IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 28, 2020, accepted September 20, 2020, date of publication October 5, 2020, date of current version October 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3028649

Short-Term Power Load Forecasting Based on Cross Multi-Model and Second Decision Mechanism

PAN ZENG, (Associate Member, IEEE), MIN JIN, (Member, IEEE), AND MD. FAZLA ELAHE
College of Computer Science and Electronics Engineering, Hunan University, Changsha 410082, China

Corresponding author: Min Jin (jinmin@hnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61773157, and in part by the Hunan Provincial Innovation Foundation for Postgraduate under Grant CX2018B231.

ABSTRACT Short-term load forecasting (STLF) plays a vital role in the reliable, secure, and efficient operation of power systems. Since electric load variation results from diverse factors, accurate and stable load forecasting remains a challenging task. To increase the forecasting accuracy and stability, in this paper, we newly propose a short-term load forecasting method based on the cross multi-model and second decision mechanism. First, we combine horizontal and longitudinal training set selection method to construct the cross training sets, which acquire both the horizontal and longitudinal characteristics of the load variation. Second, to improve the generalization ability and extend the application scope, we construct forecasting multi-models by training multiple forecasting algorithms with cross training sets. Finally, to aggregate the forecasting outputs obtained by the forecasting multi-models, we propose a second decision mechanism based on a decision multi-model and adaptive weight allocation strategy, which overcomes the limited learning ability shortcoming of single decision models and further improves the forecasting accuracy. Case studies based on electrical load data from the state of Maine, the region of New England, Singapore, and New South Wales of Australia show that both the accuracy and the stability of the proposed method are superior to the compared models.

INDEX TERMS Short-term load forecasting, multi-model, cross training set, second decision mechanism, model aggregation.

I. INTRODUCTION

The accuracy and stability of electric load forecasting are crucial for scheduling power generation, reducing cost, assessing security, and making maintenance plans in power systems. The increasing penetration of power generation from renewable energy and the diversity of power demand in the energy market make short-term load forecasting (STLF) more difficult than ever before [1], [2]. In STLF, selecting the proper training sets and building optimization forecasting models play essential roles in improving the forecasting accuracy and stability, and much research has been proposed over the past few decades.

In terms of the training set selection, there are two major methods: the horizontal training set selection method and the longitudinal training set selection method. The horizontal training set selection method is the most commonly used training set selection method, which selects continuous data samples from historical data as the training set [3], [4]. Reference [5] selects winter months as the training set when forecasting the daily peak load of January. Reference [6] uses the previous three years' data as the training set when constructing sister models. Training sets constructed by the horizontal training set selection method contain consecutive historical data samples so that the forecasting models based on them can effectively learn the load variation trend over continuous time, and thus the forecasting accuracy can be improved. Load consumption grows with the development of the economy [7]. While the growth of the load may not be evident in the short term, significant growth can be observed when we compare the current load with the load in previous years. The horizontal training set selection method cannot adequately capture such a growth trend of the load, which limits further improvement of the forecasting accuracy.

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

The longitudinal training set selection method selects discrete data samples with the same attributes (such as the same time attribute, same week attribute) as the test sample to construct the training set [8]–[10]. Reference [11] selects samples that have the same week and same month as the test sample to construct the training set and obtains a satisfactory accuracy. Training sets selected by the longitudinal training set selection method can capture the load growth trend over a long period of time. However, they cannot reflect the load variation patterns in the short term, which limits the improvement of accuracy.

In terms of the model construction, to overcome the shortcomings of limited application scope and the generalization ability of single models, multi-model methods have been extensively studied [12]–[14]. In multi-model methods, a key technology that affects the accuracy and stability is the model aggregation strategy, which can be divided into three basic categories: averaging strategies, performance-based strategies, and learning strategies. The averaging strategy is simple and easy to apply. The most natural and widely used strategy is the simple averaging strategy, which utilizes the arithmetic mean of all outputs of different models as the final forecasting output [15]. However, the simple averaging strategy is sensitive to outliers. The trimmed averaging strategy deletes the highest and the lowest values of all the outputs and calculates the arithmetic mean of the remaining outputs to obtain the forecasting result. This strategy is more robust to outliers than simple averaging [16]. Although averaging strategies are popular and work well in some cases, they treat each model equally and fail to consider the differences between different models. Unlike averaging strategies, performance-based strategies assign weights to each model according to their performance based on the validation set. Commonly used performance-based strategies include the inverse root mean squared error (IRMSE) averaging strategy [17], which assigns weights to each model according to the root mean squared error (RMSE); the exponentially weighted averaging (EWA) strategy [18], which assigns weights according to the mean squared error; and so forth. Performance-based strategies are based on a linear integration of different individual models, and they fail to consider any nonlinear relationships between them. Therefore, learning strategies have been proposed. Reference [19] proposes an online second learning method, which uses the outputs of the multi-model, the original feature variables, and the label of the corresponding samples as a new training set. The least-square support vector machine (LSSVM) is then trained by the new training set as the decision model, which is applied to obtain the final forecasting results. Since the forecasting value of multi-models and the label of the corresponding samples are used as the input, the decision model not only can learn the nonlinear relationship between individual models but also can decrease the forecasting error by learning the errors of different individual models. This method provides a new idea for the construction of the model aggregation strategy. However, the decision model used in [19] is the LSSVM, which is still a single model with the shortcomings of limited application scope and generalization ability. Facing the nonlinearity and randomness in power load variation, model aggregation strategies have become the bottleneck for improving the forecasting performance, and therefore need to be further studied.

In summary, most traditional STLF methods either select a horizontal training set or a longitudinal training set for modeling. Whereas, horizontal training sets or longitudinal training sets can not fully reflect the load variation patterns in different directions. In terms of model aggregation, the performance of existing strategies is not satisfactory when facing the nonlinearity and randomness in power load variation. Therefore, aiming at solving the above problems, in this paper we propose an STLF method based on cross multi-model and second decision mechanism (CMSDM). First, we propose a cross training set selection method, which constructs multiple training sets from both the horizontal direction and longitudinal direction to solve the problems where the horizontal training set cannot acquire the load growth characteristics across a long period of time and where the longitudinal training set cannot reflect the load variation patterns in the short term. Second, multiple cross training sets are used to train multiple learning algorithms with different characteristics to build multiple forecasting models. The ensemble of these models is called a cross forecasting multi-model, whose generalization ability is greatly enhanced and whose application scope is widely extended. Finally, a second decision mechanism based on decision multi-model and adaptive weight allocation strategy is proposed to aggregate the outputs of the cross forecasting multi-model and to obtain the final forecasting value.

The remainder of this paper is organized as follows. Section II presents the framework and the construction process of the CMSDM. Section III presents the case studies and explores the selection of the validation length of adaptive weight allocation strategy based on public datasets. Section VI concludes this research.

II. METHODOLOGY

Fig.1 presents the framework of the proposed CMSDM, which consists of three steps.

The first step is dataset collection and splitting. In this step, we first collect natural and social factor data as the feature variables, including the maximum temperature (TEMP), minimum TEMP, air pressure, holidays, etc. The historical load data are also used as feature variables. Then we construct a decision training set and a validation set for each test sample. A decision training set is defined as a set of samples which will be used for the training of decision multi-model, and the validation set will be used for selecting optimal parameters.

