IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 5, 2021, accepted May 12, 2021, date of publication May 24, 2021, date of current version June 1, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3082926

A Bottom-up Method for Probabilistic Short-Term Load Forecasting Based on Medium Voltage Load Patterns

ZHENGBANG JIANG[1], HAO WU[1], (Member, IEEE), BINGQUAN ZHU[2], WEI GU[2], YINGWEI ZHU[3], YONGHUA SONG[1,4], (Fellow, IEEE), AND PING JU[1], (Member, IEEE)

[1]Department of Electrical Engineering, Zhejiang University, Hangzhou 310027, China
[2]Zhejiang Power Corporation, Hangzhou 310007, China
[3]Jinhua Power Corporation, Jinhua 321000, China
[4]Department of Electrical and Computer Engineering, University of Macau, Macau 999078, China

Corresponding author: Hao Wu (zjuwuhao@zju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U2066601, and in part by the Zhejiang Power Corporation under Grant 5211JH1900M3.

ABSTRACT Load forecasting has always been an essential part of power system planning and operation. In recent decades, the competition of the market and the requirements of renewable integration lead more attention to probabilistic load forecasting methods, which can reflect forecasting uncertainties through prediction intervals and hence benefit decision-making activities in system operation. Moreover, with the development of smart grid and power metering techniques, power companies have collected enormous load data about electricity customers and substations. The abundant load data allow us to utilize medium voltage measurement data to achieve better accuracy in high voltage transmission substation load forecasting. In this paper, a bottom-up probabilistic forecasting method is proposed for high voltage transmission substation short-term load forecasting, in which the probability distributions of medium voltage day-ahead load forecasting values are estimated and added up to form high voltage load predictions. Two bottom-up frameworks based on load patterns collected from medium voltage outgoing lines and substations are proposed respectively, in which mismatches between load data at different levels are estimated for correcting high voltage predictions. The comparison of predictions obtained by traditional and bottom-up methods demonstrates that the proposed method obtains high voltage load forecasting more accurately and give narrower prediction intervals.

INDEX TERMS Short-term load forecasting, bottom-up, medium voltage, load pattern, probabilistic load forecasting.

I. INTRODUCTION

Load forecasting is important for power system operation and control, specifically in generation dispatch, distributed generation integration, reactive power control [1]. Traditional transmission substation short-term load forecasting (STLF) methods obtain deterministic forecasting values based on potential regularity in historical load patterns.

In last decades, forecasting uncertainties neglected in deterministic load forecasting become more and more concerned, since the rapid developments of distributed generations, energy storage devices and electric vehicles [2]. Therefore, probabilistic load forecasting (PLF) [3], [4] has gained growing attention for its ability to quantify forecasting uncertainty by constructing prediction intervals (PIs) [5]. At the present, deterministic and probabilistic forecasting methods are still being improved for different STLF scenarios [6], but it is worth mentioning that present forecasting methods for high voltage (HV) transmission substation STLF are mainly based on HV load data [7], which limit the amount of information that can be used for HV load forecasting, because an HV transmission substation itself contains limited load measurement equipments.

Fortunately, with the development of the metering system, abundant medium voltage (MV) and low voltage load data have been collected by power companies [8], [9]. Along with

The associate editor coordinating the review of this manuscript and approving it for publication was Yunfeng Wen.

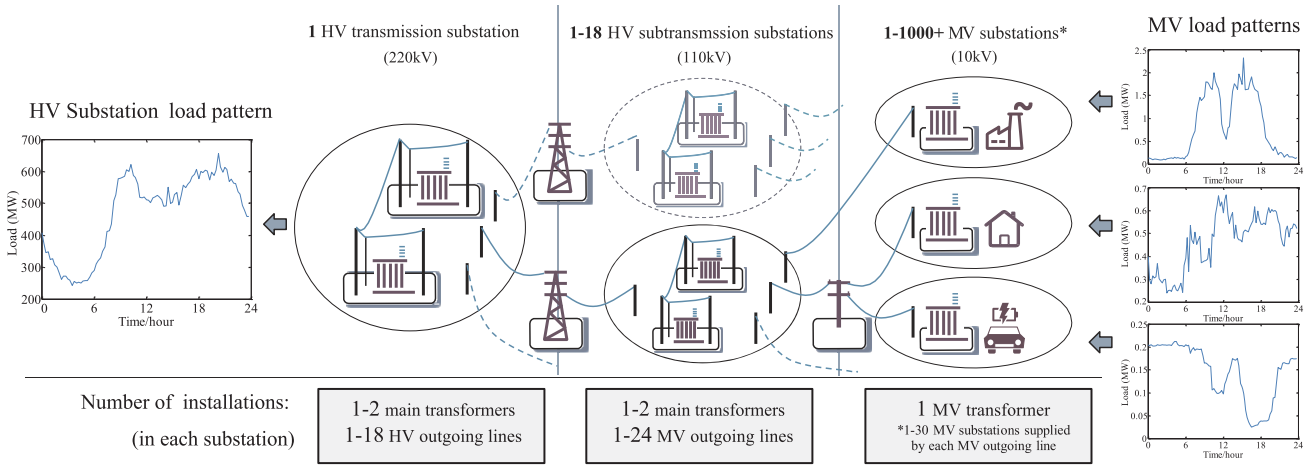


FIGURE 1. Hierarchical relationships between users and substations.

HV load data, a hierarchical structure can be constructed containing load patterns at diverse levels and hierarchical load forecasting (HLF) methods are proposed to improve the predictions of different levels [10], [11].

Combining the advantages of PLF with HLF, hierarchical probabilistic load forecasting (HPLF) is proposed to provide comprehensive probability distribution for load series at different levels of a hierarchy [1]. During qualifying match of Global Energy Forecasting Competition 2017 [12], a representative two-step method is proposed for HPLF [13], in which loads at different levels of the hierarchy are first predicted independently and then adjusted according to the inherent correlation between load data collected from different levels. In [14], the implementation of HPLF is based on load data collected from smart meters, and great attention is paid to the coherency between load predictions of bottom and upper levels. In [15], a weighted correction method based on constrained quantile regression is introduced to adjust predictive distributions at the bottom level and achieve coherent HPLFs.

As MV outgoing lines and MV substations usually supply groups of users with relatively similar load patterns [16], MV load patterns usually have clear regularities which can be extracted for HV transmission substation STLF and hence improve forecasting accuracy. Moreover, existing HPLF methods ignore mismatches between load data at different levels caused by network loss and measurement error.

Hence, with the train of thought similar to HPLF, a bottom-up load forecasting method, which includes two bottom-up frameworks based on measurement data collected from MV outgoing lines and MV substations respectively, is proposed for HV transmission substation day-ahead load forecasting. It is a bottom-up method, because MV load predictions at bottom level are first obtained individually by a PLF method based on Feed-forward Neural Network (FNN), and then aggregated for HV load predictions at upper level. As MV data are collected in many countries and regions, the proposed method has wide application prospects.

In general, focus on the utilizing of MV, HV load data and the aggregation relationship between them, this paper has the following contributions:

1) A bottom-up method is proposed to improve probabilistic load predictions by utilizing load data collected in both HV and MV levels, which can obtain HV transmission substation forecasts with better accuracy, narrower PIs and stable evaluations.
2) For different application scenarios, two effective bottom-up frameworks are proposed for probabilistic STLF based on load data collected from MV outgoing lines and MV substations respectively.
3) After probability distributions of MV load predictions being obtained, the normality of these distributions and the relevance between them have been analyzed for efficient and correct aggregation of prediction intervals.
4) Mismatches between MV and HV load data, which mainly reflect network loss and measurement error, are predicted by PLF method according to historical load data and then used in correcting HV predictions.

This paper is organized as follows: section II gives two bottom-up frameworks in terms of their data source. Section III reviews a probabilistic load forecasting method based on FNN for MV load predicting. In section IV, MV prediction results are aggregated to HV predictions. Section V presents case studies, and section VI concludes the paper.

II. BOTTOM-UP FRAMEWORKS FOR HV TRANSMISSION SUBSTATION SHORT TERM LOAD FORECASTING

A. DATA SOURCES FOR BOTTOM-UP FRAMEWORKS

Power system is a hierarchical system and Fig. 1 shows a conceptual topology connecting power customers to an HV transmission substation in China. This figure is based on both the information provided by the local power company and the system operation standard of China state grid. It shows that load measurement data are mainly collected in 220kV

HV transmission substations, 110kV HV subtransmission substations, 10kV MV substations and power customers.

For an HV transmission substation, the load data collected in it can be divided into two categories. The first category contains the load data collected from HV side of transformers. These data are commonly used for traditional HV transmission substation load forecasting. The second category contains the load data collected from substation outgoing lines. These data are equivalent to the load data collected in the HV subtransmission substations supplied by the outgoing lines, because one outgoing line in HV transmission substation usually supplies only one HV subtransmission substation.

