similarity of forecasting results obtained with the direct and indirect approaches reaffirms with real data the conclusion reported in [5], where the authors suggest that if the net demand forecasting model includes as explanatory variables all those relevant for solar energy, it might not be necessary to obtain a PV production forecast beforehand if the PV capacity remains unchanged.

5. Conclusions

This paper presents a comparative study of the results obtained in the forecast of the net load for a power substation with PV generation BTM. The data used in the study correspond to those measured during 30 months in a real substation: hourly values of net load, total load, and PV generation. The hourly data measured at the power substation were completed with a set of explanatory variables that include forecasts of weather variables, dummy variables related to the calendar, and variables related to the relative position of the sun with respect to the PV plant. Three scenarios have been considered, corresponding to PV penetration levels of 0, 17.8% and 35.6%. The forecast of the net load for the 24 h of the next day is performed using two approaches: the direct approach that consists of using a single model that provides as output the net load forecast, and the indirect approach that consists of using two forecasting models, one for the load and the other for the PV generation, obtaining the net load forecast as the difference of the forecasts of both models. Four popular ML techniques, widely used in load and PV generation forecasting, have been chosen for the development of the forecasting models needed in the study.

The results obtained for a testing period of 6 months show that the indirect approach provides better forecasts than the direct approach, however, the forecasting models developed with the direct approach, using those related to PV generation as explanatory variables, obtain results that are relatively close to those obtained with the indirect approach, which confirms conclusions postulated in previous works. This relative similarity of results may be of interest to electricity distribution industry players who wish to obtain net load forecasts but have not PV generation BTM data at their disposal.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This research is a part of the project Afianza 2021/03 funded by Government of La Rioja, Spain, and the grant RED2018-102618-T funded by MCIN/AEI/10.13039/501100011033, Spain.

References

[1] Haque MM, Wolfs P. A review of high PV penetrations in LV distribution networks: Present status, impacts and mitigation measures. Renew Sustain Energy Rev 2016;62:1195–208. http://dx.doi.org/10.1016/j.rser.2016.04.025.

[2] Erdener BC, Feng C, Doubleday K, Florita A, Hodge B-M. A review of behind-the-meter solar forecasting. Renew Sustain Energy Rev 2022;160:112224. http://dx.doi.org/10.1016/j.rser.2022.112224.

[3] Wang Y, Zhang N, Chen Q, Kirschen DS, Li P, Xia Q. Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV. IEEE Trans Power Syst 2018;33:3255–64. http://dx.doi.org/10.1109/TPWRS.2017.2762599.

[4] Cha JW, Joo SK. Probabilistic short-term load forecasting incorporating behind-the-meter (BTM) photovoltaic (PV) generation and battery energy storage systems (BESSs). Energies 2021;14. http://dx.doi.org/10.3390/en14217067.

[5] Haupt SE, Dettling S, Williams JK, Pearson J, Jensen T, Brummet T, Kosovic B, Wiener G, McCandless T, Burghardt C. Blending distributed photovoltaic and demand load forecasts. Sol Energy 2017;157:542–51. http://dx.doi.org/10.1016/j.solener.2017.08.049.

[6] Sun X, Jin C. Impacts of solar penetration on short-term net load forecasting at the distribution level. In: Proc. 2021 IEEE 4th int. electr. energy conf. CIEEC 2021. 2021, http://dx.doi.org/10.1109/CIEEC50170.2021.9510770.

[7] Rouwhorst G, Nguyen P, Slootweg H. A hybrid supervised learning model for a medium-term MV/LV transformer loading forecast with an increasing capacity of PV panels. In: 2021 IEEE madrid powertech, powertech 2021 - conf. proc.. 2021, http://dx.doi.org/10.1109/PowerTech46648.2021.9494854.

[8] Zhang N, Li Z, Zou X, Quiring SM. Comparison of three short-term load forecast models in southern California. Energy 2019;189:116358. http://dx.doi.org/10.1016/j.energy.2019.116358.

[9] Shaker H, Manfre D, Zareipour H. Forecasting the aggregated output of a large fleet of small behind-the-meter solar photovoltaic sites. Renew Energy 2020;147:1861–9. http://dx.doi.org/10.1016/j.renene.2019.09.102.

[10] Saeedi R, Sadanandan SK, Srivastava AK, Davies KL, Gebremedhin AH. An adaptive machine learning framework for behind-the-meter load/PV disaggregation. IEEE Trans Ind Inform 2021;17:7060–9. http://dx.doi.org/10.1109/TII.2021.3060898.

[11] Aponte O, McConky K. Peak electric load days forecasting for energy cost reduction with and without behind the meter renewable electricity generation. Int J Energy Res 2021;45:18735–53. http://dx.doi.org/10.1002/er.6990.

[12] Razavi SE, Arefi A, Ledwich G, Nourbakhsh G, Smith DB, Minakshi M. From load to net energy forecasting: Short-term residential forecasting for the blend of load and PV behind the meter. IEEE Access 2020;8:224343–53. http://dx.doi.org/10.1109/ACCESS.2020.3044307.

[13] Landelius T, Andersson S, Abrahamsson R. Modelling and forecasting PV production in the absence of behind-the-meter measurements. Prog Photovolt Res Appl 2019;27:990–8. http://dx.doi.org/10.1002/pip.3117.

[14] Mejia-Giraldo D, Casadiegos-Osorio A, Grajales-Espinal C, Tobon-Villa JE. Very-short term electricity demand forecasting integrating solar distributed generation. In: 2021 IEEE int. conf. eng. veracruz, ICEV 2021. 2021, http://dx.doi.org/10.1109/ICEV52951.2021.9632643.

[15] Telle JS, Maitanova N, Steens T, Hanke B, Von Maydell K, Grottke M. Combined PV power and load prediction for building-level energy management applications. In: 2020 15th int. conf. ecol. veh. renew. energies, EVER 2020. 2020, http://dx.doi.org/10.1109/EVER48776.2020.9243026.

[16] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York, NY: Springer New York; 2009, http://dx.doi.org/10.1007/978-0-387-84858-7.

[17] Chen T, Guestrin C. Xgboost. In: Proc. 22nd ACM SIGKDD int. conf. knowl. discov. data min.. New York, NY, USA: ACM; 2016, p. 785–94. http://dx.doi.org/10.1145/2939672.2939785.

[18] Awad M, Khanna R. Support vector regression. In: Efficient learning machines. Berkeley, CA: A Press; 2015, http://dx.doi.org/10.1007/978-1-4302-5990-9_4.

[19] Scholkopf B, Sung Kah-Kay, Burges CJC, Girosi F, Niyogi P, Poggio T, Vapnik V. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Trans Signal Process 1997;45:2758–65. http://dx.doi.org/10.1109/78.650102.

[20] Kuhn M, Johnson K. Applied predictive modeling. New York, NY: Springer New York; 2013, http://dx.doi.org/10.1007/978-1-4614-6849-3.

[21] Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 2014;6:10. http://dx.doi.org/10.1186/1758-2946-6-10.