The second step is the training and application of the forecasting multi-models. First, for the test sample and each sample from the decision training set, we construct a forecasting training set by selecting all the available historical samples. Then, each forecasting training set is used to train

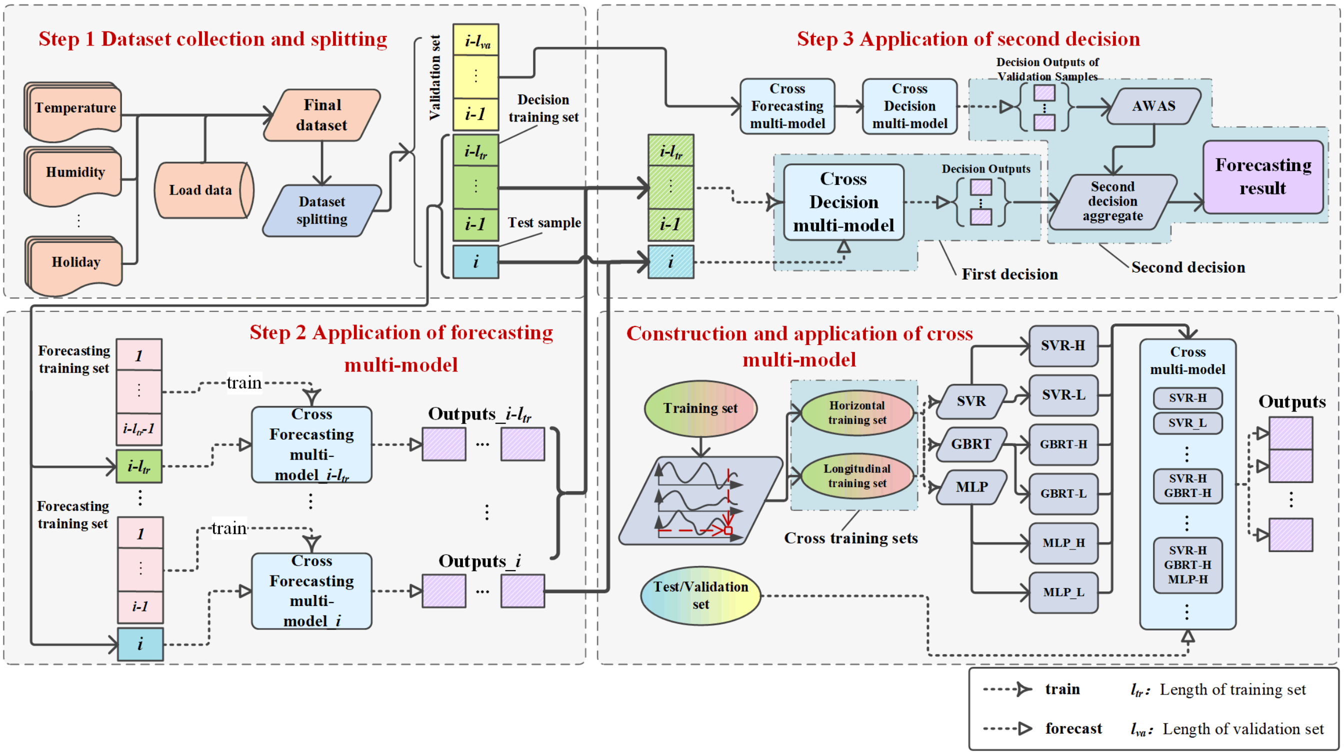


FIGURE 1. The framework of the CMSDM.

a forecasting multi-model, which is applied to make predictions for the corresponding sample, as shown in Fig.1. In this way, we will obtain multiple forecasting values for the test sample and each sample from the decision training set.

The last step is the application of the second decision mechanism. In this step, multiple forecasting values obtain in the previous step are used as additional feature variables and combined with the original ones to construct a new dataset. Then, we build a decision multi-model based on the new dataset and apply it to make predictions, which is the process of the first decision. Next, to integrate multiple outputs obtained by the decision multi-model, we assign weights to each of the individual decision models according to their performance based on the validation set, calculate the weighted summation of the outputs, and obtain the final forecasting value. This is the process of the second decision.

In terms of model construction, we propose a multi-model construction method based on the cross training set. First, we construct several horizontal training sets, which consist of a large number of continuous historical data samples, as well as longitudinal training sets, which contain discrete samples that have the same time attribute as the test sample. Then, with horizontal and longitudinal training sets, we train multiple models by applying different learning algorithms. Last, we combine these models to obtain the cross multi-model.

A. DATASET COLLECTION AND SPLITTING

There are various factors that affect the load variation, which can usually be divided into two categories: natural factors and social factors [20], [21]. Natural factors include temperature, air pressure, weather, air quality, etc., and social factors include holidays, weekdays, GDP, etc. These factors are widely used as feature variables [22]–[24]. Additionally, the current load variation is closely related to the historical load. Therefore, historical load values are also important feature variables [25], [26]. In this paper, the feature variables used are shown in (1).

$$X = [N_1, \ldots, N_i, S_1, \ldots, S_j, L_1, \ldots, L_k] \tag{1}$$

where N,S and L represent the natural, social, and historical load variables, respectively, and i,j and k represent the number of variables in each category, respectively.

As mentioned before, the construction process of the CMSDM includes two training procedures: the training of the forecasting multi-model and the training of the decision multi-model. Therefore, two groups of training sets are needed. First, the previous l_{tr} samples before the test sample are used to construct the cross training set to train the decision multi-model, and the previous l_{va} samples are used as the validation set. Then we construct forecasting training sets for the test sample and each sample of the decision training set. The forecasting training sets are then used to construct cross training sets to train the forecasting multi-models.

B. CROSS TRAINING SET SELECTION METHOD

Selecting an appropriate training set can effectively improve the performance of the forecasting model [27]. Horizontal training sets fail to acquire the load changes over a long period of time, whereas longitudinal training sets cannot reflect the load variation patterns in the short term. To overcome these shortcomings, we combine the horizontal and longitudinal training set construction methods and propose a cross training set construction method. To construct the longitudinal training set, we divide all the data samples into 85 subsets. For nonholidays, samples that have the same month and weekday are collected into the same subset. For example, all the Mondays in January of every year are in the same subset. From Monday to Sunday, January to December, there are a total of 84 subsets. All the official holidays are in another subset, resulting in 85 subsets. For each test sample, the longitudinal training set is constructed by selecting the historical data samples from the corresponding subset that contains this test sample. To better reflect the load growth trend in the longitudinal direction, the load values of the previous 7 samples are used as additional feature variables for each sample in the longitudinal training set. For the horizontal training set, all the available historical samples are used in order to fully capture the load variation trend. For example, suppose the date of the test sample is January 5th of 2015, which is a Monday; all the Mondays in January of every year before 2015 (except holidays) are used as the longitudinal training set, and all the historical samples are used as the horizontal training set. Since the training sets are constructed from the horizontal and longitudinal directions, they are called cross training sets, and models trained by them are called cross models. A model constructed by training a single algorithm with a single horizontal or longitudinal training set is called a basic cross model, and the multi-model based on cross training sets is called a cross multi-model, which is obtained by combining these basic cross models.