As shown in Fig. 1, one HV transmission substation contains 1-18 outgoing lines and hence supplies 1-18 HV subtransmission substations. One HV subtransmission substation usually contains 1-24 MV outgoing lines and one MV outgoing line can supply 1-30 MV substations. As a result, one HV transmission substation could contain up to hundreds of MV outgoing lines and supply thousands of MV substations. But in practice, an HV transmission substation usually supplies no more than one hundred MV outgoing lines and supplies about one thousand MV substations. Each MV outgoing line or MV substation supplies a relatively small area, where users are prone to share similar power consumption behaviors. Therefore, MV load patterns usually have clear regularities which can be extracted for HV transmission substation load forecasting.

Though MV load data are collected from both MV outgoing lines and MV substations, the former are more accessible while the latter are a bit harder to be accessed, because of enormous involved load data and also because of barriers between different databases. Sometimes, there is even no available MV substation load data in some regions.

Cases in this paper are based on load data (96 points per day) collected from two HV transmission substations and their subordinate MV substations or MV outgoing lines located in East China. The first substation, named substation A in the following, only realizes the full collection of load data from 56 MV outgoing lines, excluding MV substations, so only load data of MV outgoing lines are available for STLF. The second substation, named substation B in the following, realizes the full collection of load data from 996 MV substations, so load data of MV substations can be used for STLF.

Therefore, for different application scenarios, two bottom-up frameworks based on MV outgoing lines and MV substations load patterns are introduced respectively in the section II.B and section II.C. Besides, for simplicity, HV transmission substation will be called HV substation later in this paper.

B. BOTTOM-UP LOAD FORECASTING FRAMEWORK BASED ON LOAD PATTERNS OF MV OUTGOING LINES

An HV substation usually contains no more than one hundred MV outgoing lines, so the loads of all MV outgoing lines can be predicted based on historical load patterns separately within acceptable computation time, and then aggregated to HV load prediction.

Because of the network loss and measurement error, load patterns collected independently in MV level may not add up to HV substation load pattern exactly. These mismatches are called add-up error in our research, which will be discussed in section IV.A.

Given an HV substation with K MV outgoing lines, its day-ahead load prediction at time t can be written as

$$\widehat{y}_{hv,t} = \sum_{l=1}^{K} \widehat{y}_{l,t} + \varepsilon_{a,t}, \quad t = 1, \ldots, 96 \tag{1}$$

where $\widehat{y}_{hv,t}$ is the day-ahead load prediction of HV substation, $\widehat{y}_{l,t}$ is the day-ahead load prediction of the lth MV outgoing line and $\varepsilon_{a,t}$ is the add-up error.

C. BOTTOM-UP LOAD FORECASTING FRAMEWORK BASED ON LOAD PATTERNS OF MV SUBSTATIONS

Because an HV substation can supply about one thousand MV substations, it is rather time-consuming to predict loads of every MV substations. Therefore, MV load patterns are first clustered to extract their common regularities. Here, classical K-means algorithm [11], [17] is adopted, whose inputs are the normalized 96-dimension load patterns of MV substations. Before clustering, each load pattern $Y_{original} = [y_{original,1} \cdots y_{original,96}]$ in the data set are normalized by max-min method, i.e.,

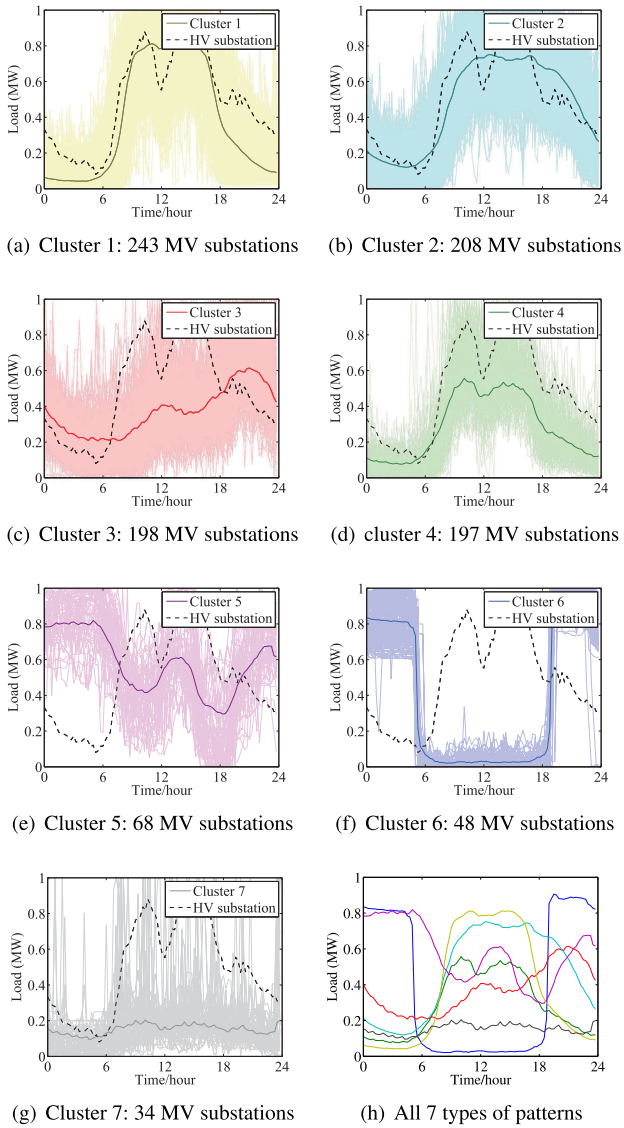
$$y_{normalized,t} = \frac{y_{original,t} - min(Y_{original})}{max(Y_{original}) - min(Y_{original})}, \tag{2}$$

where $y_{original,t}$ and $y_{normalized,t}$ are the tth elements of the original load pattern and the normalized load pattern respectively. The optimal number of clusters is chosen according to popular Davies-Bouldin index (DBI) [18], i.e., the optimal number of clusters corresponds the minimal DBI.

Because the normalized MV load patterns in each cluster share similar regularity, they can be predicted together to reduce prediction time consumption. To simplify total load handling, the unnormalized MV load patterns in each cluster are added up and predicted. Then, the HV load prediction at time t can be obtained similarly to that of (1), where K is now the number of clusters, $\widehat{y}_{l,t}$ is the load prediction of the lth cluster, and $\varepsilon_{a,t}$ is the add-up error.

D. CLUSTERING RESULT OF MV SUBSTATIONS

Clustering results of 996 MV substations supplied by HV substation B is taken as an example here. According to DBI, the optimal number of clusters is 7 and Fig. 2 shows the centroids of 7 types of MV substations. The first four clusters contain more than 80% of the MV substations supplied by HV substation B. These MV substations mainly supply residential, commercial and utility loads with day-peak and



(a) Cluster 1: 243 MV substations (b) Cluster 2: 208 MV substations

(c) Cluster 3: 198 MV substations (d) cluster 4: 197 MV substations

(e) Cluster 5: 68 MV substations (f) Cluster 6: 48 MV substations

(g) Cluster 7: 34 MV substations (h) All 7 types of patterns

FIGURE 2. 7 types of MV substation load patterns in HV substation B.

evening-peak load patterns. Meanwhile, the last three clusters contain MV substations mainly supply agricultural, street lamp and industrial loads with double-peak and night-peak load patterns. It can be seen from the figure that all 7 centroids capture the overall MV substation load patterns of corresponding clusters, and that 7 centroids are different from each other. All these features above indicated that the clustering results are proper.

The added up load patterns of each cluster, which is used in MV load prediction, are shown in Fig. 3. It can be seen from the figure that the sums of load patterns in cluster 1, 2, 4 and 6 show obvious changes with time, especially at 6:00, 12:00 and 18:00. These regularities imply that above patterns are easier to be predicted. The sums of load patterns in cluster 3, 5 and 7 slope gently, but contain more small fluctuations, which will bring uncertainties in their load predictions.

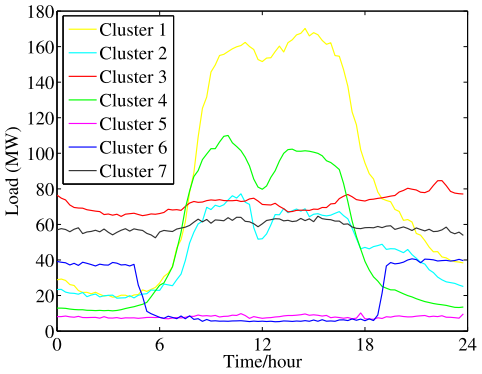


FIGURE 3. Sums of the MV substation load patterns in 7 clusters.