C. CONSTRUCTION OF THE CROSS FORECASTING MULTI-MODEL

In this step, three algorithms with different characteristics are used to construct the multi-models, namely, support vector regression (SVR), gradient boosting regression tree (GBRT), and multilayer perceptron (MLP). SVR provides a satisfactory forecasting accuracy but is sensitive to outliers. GBRT is robust to outliers, whereas its computational complexity is high. MLP has strong nonlinear learning ability, yet it is easy to overfit when the feature dimension is high and the number of training samples is small. These three algorithms are widely used in the research of load forecasting [28]–[31]. In the previous step, a longitudinal training set and a horizontal training set are constructed for the test sample and each of the decision training samples; therefore, six basic cross models can be generated by training three learning algorithms with two training sets. All the possible combinations of these basic cross models are then obtained and integrated as the cross forecasting multi-model. There are 2^6-1 possible combinations of the six basic cross models; therefore, each forecasting multi-model consists of 63 individual forecasting models. Since each of the decision training samples and the test sample need a cross forecasting multi-model, $l_{tr}+1$ cross forecasting multi-models will be built in this step. The cross forecasting multi-models are then applied to obtain multiple outputs for the test sample and each of the decision training samples.

D. SECOND DECISION MECHANISM FOR MULTI-MODEL AGGREGATION

In the previous step, multiple outputs were obtained with cross forecasting multi-models. To obtain the best aggregation results, we proposed a second decision mechanism that consists of two decision stages. In the first decision stage, aiming at making use of the strong learning ability of multi-models, we construct a decision multi-model based on different learning algorithms to aggregate the outputs of the forecasting multi-model. In the second decision stage, an adaptive weight allocation strategy (AWAS) is developed to aggregate the outputs of the decision multi-model in order to assign higher weights to models with better performance.

1) FIRST DECISION: DECISION MULTI-MODEL

After applying the cross forecasting multi-model, we can obtain multiple forecasting outputs for the test sample and the decision training samples. In this step, we combine these outputs, the original feature variables, and the label of the corresponding samples to construct a new dataset. In other words, the outputs obtained by the forecasting multi-models are used as additional feature variables to supplement the original feature variables. Then the cross decision multi-model is constructed based on the new dataset. Since the new dataset contains the outputs of forecasting multi-models, the decision multi-model based on this dataset could effectively learn the performance of each individual forecasting model, thus reducing the forecasting error.

Four training sets are constructed to train the decision multi-model, of which the first two are the same as when building the cross forecasting multi-model, namely, the horizontal training set and the longitudinal training set. The third one divides the 12 months into hot months and cold months according to temperature. When the test sample belongs to hot months, historical samples from hot months are used as the training set, and vice versa. This method essentially belongs to the horizontal training set construction method. The last method recursively uses the previous 1095 days (3 years) of the test sample as the training set, which also belongs to the horizontal training set selection method. These four training sets are referred to as the **ALL**, **LNG**, **HCD** and **REC** training sets, respectively. Next, SVR, GBRT, and MLP are individually trained using these four training sets, so 12 basic cross decision models can be obtained. Different from the construction of forecasting multi-models, in this step only the 12 basic cross decision models are used to construct the decision multi-model rather than combinations of them. In this way, a decision multi-model with 12 outputs for each test sample is constructed.

2) SECOND DECISION: ADAPTIVE WEIGHT ALLOCATION STRATEGY

Since there are 12 outputs of the decision multi-model for the test sample, the aggregation strategy becomes the key to improving the forecasting accuracy and stability. Fundamentally speaking, the aggregation of the multi-model's outputs is to calculate the weighted average of them. For a multi-model with m outputs, the outputs at time t can be represented by $\hat{y}_{1,t}$,$\hat{y}_{2,t}$,... $\hat{y}_{m,t}$, and the aggregated output can be given by:

$$\hat{y}_t^a = \sum_{i=1}^{m} \omega_{i,t}\hat{y}_{i,t} \tag{2}$$

where $\omega_{i,t}$ is the weight for the ith output at time t. Therefore, the essential step of a model aggregation strategy is to calculate the weight assigned to each output.

In the previous step, 12 outputs were obtained by the decision multi-model for the test sample. In this step, we remove the individual decision models with the highest, the second highest, the lowest and the second lowest forecasting values, and assign weights to the remaining 8 individual decision models according to their performance based on the validation set. We apply the mean absolute error (MAE) to evaluate the performance of the models, as shown in (3).

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|\hat{y}_t - y_t\right| \tag{3}$$

where $\hat{y}_t$ represents the forecasting load and y_t represents the actual load.

Then, the weights for individual models are calculated according to

$$\omega_{i,t} = \frac{\exp(-\eta MAE_i)}{\sum_{i=1}^{m}\exp(-\eta MAE_i)} \tag{4}$$

where MAE_i represents the MAE of the ith model in the validation set, and η is the learning parameter that controls the sensitivity of the algorithm. This strategy is referred to as the adaptive weight allocation strategy (AWAS). The final forecasting value is obtained by calculating the weighted average of the remaining 8 forecasting values.

III. CASE STUDY

In this section, we perform case studies based on real-world public datasets to evaluate the performance of the proposed CMSDM. We use the SVR, GBRT, and MLP package provided by scikit-learn[1]. The kernel function of SVR is Linear, and the loss function of GBRT is the least squares regression. For MLP, the solver for weight optimization is *lbfgs*, the number of hidden layers is 2, and the number of nodes in the two hidden layers are 5 and 2, respectively. These parameters are determined by the trial and error method, and other parameters are determined by the grid search method. The learning parameter η of AWAS is 0.01.

In addition to the MAE, the mean absolute percentage error (MAPE) and the root mean squared error (RMSE) are used to evaluate the forecasting performance, as shown in (5) and (6).

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \times 100\% \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{y}_t - y_t)^2} \tag{6}$$

A. CASE I: EFFECTIVENESS OF CROSS TRAINING SET

In this case, the daily peak load data of Maine[2] from 2003 to 2015 are used to demonstrate the effectiveness of the multi-model based on the cross training set. Data from 2013 to 2015 are used for testing. Natural feature variables used in this case include precipitation, maximum TEMP, average TEMP, minimum TEMP, average dew point TEMP, average air pressure, average visibility, fastest 5-second wind speed, average wind speed, TEMP at peak load time, and dew point TEMP at peak load time. Social feature variables include holidays, observances, and 7 weekday attributes. The peak load values of the previous 7 days are also used as feature variables. We compare the forecasting multi-models based on horizontal, longitudinal, and cross training sets. First, we construct the horizontal, longitudinal, and cross training sets. Then, with each of these training sets, SVR, GBRT, and MLP are trained and constructed as a multi-model separately. In this way, we can build 3 multi-models, namely, the horizontal multi-model, longitudinal multi-model, and cross multi-model. Last, we apply SVR as the decision model to aggregate the outputs of each multi-model. For example, to construct the horizontal multi-model with SVR, GBRT, and MLP, three basic models are obtained based on the horizontal training set, and the combinations of these basic models are used to construct the horizontal multi-model. Therefore, the horizontal multi-model contains 7 individual models.

The results are shown in Fig.2. The accuracy of the cross forecasting multi-model is higher than that of the horizontal and longitudinal forecasting multi-models for all three years. The accuracy obtained by the horizontal multi-model and the longitudinal multi-model are very close to each other. Horizontal training sets contain continuous data samples in the horizontal direction, whereas longitudinal training sets include discrete data samples in the longitudinal direction. Cross training sets make use of the advantages of the horizontal and longitudinal training sets by combining them, which could fully reflect the load variation trend in two directions and is of great importance for improving the forecasting accuracy.

There are two training procedures for each of these three multi-models: the training of forecasting multi-model and the training of the decision model (SVR). In this case, the time

[1] https://scikit-learn.org/

[2] https://www.iso-ne.com/isoexpress/web/reports/load-and-demand/-/tree/zone-info

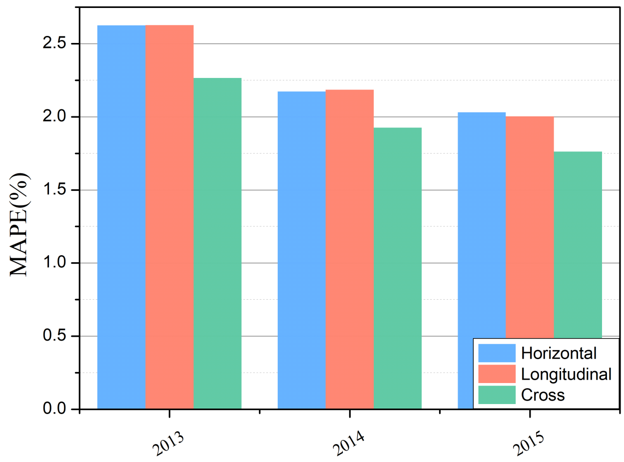


FIGURE 2. **Forecasting accuracy based on horizontal, longitudinal, and cross multi-model.**

cost of training the decision model based on these three methods are almost the same. The average time costs of forecasting one sample using horizontal, longitudinal, and cross multi-model are 2.12s, 1.98s, and 2.42s, respectively. In the procedure of training the forecasting multi-model, since the cross multi-model consists of 63 individual models, whereas the horizontal and longitudinal multi-model only contain 7 individual models, the time cost of training the cross multi-model is approximately 9 times of the time cost of training the horizontal and longitudinal multi-model. In this procedure, the time costs of forecasting one sample of these three multi-models are 1.53s, 1.49s, and 13.44s. Although the time cost of training the cross multi-model is higher than that of the horizontal and longitudinal multi-model, it is still acceptable and fully meets the actual application requirements.

B. CASE II: EFFECTIVENESS OF THE SECOND DECISION MECHANISM

In this case, we compare the forecasting performance of the proposed second decision mechanism with that of single decision models and other model aggregation strategies.

The datasets used in this case are the daily peak load data of Maine from 2003 to 2015, that of Singapore[3] from 2006 to 2018 and that of New South Wales[4] (NSW) from 2003 to 2017. For the Maine dataset, the feature variables used in this case are the same as those in Case I. For the Singapore dataset, the feature variables are precipitation, average TEMP, maximum TEMP, minimum TEMP, average wind speed, maximum wind speed, the peak load of the previous 7 days, weekday attributes, and holidays. The feature variables used for the NSW dataset are the maximum TEMP, minimum TEMP, average TEMP, maximum dew point TEMP, average dew point TEMP, minimum dew point TEMP, maximum humidity, minimum humidity, average humidity, maximum air pressure, minimum air pressure, average air pressure, electricity price, the peak load of the previous 7 days and weekday attributes. In each case, data from the last year are used for testing, and the remainder is used for constructing the training set. The length of the validation set is 7 — in other words, the previous 7 samples before the testing sample are used to calculate the weights or select the best individual models.

Since the cross decision multi-model consists of 12 individual decision models, we compare the accuracy of these 12 individual decision models with that of the proposed second decision mechanism in order to demonstrate the effectiveness of the proposed method. The model aggregation strategy used in the second decision mechanism is AWAS. We also compare AWAS with the following model aggregation strategies.

1) SIMPLE AVERAGING

Simple averaging is the most straightforward averaging strategy that calculates the arithmetic mean of all outputs. While it is simple, easy to apply, and has been widely used in multiple areas [32], [33], simple averaging is not robust enough to outliers.

2) TRIMMED AVERAGING

Trimmed averaging (TA) is an extension of simple averaging, which removes two extreme values of the outputs and calculates the arithmetic mean of the remaining outputs as the result.

3) WINDSORIZED AVERAGING

Windsorized averaging (WA) replaces the extreme values by the second highest and the second lowest outputs and calculates their arithmetic mean.

4) INVERSE ROOT MEAN SQUARED ERROR

In regression problems, RMSE is an important performance metric, and can, therefore, be used to calculate the weights. Reference [17] proposed the IRMSE strategy, which computes the weights according to

$$\omega_{i,t} = \frac{\frac{1}{RMSE_i}}{\sum_{i=1}^{m} \frac{1}{RMSE_i}} \tag{7}$$

where $RMSE_i$ represents the $RMSE$ of the ith model in the validation set.

5) MODEL COMPETITION STRATEGY

We introduce two model competition strategies for comparisons with the strategies mentioned above. According to the performance in a static validation set, the first one uses the output of the best individual model as the final output, which can be denoted by the Best-S strategy. The second one can be

[3] https://www.ema.gov.sg/statistic.aspx?sta_sid=20140826Y84sgBebjwKV

[4] https://aemo.com.au/en/energy-systems/electricity/national-electricity-market-nem/data-nem/aggregated-data

	Model	Maine, 2015			Singapore, 2018			NSW, 2017		
		MAPE(%)	MAE(MW)	RMSE(MW)	MAPE(%)	MAE(MW)	RMSE(MW)	MAPE(%)	MAE(MW)	RMSE(MW)
Single decision model	SVR-ALL	1.77%	27.06	34.93	1.27%	84.28	114.11	3.33%	316.54	451.99
	SVR-HCD	1.75%	26.80	35.07	1.41%	93.29	125.53	3.30%	314.21	452.09
	SVR-REC	1.76%	26.92	35.29	1.19%	78.53	108.15	3.86%	375.08	579.15
	SVR-LNG	5.01%	77.09	104.49	2.76%	187.14	242.92	5.75%	546.78	787.24
	GBRT-ALL	1.70%	26.13	34.66	1.18%	77.21	113.09	2.90%	276.40	419.15
	GBRT-HCD	1.67%	25.61	34.13	1.21%	79.09	120.72	2.93%	280.63	418.97
	GBRT-REC	1.85%	28.55	37.30	1.13%	74.06	107.40	2.92%	275.76	418.80
	GBRT-LNG	2.17%	33.37	44.50	1.44%	96.07	127.16	3.79%	360.58	535.80
	MLP-ALL	3.00%	46.27	78.86	1.28%	83.83	120.12	3.09%	294.82	421.60
	MLP-HCD	2.60%	40.38	63.97	1.18%	77.24	114.84	2.97%	282.62	404.88
	MLP-REC	2.00%	30.50	40.63	1.26%	82.62	113.94	3.36%	320.92	486.16
	MLP-LNG	26.91%	417.55	542.48	20.15%	1359.62	1691.76	21.15%	2057.47	3673.96
Second decision mechanism	Simple	2.70%	41.71	54.91	1.87%	124.44	155.92	3.37%	322.39	486.60
	TA	1.67%	25.75	33.97	1.05%	69.37	97.28	2.91%	276.53	409.70
	WA	1.73%	26.68	35.50	1.06%	70.05	97.58	2.94%	279.30	412.55
	IRMSE	1.65%	25.26	33.10	1.05%	69.00	97.35	2.87%	271.97	404.18
	Best-D	1.91%	29.30	41.49	1.23%	80.91	120.47	3.05%	291.09	426.72
	Best-S	2.60%	40.38	63.97	1.18%	77.24	114.84	2.97%	282.62	404.88
	AWAS	1.62%	24.89	32.51	1.05%	68.97	97.67	2.82%	268.36	403.27

FIGURE 3. Heat map of the MAPE, MAE, and RMSE values of single decision models and aggregation strategies of the second decision mechanism.

called the Best-D strategy, which uses a dynamic validation set to select the best model. The Best-S strategy only selects the best model once, whereas the Best-D selects the best model in a rolling manner. In this case, the validation set for the Best-S strategy is the previous year of the test sample. For example, in the Maine case, data from 2015 are used for testing, and data from 2014 are used as the validation set for selecting one best individual decision model.