E. MV PROBABILISTIC LOAD FORECASTING

NNs have been used for MV load forecasting with their approximation capability of nonlinear mapping [19], [20]. To cope with uncertainties in load forecasting, researchers propose NN-based PLF methods [21], where a basic load forecasting model is applied to historical data set repeatedly to obtain multiple load predictions. With the probability distribution of these predictions, model and data noise variances of the predictions are estimated and integrated to form the prediction interval (PI).

Focus on the bottom-up PLF frameworks themselves, the wide accepted feed-forward neural network (FNN) is adopted in this paper as the basic forecasting model [22]. The FNN-based PLF method will be introduced in section III.A. However, to aggregate MV predictions more effectively, not PIs but mean values and variances of MV load predictions are obtained at first. After that, they are aggregated and hence form PI at HV level. Therefore, the rest of section III mainly discusses the estimation of mean values and variances, and later, the formation of PIs will be discussed in section IV.D.

III. FNN-BASED PROBABILISTIC MV LOAD FORECASTING

A. FEED-FORWARD NEURAL NETWORK

FNN is a typical multi-layer network with one input layer, one output layer and multiple hidden layers. For simplicity, the FNN adopted in this paper has only one hidden layer, and its structure is shown in the Fig. 4. As usual, full connections exist between these layers and no connection exists in the same layer.

Given N arbitrary training set

$$\boldsymbol{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}, \tag{3}$$

where $\boldsymbol{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]$ is the input and y_i is the target of the train set. An FNN with n_r hidden nodes and activation function $g(\cdot)$ can be mathematically modeled as

$$f(\boldsymbol{x}_i; \boldsymbol{w}, b, \beta) = \sum_{j=1}^{n_r} \beta_j g(\boldsymbol{w}_j \cdot \boldsymbol{x}_i + b_j), \quad i = 1, \ldots, N, \tag{4}$$

where $\boldsymbol{w}_j = [w_{j1}, w_{j2}, \ldots, w_{jd}]^T$ is the weight vector connecting the jth hidden node and the input nodes, and β_j is

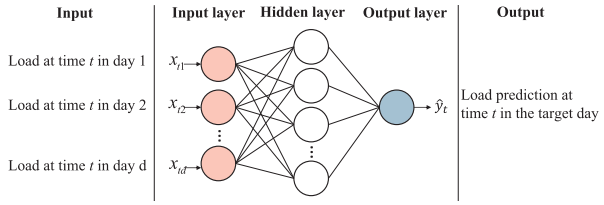


FIGURE 4. Conceptual structure of feed-forward neural network with single hidden layer.

the weight vector connecting the jth hidden node and the output node, and b_j is the threshold of the jth hidden node, and $f(\boldsymbol{x}_i; \boldsymbol{w}, b, \beta)$ is the output of FNN. Activation function $g(\cdot)$ is the most popular sigmoid function and the FNN is trained by classic back propagation method.

As it shown in Fig. 4, when a trained FNN is adopted to predict day-ahead load value at time t, the historical load values in the last d days at the time t are usually chosen as the input x_t of FNN, and d is determined in section V.B.

Given a N_{test} arbitrary test set $\{x_t, y_t\}_{t=1}^{N_{test}}$, where $\boldsymbol{x}_t = [x_{t1}, x_{t2}, \ldots, x_{td}]$ is the input, y_t is the real load value, N_{test} is the number of test sample, and for day-ahead load forecasting in our case, $N_{test} = 96$. With specific input $\boldsymbol{x}_t$, the prediction of real load value y_t can be denoted by $\widehat{y}_t = f(\boldsymbol{x}_t; \boldsymbol{w}, b, \beta)$. Then prediction errors of FNN are analyzed and quantified by variances in following subsections.

B. COMPONENTS OF THE PREDICTION ERROR

At time t, the real load value y_t can be regarded as the sum of the prediction value $\widehat{y}_t$ and a prediction error ε_t, that is

$$y_t = f(\boldsymbol{x}_t; \boldsymbol{w}, b, \beta) + \varepsilon_t = \widehat{y}_t + \varepsilon_t, \tag{5}$$

because the prediction error ε_t is mainly caused by model misspecification and data noise [5]. Therefore, ε_t can be seen as containing two components and written as

$$\varepsilon_t = \varepsilon_{m,t} + \varepsilon_{d,t}, \tag{6}$$

where $\varepsilon_{m,t}$ is the error caused by the model misspecification and $\varepsilon_{d,t}$ is the error caused by data noise.

Assuming two error components in (6) are independent and both subject to Gaussian distribution, the variance σ_t^2 of the prediction errors ε_t can be written as

$$\sigma_t^2 = \sigma_{m,t}^2 + \sigma_{d,t}^2, \tag{7}$$

where $\sigma_{m,t}^2$ is the model misspecification variance corresponding to $\varepsilon_{m,t}$, and $\sigma_{d,t}^2$ is the data noise variance corresponding to $\varepsilon_{d,t}$.

C. ESTIMATION OF MODEL MISSPECIFICATION VARIANCE

In (6), $\varepsilon_{m,t}$ is mainly caused by the randomly generated initial parameters of FNN and the so caused local minima in the training process. In addition, even if the global minimum can be reached, the misspecification of model structure also introduces non-negligible uncertainty in predictions [5].

Therefore, to estimate model misspecification variance $\sigma_{m,t}^2$, B FNNs with random weights $\boldsymbol{w}$, β and threshold b are trained with the same data set to predict day-ahead load at time t. If $\widehat{y}_{h,t}$ is the prediction given by the hth FNN, the mean value of prediction given by all B FNNs can be expressed as

$$\widehat{y}_t = \frac{1}{B}\sum_{h=1}^{B}\widehat{y}_{h,t}. \tag{8}$$

Later, $\widehat{y}_t$ will be taken as the deterministic prediction of y_t.

Based on $\widehat{y}_t$, the model misspecification variance $\sigma_{m,t}^2$ can be estimated from the outputs of the trained B FNNs as

$$\sigma_{m,t}^2 = \frac{1}{B-1}\sum_{h=1}^{B}(\widehat{y}_{h,t} - \widehat{y}_t)^2. \tag{9}$$

D. ESTIMATION OF DATA NOISE VARIANCE

It is difficult to quantify the data noise variance, considering the heteroscedasticity characteristic in measurement data noise. From [23], the historical data noise variance conditioned on the training input set variables x_i in (3) can be estimated as

$$\sigma_{d,i}^2 = E[(y_i - \widehat{y}_i)^2 | x_i]. \tag{10}$$

Specifically, based on B trained FNNs in section III.C, if $\widehat{y}_{h,i}$ is the estimation of y_i given by the hth FNN, historical data noise variance $\sigma_{d,i}^2$ can be calculated by

$$\sigma_{d,i}^2 = \frac{1}{B-1}\sum_{h=1}^{B}(y_i - \widehat{y}_{h,i})^2. \tag{11}$$

To obtain the day-ahead data noise variance at time t, a new training set $\boldsymbol{D}_{\sigma_d}$ is formed via replacing the targets y_i with $\sigma_{d,i}^2$ in (3), and given as

$$\boldsymbol{D}_{\sigma_d} = \{(\boldsymbol{x}_i, \sigma_{d,i}^2)\}_{i=1}^{N}. \tag{12}$$

Then, with input x_t and according to (4), data noise variance $\sigma_{d,t}^2$ is predicted by FNN trained with (12).

So far, MV probabilistic load prediction at any time t, which includes deterministic prediction $\widehat{y}_t$ and prediction variance σ_t^2, can be obtained. In the section IV, MV probabilistic load predictions are aggregate to HV load prediction.

IV. HV SUBSTATION PROBABILISTIC LOAD FORECASTING BASED ON BOTTOM-UP FRAMEWORKS

To accurately aggregate MV probabilistic load predictions to the HV substation probabilistic load prediction, the following three aspects should be addressed:

(1)Add-up errors mentioned in section II.B should be considered.

(2)The probability distributions of MV load predictions should be modeled.

(3)The relevances between MV load prediction distributions should be determined.

A. ADD-UP ERRORS

Add-up error ε_a mentioned in section II.B mainly originates from distribution network losses and power measurement errors. The former relate to MV load patterns, which change with time regularly, and hence can be predicted. The later usually introduce small irregular variations in the pattern of ε_a, which are hard to predict and will bring uncertainties to the predictions.