Fig.3 presents the results. We can see that the accuracy of the MLP-LNG model in all three cases is very poor. The reason for the poor accuracy is that the number of samples of the longitudinal training set is relatively small and that the feature dimension is large. Therefore, the MLP-LNG model has overfitted the training data. Here we do not remove the MLP-LNG model from the decision multi-model. Instead, we keep it as noise to test the robustness of the model aggregation strategies.

According to the results based on the three datasets, we can see that the accuracy obtained by IRMSE and AWAS are better than any of the single decision models. TA and WA also achieve a satisfactory forecasting performance, which shows the effectiveness of the second decision mechanism. The proposed second decision mechanism could combine the outputs of multiple decision models and overcome the shortcoming of limited learning ability. Due to the overfitting problem of the MLP-LNG, the performance of simple averaging is poor. The forecasting accuracy of the Best-S and Best-D are also at the medium level, suggesting that model competition strategy may not be the best model aggregation strategy. Finally, comparing TA, IRMSE, and AWAS, we can see that the accuracy of these strategies in the Singapore dataset are almost the same, whereas in the Maine and NSW datasets, AWAS outperforms TA and IRMSE. AWAS discards four extreme values of the outputs and assigns weights to each of the individual models according to their performance based on the validation set. The application of AWAS could assign higher weights to models with better performances and reduce the influence of outliers.

TABLE 1. The time costs (s) of model aggregation strategies used in Case II.

	Maine,2015	Singapore,2018	NSW,2017
Simple	3.22×10^{-5}	3.38×10^{-5}	3.07×10^{-5}
TA	5.23×10^{-3}	5.35×10^{-3}	1.33×10^{-3}
WA	7.2×10^{-3}	7.37×10^{-3}	1.84×10^{-3}
IRMSE	0.43	0.44	0.11
Best_D	2.68×10^{-2}	2.66×10^{-2}	6.68×10^{-3}
Best_S	6.52×10^{-4}	6.42×10^{-4}	1.34×10^{-4}
AWAS	0.44	0.46	0.11

Table 1 shows the time cost of all the model aggregation strategies used in this case. We can see that the time costs of averaging strategies are significantly lower than performance-based strategies. Since IRMSE, Best-D, and AWAS use a dynamic validation set when aggregating the outputs, the time cost of these three methods are higher than other strategies. Although the time cost of AWAS the highest, it still meets the application requirements.

Fig.4 presents the normalized daily peak load of Maine, Singapore, and NSW. The load variation patterns of these three regions are significantly different. Singapore is an island country in Southeast Asia, where the variation range of temperature throughout the entire year is not broad, and thus, the electric load consumption does not vary much; in addition, the difference in the load variation in different seasons is not obvious. However, there is a significant seasonal periodicity of the load variation in Maine and NSW. The daily peak load of NSW is higher in January, February, May to August and December, and lower in other months. The load in Maine changes more in different seasons, with a higher load consumption from January to March, May to July, and December. Fig.5 shows the scatter plot of the normalized daily peak load and the average temperature in these three

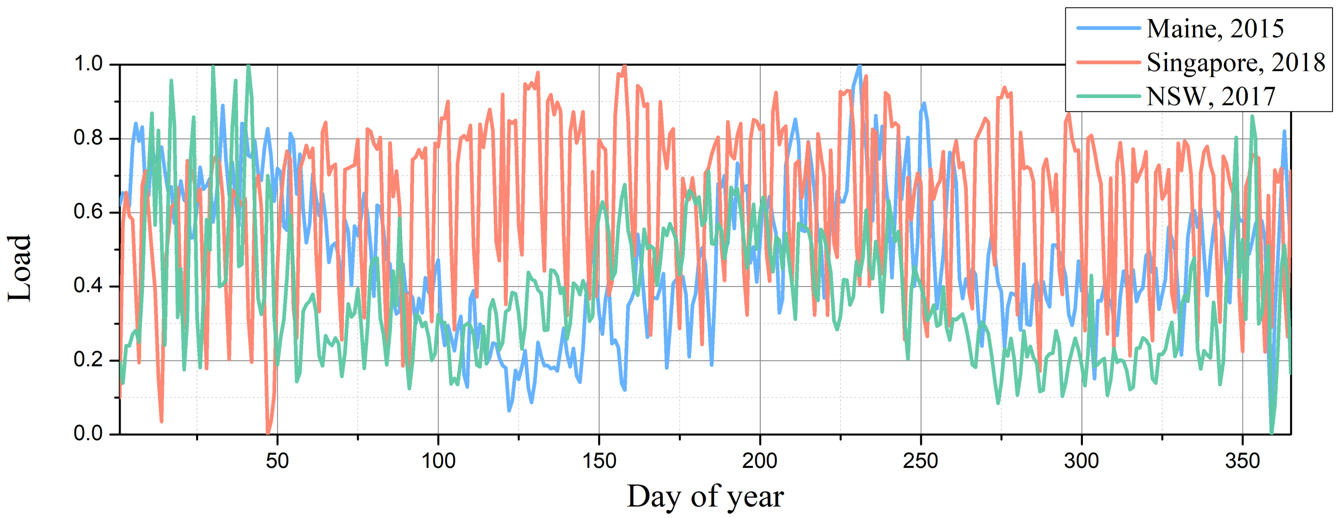


FIGURE 4. Normalized daily peak load of Maine, Singapore, and NSW.

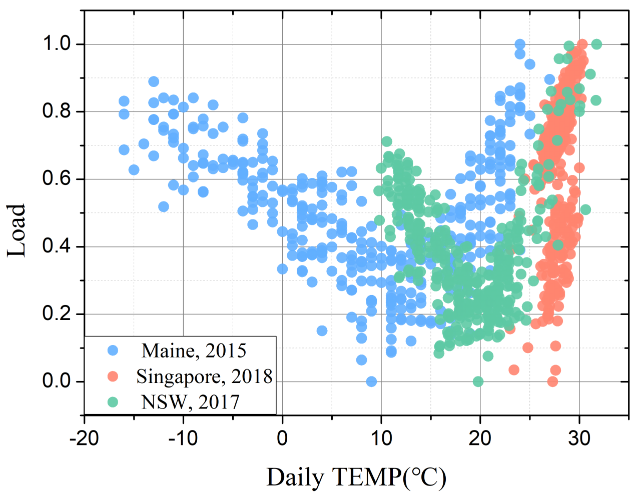


FIGURE 5. Scatter plot of the normalized daily peak load and average temperature of Maine, Singapore, and NSW.

TABLE 2. Average values and standard deviations of MAPE, MAE and RMSE obtained by IRMSE, Best-D and AWAS based on different validation set lengths.

		IRMSE	Best_D	AWAS
MAPE(%)	AVG	1.6478	1.7803	**1.6253**
	STD	4.36×10^{-2}	14.19×10^{-2}	$\mathbf{9.85\times10^{-6}}$
MAE(MW)	AVG	25.278	27.3100	**24.9176**
	STD	0.6708	2.1818	**0.0161**
RMSE(MW)	AVG	33.6019	37.7181	**32.5059**
	STD	3.2536	12.1139	**0.0167**

regions. It is obvious that the load and temperature of Singapore show a linear relationship and fluctuate little throughout the entire year. The relationship between the load and temperature of Maine and NSW are similar, showing a V shape, and the load is higher when the temperature is lower or higher. The difference is that, in Maine, the load consumption trough occurs when the temperature is approximately 12°C, whereas in NSW, the trough occurs when the temperature is approximately 20°C. This outcome is clearly related to the geographical location, climate, environment, residents' living habits, etc. It can also be seen from the figure that Maine has a significant temperature fluctuation throughout the year, the minimum temperature is approximately −20°C and the maximum is approximately 30°C. Whereas the minimum and maximum temperatures of NSW are approximately 10°C and 32°C, and those of Singapore are approximately 22°C and 30°C, respectively. Maine is located in New England in the United States, Singapore is in Southeast Asia, and New South Wales is in the southeast of Australia. These three regions have completely different characteristics in terms of weather, climate, load variation pattern, residents' living habits, geographical location, etc. The experimental results based on these three datasets can fully illustrate that the effectiveness of the second decision mechanism based on the decision multi-model and AWAS is dataset-independent, showing that this method has excellent generalization ability and wide application scope.

C. CASE III: INFLUENCE OF THE VALIDATION SET LENGTH AND LEARNING PARAMETER η

For AWAS, IRMSE, and Best-D, which need a dynamic validation set to calculate weights, in the previous case studies the length of the validation set was 7. In this case, to study the influence of different validation set lengths on the forecasting accuracy of CMSDM and to explore the best length of the validation set, we run the experiment 90 times. The initial length of the validation set is set to 1, and in each iteration, we increase the length of the validation set by 1 and compare the forecasting performances of these three model aggregation strategies. The dataset used in this case is the daily peak load of Maine, and the data from 2015 are used for testing.

Table 2 presents the average values and standard deviations of the MAPE, MAE, and RMSE obtained by three strategies with different validation set lengths. From the table, we can see that the average values of these three performance metrics for the AWAS are lower than those of IRMSE and Best-D, and the standard deviations are much lower. The results show that AWAS not only has higher forecasting accuracy but also higher stability than the other strategies, demonstrating that AWAS is not sensitive to the validation set length and could provide high accuracy, high stability, and robustness aggregation results.

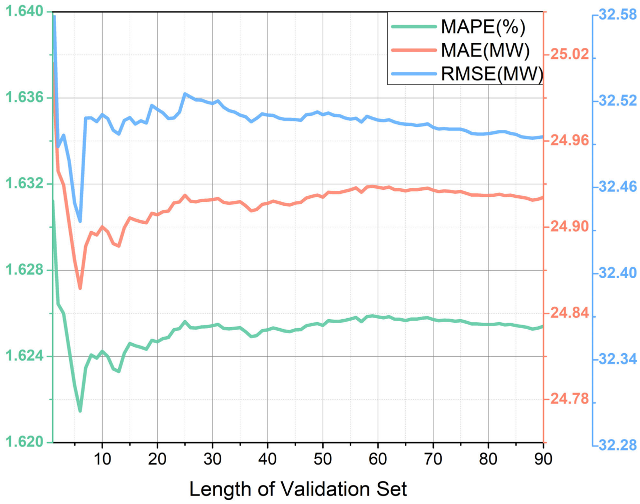


FIGURE 6. MAPE, MAE, and RMSE obtained by AWAS based on different validation set lengths.

Fig.6 shows the MAPE, MAE, and RMSE obtained by AWAS with different validation set lengths. The results show that when the validation set length is 6, AWAS obtains the best forecasting performance. When the validation set length is larger than 6, the forecasting performance will decrease slowly, but the fluctuation in the accuracy is not obvious. Nevertheless, the difference between the maximum and minimum values of the three metrics are 0.0097%, 0.1565 MW, and 0.1785 MW, respectively. Considering the average values and standard deviations of MAPE, MAE, and RMSE, we believe that the length of the validation set has little influence on the AWAS's accuracy.

The value of the learning parameter η controls the sensitivity of AWAS. When $\eta = 0$, all the individual models get the same weight, and with the increase of η, individual models with better performance will get higher weights. To explore the influence of η, we set the value of η from 0, 0.01, 0.02 to 1.00, run the experiment 101 times, and compare the MAPE, MAE and RMSE obtained in each iteration. We can see from Fig.7 that the variation of MAPE and MAE are similar to each other, they decrease with the increase of η when it is less than 0.32, and after that, the value of MAPE and MAE increase. When the value of η is 0.32, the lowest MAPE and MAE could be obtained. Whereas for RMSE, the variation pattern

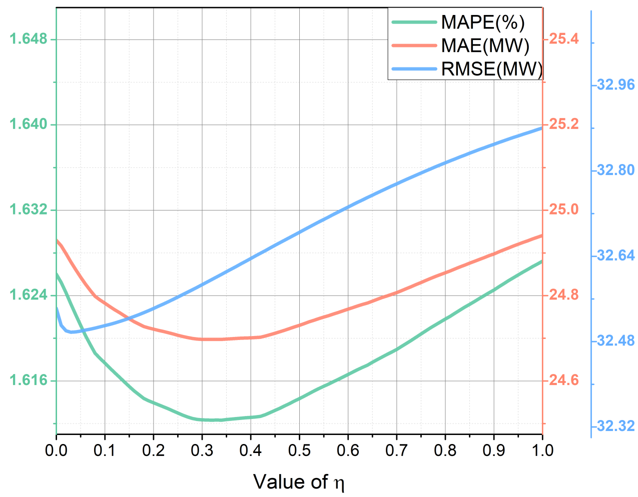


FIGURE 7. MAPE, MAE, and RMSE obtained by AWAS based on different values of η.

is different, and the lowest RMSE could be observed when the value of η is 0.03. Nevertheless, the accuracy obtained by AWAS with different values of η, in this case, is still higher than that obtained by other strategies. This case shows that the stability and accuracy of the AWAS are superior to other strategies.

D. CASE IV: COMPARISON BETWEEN CMSDM AND OTHER STATE-OF-THE-ART METHODS

In this case, we compare the proposed CMSDM with the method proposed by [19], [34], and [35], which can be referred to as the HEFM, BART, and Bi-LSTM method based on AM and RU. The datasets used in this case are the daily total load of New England[5] from 2003 to 2015 and the half-hourly load of NSW from January 1, 2009, to January 6, 2010.

1) COMPARISON WITH HEFM AND BART

HEFM is an online second learning method based on multi-models, and it uses LSSVM as the decision model in the decision stage. Bayesian additive regression trees (BART) is a Bayesian sum-of-tree model [36]. In STLF, BART is considered to be an accurate model that could effectively capture the nexus between electricity consumption and climate variability [34]. In this case, we compare the CMSDM with HEFM and BART based on the dataset of the daily total demand of New England. The feature variables used are dry bulb TEMP, dew point TEMP, holidays, observances, date index, weekday attributes, and the load values of the previous 7 days, and the data from 2015 are used for testing. The validation set length is 6.