The predictable network losses are accounted as major part of add-up error ε_a, and the power measurement errors can be seen as data noise. Therefore, the FNN-based PLF method introduced in section III is utilized here to obtain the probabilistic prediction of ε_a based on its historical values. Specifically, similar to (3), an add-up error training set $\boldsymbol{D}_{\varepsilon_a}$ with N arbitrary is given as

$$\boldsymbol{D}_{\varepsilon_a} = \{\boldsymbol{\varepsilon}_{a,i}, y_{\varepsilon_a,i}\}_{i=1}^{N}, \tag{13}$$

where the ith input $\boldsymbol{\varepsilon}_{a,i} = [\varepsilon_{a,i1}, \varepsilon_{a,i2}, \ldots, \varepsilon_{a,id}]$ is historical add-up error and target $y_{\varepsilon_a,i}$ is the corresponding day-ahead add-up error. Then deterministic prediction $\widehat{y}_{\varepsilon_a,t}$ and prediction variance $\sigma^2_{\varepsilon_a,t}$ of day-ahead add-up error at time t can be obtained according to (4)-(12).

Taking the HV substation A with 56 MV outgoing lines as an example, daily add-up errors between HV substation load pattern and the sum of outgoing line load patterns in 1095 days from April 2016 to April 2018 are shown in Fig. 5. To measure the accuracy of add-up error prediction based on FNN, mean absolute percentage error (MAPE) is used as an index to evaluate the prediction errors, which is defined by

$$MAPE = \frac{100\%}{N_{test}} \sum_{t=1}^{N_{test}} \left| \frac{y_t - \widehat{y_t}}{y_t} \right|, \tag{14}$$

where y_t and $\widehat{y_t}$ are real value and prediction value. The minimum and average MAPE of add-up error predictions are 6.08% and 11.61%. For comparison, the MAPEs of day-ahead load predictions obtained by FNN is about 5%-10% [24]. The MAPEs of add-up error predictions are slightly larger but still acceptable, which imply it is feasible to predict add-up error according to its historical values.

B. PROBABILITY DISTRIBUTION OF MV LOAD PREDICTIONS

The probability distributions of MV load predictions are hard to be investigated theoretically, therefore they are studied and modeled statistically here. More specifically, the MV load pattern are predicted by FNN for many times, and then the prediction distribution are formed empirically through the obtained predictions. As examples, two outgoing lines with typical and untypical load patterns are studied in the following.

Load patterns of these two outgoing lines on August 27 2018 are predicted 1000 times by FNN to form the empirical distributions of predictions, and Fig. 6(a) shows prediction distributions of the first outgoing line. The mean value of

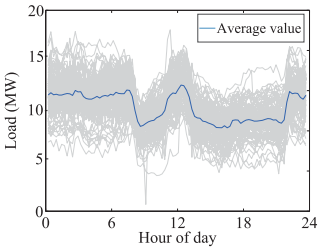
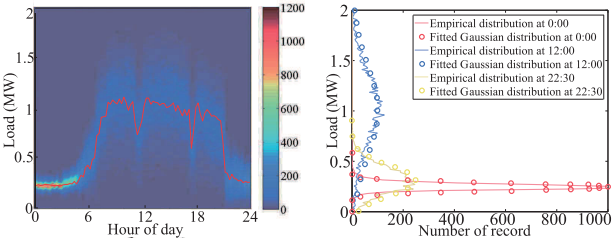
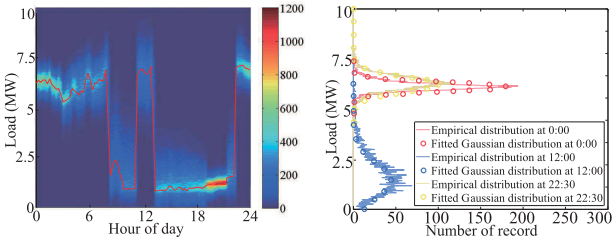


FIGURE 5. Historical daily add-up errors in 2016-2018.



(a) General distribution (b) Distributions at specific times

FIGURE 6. Probability distributions of predictions in outgoing line with typical load pattern.



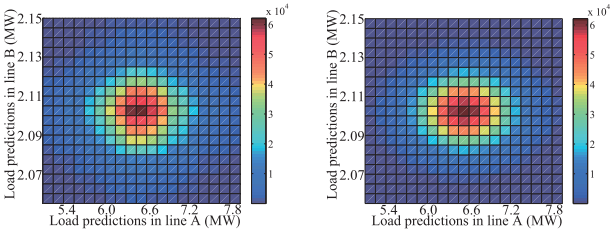
(a) General distribution (b) Distributions at specific times

FIGURE 7. Probability distributions of predictions in outgoing line with untypical load pattern.

these predictions are also plotted with red line in this figure. It can be seen that most of the predictions are concentrated around their mean value, which accords with the characteristic of Gaussian distribution.

Therefore, the corresponding Gaussian distributions are calculated according to mean value and variance of above 1000 predictions to fit the empirical prediction distribution. Fig. 6(b) shows part of the distributions in Fig. 6(a) at three specific times. There are 3 pairs of curves in Fig. 6(b) corresponding to 3 specific times at 0:00, 12:00, and 22:30 respectively. The dotted curves show the fitted Gaussian distributions calculated according to corresponding mean value and variance of predictions, while the other curves in the pairs show the empirical distributions of predictions. In Fig. 6(b), each pair of curves are broadly coincide with each other, which imply the predictions at all three times are subject to Gaussian distribution. Similarly, Fig. 7 shows that the predictions of untypical load pattern are also subject to Gaussian distribution.

In addition to these examples above, prediction distributions of all outgoing lines in HV substation A are tested,



(a) Empirical joint distribution (b) Calculated joint distribution

FIGURE 8. Empirical joint distribution and calculated joint distribution. Figure(a) shows the empirical joint distribution formed through repeated experiments, and figure (b) shows the joint distribution calculated according to joint distribution function.

and the MAPE between all the empirical and fitted Gaussian distributions is 1.34%. The minor MAPE further verifies the predictions obtained by FNN are generally subject to Gaussian distribution.

C. RELEVANCE TEST OF MV LOAD PREDICTIONS

Independence of MV load predictions is tested as the first step of relevance test. Mutual independence of two Gaussian variables can be tested according to the independence theorem [25]: for two mutually independent Gaussian variables y_1 and y_2, their joint distribution function is equal to the product of their marginal distribution functions, and can be expressed as

$$f(y_1, y_2) = \frac{1}{(2\pi\sigma_1\sigma_2)} e^{-\frac{(y_1-\widehat{y}_1)^2}{2\sigma_1^2} - \frac{y_2-\widehat{y}_2)^2}{2\sigma_2^2}}, \tag{15}$$

where $\widehat{y}_1, \widehat{y}_2, \sigma_1^2$ and σ_2^2 are expectations and variances of the two Gaussian variables. According to above theorem, if two Gaussian variables are mutually independent, the joint distribution of them can be calculated through their mean values and variances. Hence, if the calculated joint distribution is same as the empirical joint distribution obtained according to repeated experiments, above two Gaussian variables can be verified to be mutually independent.

For example, mutually independence of load predictions of two outgoing line A and line B, at 12:00 on August 27 2018, are tested as follows. First, 100000 times load predictions of two outgoing lines are used to form the empirical joint distribution, and a sample matrix of this distribution is shown in Fig. 8(a).

Then, based on load predictions in above outgoing lines, the joint distribution function is calculated according to (15), where the load predictions in line A and line B is taken as y_1 and y_2, $\widehat{y}_1$ and $\widehat{y}_2$ are 0.2116 and 6.1983, σ_1 and σ_2 are 0.6338 and 0.5062. The sample matrix formed through calculated joint distribution function is shown in Fig. 8(b). The MAPE between two joint distribution sample matrices is 3.03%, which implies the mutual independence of load predictions of outgoing line A and B.

Furthermore, load predictions of all 56 outgoing lines are paired and tested. The average MAPE between the empirical and calculated joint distributions of these pairs is 3.16%. Hence, predictions in each two outgoing lines can be regarded as mutual independent.