A comparison of the MAPEs obtained by the three methods are shown in Fig.8. As we can see, CMSDM outperforms

[5]https://www.iso-ne.com/isoexpress/web/reports/load-and-demand/-/tree/zone-info

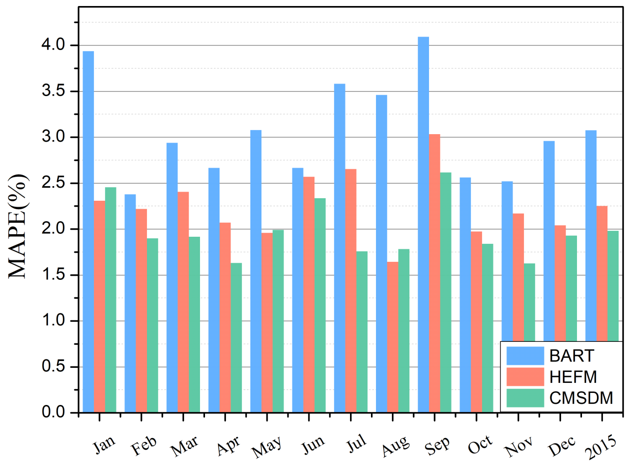


FIGURE 8. MAPE of BART, HEFM, and CMSDM.

TABLE 3. MAPE and RMSE obtained by the CMSDM and Bi-LSTM methods based on AM and RU.

Model	MAPE(%)	RMSE(MW)
Bi-LSTM method based on AM and RU	1.030	103.53
CMSDM	**0.5979**	**63.00**

HEFM and BART in 9 out of 12 months, and in January, May, and August, the MAPE of the CMSDM is slightly higher than that of HEFM. The MAPE of CMSDM, HEFM, and BART for 2015 are 1.98%, 2.25%, and 3.08%, respectively. The CMSDM outperforms HEFM and BART and increases the forecasting accuracy by 12.0% and 35.7%, respectively. The performance of CMSDM and HEFM is better than that of BART, suggesting that multi-model methods could obtain better results than single model methods. Comparing HEFM with CMSDM, in the decision stage the decision model used by HEFM is LSSVM, which is a single model, whereas CMSDM applies the second decision mechanism based on the decision multi-model and AWAS in the decision stage, which improves the learning ability of the nonlinear relationship and adjusts the forecasting results according to the performance of individual models and finally improves the forecasting performance. The average time costs of CMSDM, HEFM, and BART for each sample are 63.4s, 46.4s, and 142.2s, and all of them are entirely acceptable.

2) COMPARISON WITH BI-LSTM METHOD BASED ON AM AND RU

The Bi-LSTM method based on AM and RU is a deep learning method proposed by [35], which combines the attention mechanism, rolling update, and bi-directional long short-term memory neural network. Case studies based on NSW half-hourly data show that the Bi-LSTM method based on AM and RU is superior to other methods. Reference [35] used 9 feature variables, including the previous 3 historical load values, dry bulb TEMP, dew point TEMP, wet bulb TEMP, humidity, electricity price, and load value at the same time of the previous day. The data from December 31, 2009 to January 6, 2010, are used for testing. To compare with the Bi-LSTM method based on AM and RU, we use the same feature variables and apply the CMSDM to the same dataset used by [35]. In this case, the horizontal training set contains the previous 1000 samples of the test sample, and the longitudinal training set contains all historical samples at the same time as the test sample. The cross training set is constructed by combining the horizontal training set and the longitudinal training set. For the decision multi-model, in addition to the two training sets construction method used for forecasting multi-model, another training set construction method that recursively used the previous 3000 historical samples at the same time as the test sample is applied to construct the cross training set. The validation set length is 6.

The comparison result is shown in Table 3, and the forecasting accuracy of Bi-LSTM method based on AM and RU is directly cited from [35]. From Table 3 we can see that the proposed CMSDM outperforms the Bi-LSTM method based on AM and RU and improves the forecasting accuracy by 42.0% and 39.1% in terms of the MAPE and RMSE, respectively. Although the proposed CMSDM is not a deep learning method, it can still obtain better forecasting performance and shows better learning ability. According to [35], the time cost of the Bi-LSTM method based on AM and RU is 24.4s. In this case, the time cost of CMSDM for making a prediction for one sample is 32.7s. However, CMSDM is based on the one-step-ahead forecasting principle to obtain the best forecasting accuracy. Hence the overall time cost is significantly higher than the method proposed by [35]. Nevertheless, since this case is based on a half-hourly load dataset, the time cost of CMSDM is still acceptable and fully meets the application requirements.

Table 4 shows the comparison between CMSDM, HEFM, BART, and Bi-LSTM method based on AM and RU proposed by [35]. We can see from the comparison that the proposed CMSDM is different from other methods in multiple aspects. In terms of forecasting performance, the comparison results of the CMSDM and two single model methods, namely, the BART and Bi-LSTM method based on AM and RU, demonstrate the advantages of multi-model methods.

TABLE 4. The difference of the CMSDM with other methods.

	Category	Training set selection	Decision mechanism
BART	Shallow learning, single model	Horizontal training set	——
Bi-LSTM method based on AM and RU	Deep learning, single model	Horizontal training set	——
HEFM	Multi-model	Bootstrap sampling	Single decision model
CMSDM	Multi-model	Cross training set	Second decision mechanism

Compared with HEFM, the CMSDM applies the second decision mechanism, which overcomes the shortcoming of single decision models' insufficient ability to learn nonlinear relationships, and makes full use of the forecast capabilities of different models and finally improves the forecasting accuracy and stability. Case studies based on real-world datasets from different regions show that the CMSDM is not limited by datasets and can achieve accurate and robust forecasting results in different regions and time periods.

IV. CONCLUSION

This paper proposes an STLF method based on cross multi-model and second decision mechanism. The cross training set constructed in the CMSDM combines the advantages of the horizontal training set and the longitudinal training set. Models built based on the cross training set can fully learn the load variation trend, and the forecasting performance can be improved. The second decision mechanism is used to aggregate the outputs of the forecasting multi-model, which could significantly improve the learning ability of the nonlinear relationship, adaptively adjust the aggregation results of multi-model outputs, and effectively improve the forecasting accuracy and stability of the CMSDM. Case studies were conducted based on real-world datasets, including load data from the state of Maine, the region of New England, Singapore, and New South Wales of Australia. Since these datasets are from different regions, where the climate, residents' living habits, weather, and geographical environment are completely different from one another, the experimental results based on these datasets could demonstrate the strong generalization ability and wide application scope of the CMSDM.

The model aggregation strategy is the key to improving the forecasting accuracy of multi-model methods. We compare 8 model integration strategies based on the load data from Maine. The case study shows that simple averaging is not sufficiently robust when facing outliers. The robustness and accuracy of TA and WA are better than those of simple averaging. The IRMSE assigns weights to individual models based on their performance and provides satisfactory results, but it is not sufficiently stable or robust. Model competition strategies cannot provide an accurate aggregation output. Compared with these strategies, the aggregation results of AWAS are more accurate and robust, and less sensitive to outliers or the validation set lengths. The application of the AWAS could effectively enhance the accuracy and stability of multi-model methods.

The proposed CMSDM is not limited to STLF. We can also apply this method to other regression problems, such as traffic flow forecasting and air quality forecasting. With a few modifications, the framework of the CMSDM can also be applied to solve classification problems. However, when facing different problems and datasets with different characteristics, how do we determine the best length of the validation set? How do we select learning algorithms to construct the multi-model? These problems are worthy of further research.

REFERENCES

[1] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets: Operational Problems*, vol. 205. Springer, 2013.

[2] A. Motamedi, H. Zareipour, and W. D. Rosehart, "Electricity price and demand forecasting in smart grids," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 664–674, Jun. 2012.

[3] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[4] L. Yin, Z. Sun, F. Gao, and H. Liu, "Deep forest regression for short-term load forecasting of power systems," *IEEE Access*, vol. 8, pp. 49090–49099, 2020.