D. AGGREGATION OF MV PROBABILISTIC LOAD PREDICTIONS

Because the load predictions of MV outgoing lines and substations are independent and subject to Gaussian distribution, MV probabilistic load predictions can be aggregated by simply adding up deterministic predictions $\widehat{y}_t$ and prediction variances σ_t^2 [26]. More specifically, if $\widehat{y}_{l,t}$ is the day-ahead deterministic prediction of the lth MV outgoing line or MV substation cluster, day-ahead deterministic load prediction of HV substation can be expressed as

$$\widehat{y}_{hv,t} = \sum_{l=1}^{K} \widehat{y}_{l,t} + \widehat{y}_{\varepsilon_a,t}, \quad t = 1, \ldots, 96. \tag{16}$$

Meanwhile, if $\sigma_{l,t}^2$ is the variance of the load predictions of the lth MV outgoing line or MV substation cluster, the variance of HV substation day-ahead load predictions can be expressed as

$$\sigma_{hv,t}^2 = \sum_{l=1}^{K} \sigma_{l,t}^2 + \sigma_{\varepsilon_a,t}^2, \quad t = 1, \ldots, 96. \tag{17}$$

With deterministic prediction $\widehat{y}_{hv,t}$ and prediction variance $\sigma_{hv,t}^2$, the PI of the target HV substation with $100(1-\alpha)\%$ nominal confidence (PINC) at each time t is a stochastic interval I_t^α expressed as

$$I_t^\alpha = [L_t^\alpha, U_t^\alpha], \tag{18}$$

where the coverage rate $P(y_t \in I_t^\alpha) = 100(1-\alpha)\%$. The lower bound L_t^α and the upper bound U_t^α can be obtained by

$$L_t^\alpha = \widehat{y}_{hv,t} - z_{1-\alpha/2}\sqrt{\sigma_{hv,t}^2}, \tag{19}$$

$$U_t^\alpha = \widehat{y}_{hv,t} + z_{1-\alpha/2}\sqrt{\sigma_{hv,t}^2}, \tag{20}$$

where $z_{1-\alpha/2}$ is the critical value of the standard Gaussian distribution [5], which depends on the desired confidence level $100(1-\alpha)\%$.

E. PROCEDURE OF PLF WITH BOTTOM-UP FRAMEWORKS

To obtain HV probabilistic load prediction at time t, the procedure of PLF with bottom-up frameworks can be summarized as

1) Bottom-up framework should be chosen according to available MV load data. If MV substation load data are available, go to step (2). If there are only outgoing line load data, go to step (3);
2) Based on section II.C, MV substation load patterns are clustered with K-means algorithm, and the sum of load patterns in each cluster is calculated;
3) Based on PLF method introduced in section III, deterministic MV load predictions $\widehat{y}_{l,t}$ and corresponding prediction variances $\sigma_{l,t}^2$ are obtained according to (8)-(11);

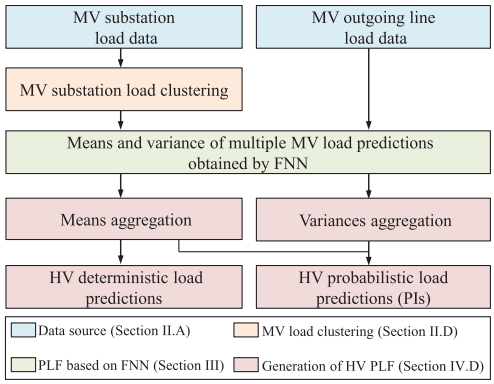


FIGURE 9. MAE of different parameters in FNN.

4) Based on the method proposed in section IV.A, the deterministic prediction of add-up error $\widehat{y}_{\varepsilon_a,t}$ and corresponding prediction variance $\sigma^2_{\varepsilon_a,t}$ are obtained;
5) The deterministic predictions and prediction variances of MV loads and add-up error are aggregated to calculate the deterministic load prediction $\widehat{y}_{hv,t}$ and prediction variance $\sigma^2_{hv,t}$ of HV substation according to (16) and (17), and then PIs are obtained according to (18)-(20).

The flowchart of the proposed method is shown in Fig. 9.

V. CASE STUDIES

A. CASES INTRODUCTION

Two cases are organized based on load data collected in HV transmission substations A and B from January 2016 to December 2018, to test bottom-up frameworks based on MV outgoing line and MV substation load data respectively. Predictions obtained by FNN-based PLF with and without bottom-up frameworks are analyzed and compared to show the effectiveness of bottom-up frameworks. The load patterns of MV outgoing lines and MV substations are provided by local power dispatch center and marketing department respectively.

The first case is organized based on load patterns of MV outgoing lines supplied by HV substation A, to test the bottom-up framework based on load patterns of MV outgoing lines. The average load of these outgoing lines is 1.216 MW.

The second case is organized based on load patterns of MV substations supplied by HV substation B, to test the bottom-up framework based on load patterns of MV substations. The average load of these MV substations is 0.354 MW.

B. EVALUATION INDEXES

Besides MAPE defined as (14), mean absolute error (MAE), root mean square error (RMSE), and R2 score are also adopted to evaluate deterministic load predictions obtained by different methods, which are defined by

$$MAE = \frac{1}{N_{test}} \sum_{t=1}^{N_{test}} |y_t - \widehat{y}_t|, \tag{21}$$

$$RMSE = \frac{1}{N_{test}} \sqrt{\sum_{t=1}^{N_{test}} (y_t - \widehat{y}_t)^2}, \tag{22}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{N_{test}} (y_t - \widehat{y}_t)^2}{\sum_{t=1}^{N_{test}} (y_t - \bar{y}_t)}, \tag{23}$$

where $\bar{y}_t = (\sum_{t=1}^{N_{test}} y_t)/N_{test}$.

For probabilistic load predictions, PI coverage probability (PICP), average coverage error (ACE), prediction interval normalized average width (PINAW) and quantile score (QS) are adopted to evaluate the quality of PIs. PICP is used to measure the proportion of real values lying within the bounds of PIs, which is defined by

$$PICP = \frac{1}{N_{test}} \sum_{t=1}^{N_{test}} c_t, \tag{24}$$

where c_t is the indicator of PICP and is defined by

$$c_t = \begin{cases} 1, & y_t \in I_t^\alpha \\ 0, & y_t \notin I_t^\alpha. \end{cases} \tag{25}$$

Meanwhile, the PICP should be close to its corresponding PINC. So another evaluating index, ACE [1], is defined by

$$ACE = PICP - PINC. \tag{26}$$

Smaller absolute value of ACE infer implies stable and effective PIs.

In practice, there may be PIs with high PICP but large width. Therefore, PINAW [27] is adopted to measure the width of PIs, which is defined by

$$PINAW = \frac{1}{N_{test} R} \sum_{t=1}^{N_{test}} (U_t^\alpha - L_t^\alpha), \tag{27}$$

where R is the difference between the maximum and the minimum real load values in test case.

Moreover, quantile score is adopted to evaluate the obtained PIs. At time t, the quantile score [28] of a quantile forecast q_t at the level $\alpha \in (0, 1)$ can be defined by

$$QS_t^\alpha(q_t, y_t) = 2(\mathbb{I}\{y_t < q_t\} - \alpha)(q_t - y_t). \tag{28}$$

In this paper, the upper bound U_t^α and lower bound L_t^α of PI with $100(1-\alpha)\%$ PINC is equal to q_t at level α and level $1-\alpha$ respectively. Therefore, the QS of the PI with $100(1-\alpha)$ PINC can defined by

$$QS_t^\alpha = \begin{cases} -2\alpha(U_t^\alpha - y_t) - 2\alpha(L_t^\alpha - y_t), & y_t > U_t^\alpha \\ 2(1-\alpha)(U_t^\alpha - y_t) - 2\alpha(L_t^\alpha - y_t), & y_t \in I_t^\alpha \\ 2(1-\alpha)(U_t^\alpha + L_t^\alpha - 2y_t), & y_t < L_t^\alpha \end{cases} \tag{29}$$

Similar to [29], in order to evaluate the overall performance of a method in providing reliable PIs with different PINC, QS_t^α of PIs with PINC from 1% to 99% is obtained. Then, for a test case with N_{test} samples, the overall quantile score QS can be obtained and written as

$$QS = \frac{1}{N_{test}} \sum_{t=1}^{N_{test}} \sum_{i=1}^{99} QS_t^{\alpha_i}. \tag{30}$$

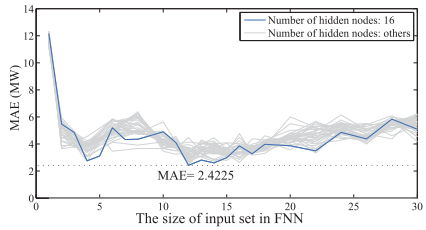


FIGURE 10. MAE of different parameters in FNN.

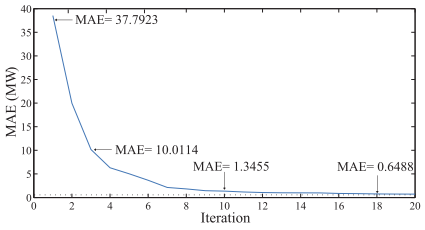
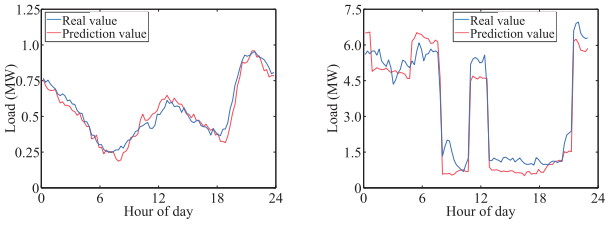


FIGURE 11. MAE of different iterations in training process of FNN.



(a) Typical (b) Untypical

FIGURE 12. Day-ahead load patterns of two outgoing lines predicted by FNN.

C. DETERMINATION OF PARAMETERS IN FNN

The number of hidden nodes n_h and the size of input set d are two important parameters in FNN. For normal weekdays and special dates such as weekends, two methods are used to determine the input of the FNN, and an exhaustive search is carried out to optimize the corresponding n_h and d. The validation set used to optimize the parameters is the last two weeks before the target day.

For weekdays, historical load data in previous d days are chosen as the input of the FNN. For special days like weekends, load data in similar days are chosen as the input of the FNN. Pearson coefficients [30] between the load patterns of weekends in the validation set and their adjacent days are calculated to measure the correlations among them and lead the selection of similar days. Then the load data in the most similar d days are selected to form the input set.

In the exhaustive search, the values of n_h and d are set to 1 to 30, and 30×30 combinations of these values are applied to FNNs. For weekday or weekend load forecasting, MV load predictions of weekdays or weekends in the validation set are given by these FNNs and the average MAEs of predictions obtained by FNN with specific n_h and d are assessed. The optimal combination of n_h and d corresponds to the minimal MAE. Take parameter optimization in a weekday in summer as an example, as seen in Fig. 10, when $n_h = 16$ and $d = 12$, MAE of the accessed predictions reaches its minimum 2.4225.

For the training of FNN, Fig. 11 shows the average training process of FNN in our case, where the MAE decreases 73.51% in first 5 iterations, and on average, MAE drops to minimum and remains stable within 20 iterations.

D. HV LOAD FORECASTING BASED ON OUTGOING LINE LOAD PATTERNS

Bottom-up framework based on load patterns of MV outgoing lines is tested in this section. Load pattern of HV substation A in August 27 2018 is predicted and analyzed here, because substation load patterns fluctuate greatly due to the climate variations in summer, which can better reflect performances of the proposed method.

To verify the validity of FNN, deterministic load predictions $\widehat{y}_l$ of two MV outgoing lines with typical and untypical load pattern are shown in Fig. 12. For load prediction of the outgoing line with typical load pattern, the MAPE is 5.06% and the MAE is 0.0268. For the outgoing line with untypical load pattern, the MAPE is 5.41% and the MAE is 0.287. The minor prediction errors show that the FNN is effective in MV load forecasting.

TABLE 1. Performance comparison of proposed bottom-up method based on load patterns of MV outgoing lines and traditional method.

Method	Deterministic predictions				PIs with 90% PINC			QS
	MAPE	MAE	R2	RMSE	ACE	PICP	PINAW	
Bottom-up	**2.43%**	**2.4495**	**0.8442**	**0.3144**	**-0.42%**	**89.58%**	**15.1157**	162.00
traditional	3.96%	3.9235	0.5502	0.5533	7.90%	97.90%	23.1059	245.51

TABLE 2. Performances of the proposed method in one-week rolling predicting test.

	Aug. 27	Aug. 28	Aug. 29	Aug. 30	Aug. 31	Sept. 1	Sept. 2
MAPE	**2.47%**	**2.24%**	**2.25%**	**2.42%**	2.76%	2.87%	3.51%
ACE	**-0.42%**	**-1.46%**	**2.70%**	**-1.46%**	5.83%	-13.96%	-3.55%

TABLE 1 compares the performances of bottom-up method and traditional PLF method, where the later is actually FNN-based PLF method only based on HV substation load patterns. For deterministic predictions, the proposed method obtained more accurate predictions with smaller MAPE, MAE, RMSE and higher R2 score. For PIs with 90% PINC, the proposed method obtains more reliable PIs with PICP closer to 90% and ACE closer to zero. According to PINAW, it should also be noted that PIs obtained by bottom-up method are significantly narrowed by 34.58%. Moreover, the QS of the proposed method is higher than that of the traditional method.

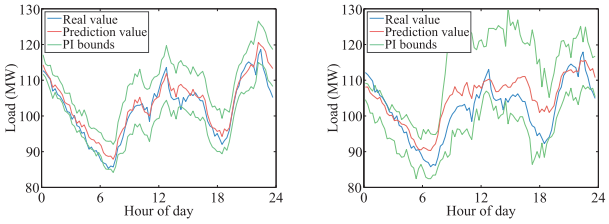
Fig. 13 shows the predictions of HV substation load pattern obtained by bottom-up PLF method and traditional PLF method. Compared with the traditional method, deterministic load prediction $\widehat{y}_{hv}$ given by bottom-up method conforms to the real load pattern with only slight fluctuations. Moreover, the bottom-up PLF method obtains obviously narrower PIs.

TABLE 3. Performance comparison of the proposed method and other popular probabilistic STLF methods in probabilistic predictions for cases in different seasons.

Season	Method	99% PINC			90% PINC			50% PINC			10% PINC			QS
		ACE	PICP	PINAW	ACE	PICP	PINAW	ACE	PICP	PINAW	ACE	PICP	PINAW	
Summer	Bottom-up	**1.00%**	**100%**	**0.4458**	**-0.42%**	**89.58%**	**0.2846**	**6.25%**	**56.25%**	**0.1415**	**1.62%**	**11.62%**	**0.0264**	**162.00**
	persistence	**1.00%**	**100%**	1.2235	2.71%	92.71%	0.7813	32.29%	82.29%	0.3417	17.08%	27.08%	0.0504	254.51
	ESM	**1.00%**	**100%**	1.6577	-7.71%	82.29%	0.3909	-35.02%	14.98%	0.2659	-5.83%	4.27%	0.0186	448.41
	BELM	-1.08%	97.92%	0.6055	**-0.42%**	**89.58%**	0.3867	-25.00%	25.00%	0.1594	-3.75%	6.25%	0.0277	207.63
	RBM	-6.29%	92.71%	0.4762	-1.46%	88.54%	0.4720	8.33%	58.33%	0.1952	-2.71%	7.29%	0.0363	182.14
	LQREC	**1.00%**	**100%**	0.6967	3.75%	93.75%	0.4449	-35.42%	14.58%	0.1824	-1.67%	8.33%	0.0340	181.34
Autumn	Bottom-up	**1.00%**	**100%**	**0.5491**	1.64%	91.64%	**0.3507**	-2.92%	47.08%	**0.1441**	-3.75%	6.25%	**0.0268**	**181.49**
	persistence	-1.08%	97.92%	1.4713	-2.71%	92.71%	0.9395	-15.63%	34.37%	0.3160	12.92%	22.92%	0.0589	263.27
	ESM	**1.00%**	**100%**	2.0890	-2.50%	87.50%	0.4853	-8.33%	41.67%	0.1525	5.62%	15.62%	0.0283	248.02
	BELM	-1.08%	97.92%	0.5822	**-1.46%**	**88.54%**	0.3718	-17.71%	32.29%	0.1952	-4.79%	5.21%	0.0364	206.57
	RBM	-8.37%	91.63%	0.7098	-24.38%	65.61%	0.4532	-20.83%	29.17%	0.1856	-6.88%	3.12%	0.0346	217.96
	LQREC	-3.17%	96.82%	1.0622	-16.04%	73.96%	0.6783	**1.04%**	**51.04%**	0.2781	**-0.42%**	**9.58%**	0.0518	219.82
Winter	Bottom-up	-1.08%	97.92%	**0.4431**	0.63%	90.63%	**0.2830**	**-2.08%**	**47.92%**	**0.1207**	**-0.42%**	**9.58%**	**0.0225**	**102.21**
	persistence	-6.29%	92.71%	1.4680	4.58%	85.42%	0.9374	-15.63%	35.37%	0.1807	-4.79%	5.21%	0.0337	206.57
	ESM	**1.00%**	**100%**	2.1482	**-0.42%**	**89.58%**	0.4791	3.13%	53.13%	0.1410	9.79%	19.79%	0.0196	147.12
	BELM	-3.16%	95.83%	0.9571	-2.50%	87.50%	0.6112	-12.50%	37.50%	0.1734	-0.63%	9.37%	0.0323	123.61
	RBM	-2.12%	96.88%	1.0867	-6.67%	83.33%	0.6335	-15.63%	34.37%	0.2791	-4.79%	5.21%	0.0456	126.62
	LQREC	**1.00%**	**100%**	1.1360	4.79%	94.79%	0.7254	8.33%	58.33%	0.2975	4.58%	14.58%	0.0554	119.20

TABLE 4. Performance comparison of proposed method and other methods in deterministic predictions and calculation times for cases in different seasons.