[5] B.-J. Chen, M.-W. Chang, and C.-J. Lin, "Load forecasting using support vector machines: A study on EUNITE competition 2001," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1821–1830, Nov. 2004.

[6] B. Liu, J. Nowotarski, T. Hong, and R. Weron, "Probabilistic load forecasting via quantile regression averaging on sister forecasts," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 730–737, Mar. 2017.

[7] J. Moral-Carcedo and J. Pérez-García, "Integrating long-term economic scenarios into peak load forecasting: An application to Spain," *Energy*, vol. 140, pp. 682–695, Dec. 2017.

[8] L. Xiao, W. Shao, C. Wang, K. Zhang, and H. Lu, "Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting," *Appl. Energy*, vol. 180, pp. 213–233, Oct. 2016.

[9] L. Xiao, W. Shao, M. Yu, J. Ma, and C. Jin, "Research and application of a combined model based on multi-objective optimization for electrical load forecasting," *Energy*, vol. 119, pp. 1057–1074, Jan. 2017.

[10] W. Yang, J. Wang, and R. Wang, "Research and application of a novel hybrid model based on data selection and artificial intelligence algorithm for short term load forecasting," *Entropy*, vol. 19, no. 2, p. 52, Jan. 2017.

[11] J. Wang, W. Yang, P. Du, and Y. Li, "Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system," *Energy*, vol. 148, pp. 59–78, Apr. 2018.

[12] 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.

[13] Q. Xu, X. Yang, and X. Huang, "Ensemble residual networks for short-term load forecasting," *IEEE Access*, vol. 8, pp. 64750–64759, 2020.

[14] S. Shan, B. Cao, and Z. Wu, "Forecasting the short-term electricity consumption of building using a novel ensemble model," *IEEE Access*, vol. 7, pp. 88093–88106, 2019.

[15] J. Nowotarski, E. Raviv, S. Trück, and R. Weron, "An empirical comparison of alternative schemes for combining electricity spot price forecasts," *Energy Econ.*, vol. 46, pp. 395–412, Nov. 2014.

[16] J. Nowotarski, B. Liu, R. Weron, and T. Hong, "Improving short term load forecast accuracy via combining sister forecasts," *Energy*, vol. 98, pp. 40–49, Mar. 2016.

[17] F. X. Diebold and P. Pauly, "Structural change and the combination of forecasts," *J. Forecasting*, vol. 6, no. 1, pp. 21–40, 1987.

[18] N. Littlestone and M. K. Warmuth, "The weighted majority algorithm," *Inf. Comput.*, vol. 108, no. 2, pp. 212–261, Feb. 1994.

[19] M. Zhou and M. Jin, "Holographic ensemble forecasting method for short-term power load," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 425–434, Jan. 2019.

[20] Y.-M. Wi, S.-K. Joo, and K.-B. Song, "Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 596–603, May 2012.

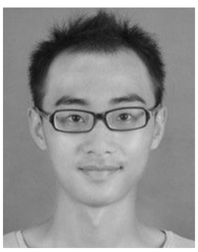
[21] J. Zhang, Y.-M. Wei, D. Li, Z. Tan, and J. Zhou, "Short term electricity load forecasting using a hybrid model," *Energy*, vol. 158, pp. 774–781, Sep. 2018.

[22] M. R. Haq and Z. Ni, "A new hybrid model for short-term electricity load forecasting," *IEEE Access*, vol. 7, pp. 125413–125423, 2019.

[23] Z. Yu, Z. Niu, W. Tang, and Q. Wu, "Deep learning for daily peak load forecasting—A novel gated recurrent neural network combining dynamic time warping," *IEEE Access*, vol. 7, pp. 17184–17194, 2019.

[24] P. Zeng, C. Sheng, and M. Jin, "A learning framework based on weighted knowledge transfer for holiday load forecasting," *J. Modern Power Syst. Clean Energy*, vol. 7, no. 2, pp. 329–339, Mar. 2019.

[25] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3943–3952, Jul. 2019.
[26] L. Han, Y. Peng, Y. Li, B. Yong, Q. Zhou, and L. Shu, "Enhanced deep networks for short-term and medium-term load forecasting," *IEEE Access*, vol. 7, pp. 4045–4055, 2019.
[27] P. Zeng and M. Jin, "Peak load forecasting based on multi-source data and day-to-day topological network," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 6, pp. 1374–1381, Mar. 2018.
[28] Z. Zhang, W.-C. Hong, and J. Li, "Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm," *IEEE Access*, vol. 8, pp. 14642–14658, 2020.
[29] E. Ceperic, V. Ceperic, and A. Baric, "A strategy for short-term load forecasting by support vector regression machines," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4356–4364, Nov. 2013.
[30] S. Ben Taieb and R. J. Hyndman, "A gradient boosting approach to the Kaggle load forecasting competition," *Int. J. Forecasting*, vol. 30, no. 2, pp. 382–394, Apr. 2014.
[31] M. Askari and F. Keynia, "Mid-term electricity load forecasting by a new composite method based on optimal learning MLP algorithm," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 5, pp. 845–852, Mar. 2020.
[32] V. Genre, G. Kenny, A. Meyler, and A. Timmermann, "Combining expert forecasts: Can anything beat the simple average?" *Int. J. Forecasting*, vol. 29, no. 1, pp. 108–121, Jan. 2013.
[33] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *Int. J. Forecasting*, vol. 30, no. 4, pp. 1030–1081, Oct. 2014.
[34] P. Alipour, S. Mukherjee, and R. Nateghi, "Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: A study applied to the Texas region," *Energy*, vol. 185, pp. 1143–1153, Oct. 2019.
[35] S. Wang, X. Wang, S. Wang, and D. Wang, "Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 109, pp. 470–479, Jul. 2019.
[36] H. A. Chipman, E. I. George, and R. E. McCulloch, "BART: Bayesian additive regression trees," *Ann. Appl. Statist.*, vol. 4, no. 1, pp. 266–298, Mar. 2010.



PAN ZENG (Associate Member, IEEE) received the B.E. degree in intelligent electrical building from the Qingdao University of Technology, Qingdao, China, in 2012, and the M.S. degree in software engineering from Hunan University, Changsha, China, in 2016, where he is currently pursuing the Ph.D. degree with the College of Computer Science and Electronics Engineering.

His research interests include artificial intelligence and power load forecasting.

MIN JIN (Member, IEEE) received the B.E. degree in automation and the M.S. degree in control theory and engineering from the Central South University of Technology, Changsha, China, in 1995 and 1997, respectively, and the Ph.D. degree in control theory and engineering from Central South University, Changsha, in 2000.

She was a Senior Engineer with Dongfang Electronics Information Industry Company, Yantai, China, from 2000 to 2004. From 2010 to 2012, she was a Visiting Scholar with the Department of Electrical and Computer Science, Georgia Institute of Technology, Atlanta, USA. Since 2004, she has been a Professor with the College of Computer Science and Electronics Engineering, Hunan University. Her research interests include artificial intelligence, industrial big data, and embedded systems and their engineering applications.

MD. FAZLA ELAHE received the B.E. degree in software engineering from the Shenyang University of Chemical Technology, Shenyang, China, in 2012, and the M.S. degree in computer software and theory from the Nanjing University of Science and Technology, Nanjing, China, in 2017. He is currently pursuing the Ph.D. degree with the College of Computer Science and Electronics Engineering, Hunan University.

Since 2013, he has been a Lecturer with the Department of Software Engineering, Daffodil International University, Bangladesh. His research interests include data mining, artificial intelligence, power load forecasting, and bioinformatics.