Season	Method	MAPE	MAE	R2	RMSE	Time
Summer	Bottom-up	**2.47%**	**2.4887**	**0.8428**	**0.3194**	200.18s
	persistence	3.36%	3.3872	0.6688	0.4570	21.47s
	ESM	3.11%	2.9670	0.7472	0.3565	40.65s
	BELM	2.61%	2.5872	0.8194	0.3385	**12.94s**
	RBM	2.77%	2.8532	0.8120	0.3866	234.25s
	LQREC	2.56%	2.5604	0.8403	0.3270	248.43s
Autumn	Bottom-up	3.24%	3.1978	0.7029	0.3039	183.30s
	persistence	5.48%	3.4773	0.4929	0.4238	21.02s
	ESM	3.25%	2.5229	0.7091	0.3110	40.89s
	BELM	**2.68%**	**2.1780**	**0.8182**	**0.2396**	**12.67s**
	RBM	7.95%	7.6034	0.3353	0.5682	232.140s
	LQREC	2.72%	2.0905	0.8937	0.2791	244.50s
Winter	Bottom-up	**2.55%**	**2.0772**	**0.8292**	**0.2635**	169.98s
	persistence	2.57%	2.1490	0.7482	4.0257	21.17s
	ESM	2.96%	2.6653	0.8142	0.3353	43.29s
	BELM	2.85%	2.5496	0.8116	0.3106	**8.61s**
	RBM	3.27%	2.8241	0.5912	0.5474	226.23s
	LQREC	2.88%	2.5556	0.8058	0.2927	236.62s



(a) Bottom-up PLF (b) Traditional PLF

FIGURE 13. PIs with nominal confidence 90% obtained by bottom-up and traditional methods in summer based on MV outgoing line load patterns.

The proposed method is also tested by predicting load not one day but several days ahead. A one-week rolling load predictions of HV substation A from August 27 to September 2 in 2018 is obtained by the proposed method, where deterministic predictions in the previous days are included in the input set to obtained load prediction in the next day. For example, to obtain load predictions on August 28, historical load data on August 27 in training and input set will be replaced by corresponding deterministic load predictions.

The performances of the proposed method in rolling test are evaluated by MAPE and ACE in TABLE 2. For the deterministic predictions, the MAPEs are stable and within 2.5% in first four days, and then gradually increase with time. For PIs, the ACEs are within 3% in the first four days, and begin to lose its accuracy since the fifth day.

E. PERFORMANCE OF THE PROPOSED METHOD OVER SEASONS

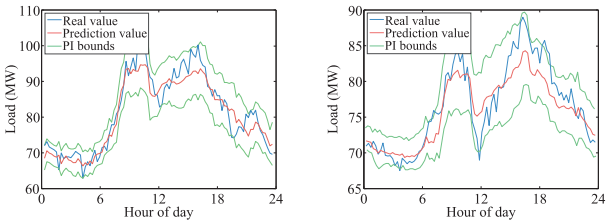
Outgoing line based bottom-up PLF method is adopted to predict load patterns of HV substation A in different seasons, to test the effectiveness of the proposed method in dealing with different data sets. The load predictions of HV substation A on October 22 2018 (autumn) and December 17 (winter) are shown in Fig. 14(a) and Fig. 14(b) respectively.

In order to better show the performances of the proposed method, a comprehensive table including performance comparisons between cases in different seasons is shown as TABLE 3 and TABLE 4. The proposed bottom-up PLF method is compared with five popular PLF methods including persistence method [31], exponential smoothing method (ESM) [32], bootstrap-based extreme learning machine method (BELM) [5], reconciled boosted models (RBM) [13] and linear quantile regression and empirical copulas based day-ahead HPLF method (LQREC) [15].

As it is shown in TABLE 3, for PIs, all PINAWs of the proposed method are smaller than the other methods. For example, compared with persistence, PIs are narrowed more than 50% on average. At the same time, QSs of the proposed method are smaller than the other methods in all three seasons, which indicate the high reliability of the constructed PIs. In terms of PICPs and ACEs, all PICPs of the proposed method are close to corresponding PINC, and all ACEs of the proposed method are close to zero, especially in summer and winter. For example, in summer, the proposed method

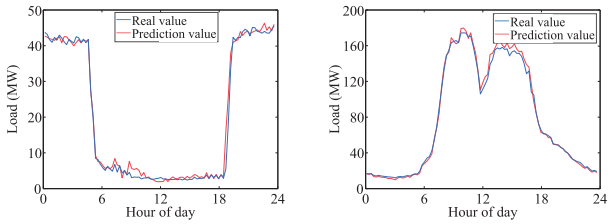
TABLE 5. Performance comparison of traditional method and proposed bottom-up method based on load patterns of MV substations (Bottom-up[a] is performed with the clustering step introduced in section II.D while Bottom-up[b] is not).

Method	Deterministic predictions				PIs with 90% PINC			QS	Time
	MAPE	MAE	R2	RMSE	ACE	PICP	PINAW		
Bottom-up[a]	**1.80%**	**7.7131**	**0.9667**	**1.5112**	6.88%	**96.88%**	0.1504	**1156.10**	385.56s
Bottom-up[b]	2.84%	8.1529	0.9628	1.5947	-16.04%	83.96%	**0.1070**	1819.59	23647.55s
traditional	3.23%	13.949	0.9611	1.6340	**-5.62%**	84.38%	0.2102	1399.38	**13.79s**



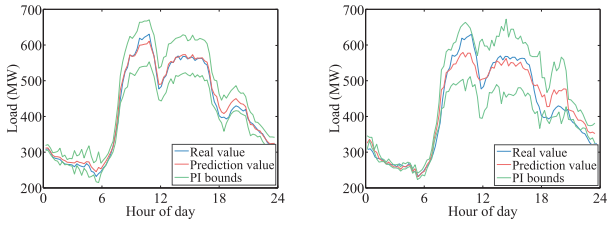
(a) Autumn (b) Winter

FIGURE 14. PIs with nominal confidence 90% of HV substation A in autumn and winter.



(a) Cluster 1 (b) Cluster 2

FIGURE 15. Day-ahead load patterns of two clusters of MV substations predicted by FNN.



(a) Bottom-up PLF (b) Traditional PLF

FIGURE 16. PIs with nominal confidence 90% obtained by bottom-up and traditional methods in summer based on MV substation load patterns.

has absolute ACEs at all confidences around 1%, which are smaller than or equal to those of the other five methods. However, in autumn, PICPs and ACEs of BELM and LQREC are better than those of the proposed method.

As it is shown in TABLE 4, for deterministic predictions, the proposed method obtains predictions with small MAPEs, MAEs, RMSEs and high R2 scores in all three seasons, especially in summer and winter. For calculation time, both the proposed method and LQREC are rather time consuming when compared with persistence, ESM and BELM. It is because that the proposed method and LQREC need to predict loads of every MV outgoing line and then form HV substation load predictions, while the other methods directly predict the loads of HV substation. However, it should be noted that the proposed method and LQREC are focus on day-ahead PLF problems, and hence calculation time will not limit the application of them significantly. Moreover, the proposed method is performed serially in this test. Since load predictions of different MV outgoing lines and MV substations are obtained independently, MV load forecasting can be performed in parallel to reduce the total time consumption.

In general, for probabilistic predictions, the proposed method outperforms other approaches with the narrower PIs and higher QS in all three seasons, and ACE closed to zero in nearly 60% circumstances. For deterministic predictions, the proposed method provides reliable deterministic predictions in all three seasons and outperforms other approaches in summer and winter.

F. HV LOAD FORECASTING BASED ON MV SUBSTATION LOAD PATTERNS

Bottom-up framework based on MV substation load patterns is tested in this section, where the load pattern of HV substation B in August 27 2018 is predicted and analyzed. The MV substations supplied by HV substation B have been divided into 7 clusters according to their load patterns in section II.D.

The day-ahead load patterns of cluster 1 and 2 are predicted by FNN and shown in Fig. 15 as an example. For the load prediction of cluster 1, the MAPE is 4.90% and the MAE is 0.772. For the load prediction of cluster 2, the MAPE is 2.33% and the MAE is 2.187. The minor prediction errors show that the classic FNN is effective to predict day-ahead load patterns of substation clusters.

To illustrate the motivation of applying clustering in this bottom-up framework, experimental comparisons between bottom-up[a] framework with clustering and bottom-up[b] framework without clustering have been provided in TABLE 5. As seen in TABLE 5, compared with the framework without clustering, the framework with clustering obtains better deterministic predictions and more reliable PIs with better ACE and QS. More importantly, the calculation time of the framework without cluster are about 60 times longer than that with clustering. In summary, better accuracy and relatively shorter calculation time motivate the usage of clustering rather than simply combining the load predictions of every MV substations.

Compared with the traditional method, as seen in TABLE 5, the proposed bottom-up[a] method with clustering obtained deterministic prediction with smaller MAPE, MAE, RMSE and higher R2 score. At the same time, though two methods obtain PIs with similar ACEs, PIs obtained by the proposed method is significantly narrowed by 35.11%

according to PINAW. Moreover, the QS of the proposed method is higher than that of the traditional method.

The load predictions obtained by bottom-up[a] method with clustering and traditional PLF method are shown in Fig. 16. It can be seen that deterministic load prediction given by bottom-up method conforms to the real load pattern, and the PIs obtained by the proposed method are obviously narrower than that obtained by traditional method.

Compared with bottom-up framework based on load patterns of MV outgoing lines, the framework based on MV substation load pattern obtains PIs with higher PICP, which implies the PIs cover more real values than expected. It may be caused by the smooth load patterns of HV substation B.

VI. CONCLUSION

A bottom-up probabilistic load forecasting method based on MV load data is proposed for HV substation STLF, which contains two specific bottom-up frameworks utilizing load data collected from MV outgoing line and MV substation respectively. In the proposed frameworks, predictions of MV loads and add-up error are obtained and aggregated to HV load predictions. Before the aggregation, probability distributions of MV load predictions are verified to be independent Gaussian distributions, which enable us to efficiently and correctly aggregate HV load prediction from MV ones.

Comprehensive numerical cases in different seasons have been studied based on MV load data in East China. Compared with traditional PLF methods, bottom-up method endows load predictions with better accuracy and narrower prediction intervals. This implies the proposed method has great potential for providing better HV transmission substation load predictions, which can help to establish generation schedules, distribute generation, analyze load flow and monitor overloads of HV transformers or transmission lines.

In future studies, more state-of-the-art NN models rather than FNN will be considered in the bottom-up PLF frameworks to further improve the prediction accuracy.

REFERENCES

[1] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecasting*, vol. 32, no. 3, pp. 914–938, Jul. 2016.
[2] N. Ding, C. Benoit, G. Foggia, Y. Besanger, and F. Wurtz, "Neural network-based model design for short-term load forecast in distribution systems," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 72–81, Jan. 2016.
[3] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: A big data approach," *Int. J. Forecasting*, vol. 32, no. 3, pp. 585–597, Jul. 2016.
[4] Y. Wang, Q. Xia, and C. Kang, "Unit commitment with volatile node injections by using interval optimization," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1705–1713, Aug. 2011.
[5] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1033–1044, May 2014.
[6] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, Jan. 2001.
[7] C. S. Chen, Y. M. Tzeng, and J. C. Hwang, "The application of artificial neural networks to substation load forecasting," *Electr. Power Syst. Res.*, vol. 38, no. 2, pp. 153–160, Aug. 1996.
[8] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of low voltage network templates—Part I: Substation clustering and classification," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 1–9, Nov. 2015.
[9] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, "Smart grid technologies: Communication technologies and standards," *IEEE Trans. Ind. Informat.*, vol. 7, no. 4, pp. 529–539, Nov. 2011.
[10] B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I. Jennett, "Incorporating practice theory in sub-profile models for short term aggregated residential load forecasting," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1591–1598, Jul. 2017.
[11] Y. Wang, Q. Chen, M. Sun, C. Kang, and Q. Xia, "An ensemble forecasting method for the aggregated load with subprofiles," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3906–3908, Jul. 2018.
[12] T. Hong, J. Xie, and J. Black, "Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting," *Int. J. Forecasting*, vol. 35, no. 4, pp. 1389–1399, Oct. 2019.
[13] C. Roach, "Reconciled boosted models for GEFCom2017 hierarchical probabilistic load forecasting," *Int. J. Forecasting*, vol. 35, no. 4, pp. 1–18, Dec. 2018.
[14] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, "Hierarchical probabilistic forecasting of electricity demand with smart meter data," *J. Amer. Stat. Assoc.*, vol. 116, no. 533, pp. 27–43, Jan. 2021.
[15] T. Zhao, J. Wang, and Y. Zhang, "Day-ahead hierarchical probabilistic load forecasting with linear quantile regression and empirical copulas," *IEEE Access*, vol. 7, pp. 80969–80979, 2019.
[16] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 911–918, Mar. 2015.
[17] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Appl. Statist.*, vol. 28, no. 1, pp. 100–108, 1979.
[18] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vols. PAMI–1, no. 2, pp. 224–227, Apr. 1979.
[19] T. S. Mahmoud, D. Habibi, M. Y. Hassan, and O. Bass, "Modelling self-optimised short term load forecasting for medium voltage loads using tunning fuzzy systems and artificial neural networks," *Energy Convers. Manage.*, vol. 106, pp. 1396–1408, Dec. 2015.
[20] S. Chemetova, P. Santos, and M. VentimNeves, "Load forecasting in electrical distribution grid of medium voltage," *Technol. Innov. Cyber-Phys. Syst.*, vol. 470, no. 1, pp. 340–349, Apr. 2016.
[21] Z. Cao, C. Wan, Z. Zhang, F. Li, and Y. Song, "Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 1881–1897, May 2020, doi: 10.1109/TPWRS.2019.2946701.
[22] E. Doveh, P. Feigin, D. Greig, and L. Hyams, "Experience with FNN models for medium term power demand predictions," *IEEE Trans. Power Syst.*, vol. 14, no. 2, pp. 538–546, May 1999.
[23] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
[24] C. N. Lu, H.-T. Wu, and S. Vemuri, "Neural network based short term load forecasting," *IEEE Trans. Power Syst.*, vol. 8, no. 1, pp. 336–342, Feb. 1993.
[25] M. D. Goodin and S. Douglas, *The Cambridge Dictionary of Statistics*. London, U.K.: Cambridge Univ. Press, 2006.
[26] D. S. Lemons, *An Introduction to Stochastic Processes in Physics*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2002.
[27] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 303–315, Feb. 2014.
[28] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *J. Amer. Stat. Assoc.*, vol. 102, no. 477, pp. 359–378, Mar. 2007.
[29] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, "Direct quantile regression for nonparametric probabilistic forecasting of wind power generation," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2767–2778, Jul. 2017.
[30] J. Hauke and T. Kossowski, "Comparison of values of Pearson's and spearman's correlation coefficients on the same set of data," *Quaestiones Geographicae*, vol. 30, no. 2, pp. 87–93, Jun. 2011.
[31] G. Sideratos and N. D. Hatziargyriou, "Probabilistic wind power forecasting using radial basis function neural networks," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1788–1796, Nov. 2012.
[32] A. Lau and P. McSharry, "Approaches for multi-step density forecasts with application to aggregated wind power," *Ann. Appl. Statist.*, vol. 4, no. 3, pp. 1311–1341, Sep. 2010.

ZHENGBANG JIANG received the B.E. degree in electrical engineering from Sichuan University, Chengdu, China, in 2016. He is currently pursuing the Ph.D. degree in electrical engineering in Zhejiang University. His research interests include the load clustering and load forecasting in distribution grid.

HAO WU (Member, IEEE) received the B.E. degree from Shanghai Jiao Tong University, Shanghai, China, in 1996, and the master's and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 1999 and 2003, respectively. He joined Zhejiang University in 2002. From 2003 to 2004, he visited The HongKong Polytechnic University and the University of Wisconsin-Madison, from 2009 to 2011. His current research interests include power system operation and stability, uncertainty analysis, cascading failure, and load modeling.

BINGQUAN ZHU received the B.E. degree from Zhejiang University, Hangzhou, China. He is currently an Engineer with Zhejiang Power Corporation. His research interest includes power system operation and dispatch.

WEI GU is currently a Senior Engineer with Zhejiang Power Corporation. His research interest includes power system operation analysis and dispatch.

YINGWEI ZHU is currently an Engineer with Jinhua Power Corporation. His research interests include power systems operation and load forecasting.

YONGHUA SONG (Fellow, IEEE) received the B.E. degree from the Chengdu University of Science and Technology, Chengdu, China, in 1984, and the Ph.D. degree from the China Electric Power Research Institute, Beijing, China, in 1989. He is currently the Rector with the University of Macau, and an Adjunct Professor with Zhejiang University. In 2004, he was elected as a Fellow of the Royal Academy of Engineering, U.K. His research interests include electricity economics and power systems operation.

PING JU (Member, IEEE) received the B.Eng. and M.Sc. degrees in electrical engineering from Southeast University, Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 1988. He is currently a Chair Professor with Zhejiang University. His current research interests include load modeling, flexible load dispatch, and integrated energy systems.