A Variable Weight Combined Model Based on Time Similarity and Particle Swarm Optimization for Short-term Power Load Forecasting

Huafeng Xian, Jinxing Che

***Abstract*—Short-term power load forecasting is an important factor affecting the security and economic operation of modern power systems. In order to avoid the limitation of a single model, this paper proposes a novel variable weight combined model that combines three single models of support vector regression (SVR), linear regression (LR) and random forest (RF). In the proposed model, a weight table is created based on time similarity, which contains weight information of 336 timestamps for determining the weight of each model. Then, particle swarm optimization (PSO) is employed to complete the modeling of the variable weight combined model. To evaluate the forecasting ability of the combined model, this paper takes the half-hourly power data of New South Wales as an example. Experimental analysis shows that all the evaluation metrics of the proposed model are better than the three single models and combined model based on average weight.**

***Index Terms*—Short-term power load forecasting, combined model, variable weight, time similarity, particle swarm optimization**

I. INTRODUCTION

Power load forecasting is an important factor that affects the security and economic operation of modern power systems. On the user side, the power demand directly determines the power production, and the high-precision power forecast can provide scientific guidance for many large-scale factories and enterprises in purchasing electricity to avoid waste of resources. As far as the power company is concerned, estimating electricity consumption in advance allows it to make reasonable arrangements for the operation of the power grid to ensure that the production cost of the power plant is minimized when the power generation and supply are reliable. Therefore, it is of great practical significance to improve the accuracy of short-term power load forecasting.

Researchers have proposed a number of methods to predict power load. Generally speaking, these methods can be classified into three major categories: traditional methods, artificial intelligence methods, and ensemble methods.

Traditional methods include simple linear regression [1], multiple linear regression [2], [3], autoregressive integrated moving average (ARIMA) [4]–[7], kalman filtering [8], etc. For example, Nepal et al. [6] propose an ARIMA model with clustering technology, which provides sufficient time for management authorities to design peak load reduction strategies. Amral et al. [2] design a linear regression forecasting system for the South Sulewesi’s Power Market. Due to a plenty of random influence factor in the fitting process, can not avoid completely producing prediction failures only through fitting parameter of linear [3], [9], [10]. Traditional ideas and tools may have been unable to meet the modern needs technically; some analysis methods of considering random factor are gradually paid attention.

Recently, the analysis methods of considering random factor have been widely applied to power load field, such as SVR [11]–[13], neural networks (NNs) [14]–[19], fuzzy logic [20]–[22], expert systems [23], etc. These methods are called artificial intelligence methods. Among them, the excellent performance of SVR has attracted much attention. SVR has a strict theoretical and mathematical foundation. The generalization ability of SVR is superior to neural networks, and the algorithm has global optimality [24]. In addition, some optimization methods are often used to help the learning of SVR. Wang et al. [25] use SVR model for annual load forecasting, its parameters are determined by the differential evolution algorithm. Yang et al. [26] propose a sequential grid method based on support vector regression (SGA-SVR). The above two works prove that their method is superior to SVR with default parameters and traditional methods. However, each method has unavoidable shortcomings and cannot always provide the desired performance. For example, neural networks are more likely to return local optimal solutions rather than global optimal solutions. The parameters of SVR have a great influence on its performance.

Ensemble methods have good robustness, including random forest (RF) [27]–[30], gradient boosting decision trees (GBDT) [31], and extreme gradient boosting (XGBoost) [32]–[34], etc. RF is based on the decision tree as learner to construct bagging integration, and further randomly selects features and randomly selects samples during the training of decision tree [35]. These methods have been used for power load forecasting and have achieved certain results. For example, Johannesen et al. [27] use random forest to predict the power demand of the Sydney region, the forecasting error of ahead half-hourly forecasting varied between 1% and 2%. Barta et al. [31] put

Manuscript received March 29, 2021; revised August 21, 2021. The work was supported by the National Natural Science Foundation of China (Grant No. 71971105) and the National Statistical Science Research Project (Grant No. 2020LZ03) and “Thousand talents plan” for high level talents in Jiangxi Province (Grant No. jxsq2019201064). The work was also supported by the Humanities and Social Sciences Projects in Jiangxi Universities (Grant No. JC20128).

Huafeng Xian is a postgraduate student of School of Information Engineering, Nanchang Institute of Technology, Nanchang 330099, Jiangxi, China (e-mail: 741419399@qq.com).

Jinxing Che is an associate professor of College of Science, Nanchang Institute of Technology, Nanchang 330099, Jiangxi, China (corresponding author, e-mail: jinxingche1@163.com).

forward a forecasting model based on GBDT, and the forecasting accuracy of the model is about 20% higher compared with a mature method. Ren et al. [36] propose an XGBoost model that integrates historical load and exogenous variables, which has smaller and smoother prediction errors compared with SVR and random forest.

So far, the above methods have been widely used to predict power load in previous studies. However, they show different prediction effects according to different data characteristics. For example, for time series, one algorithm has the optimum forecasting effect in one segment, while another algorithm can show the optimum forecasting results for other segments. Therefore, in this case, one approach to enhance performance is to use two or more prediction methods. The strategy can be called the combined method, which was first proposed by Bates and Granger [37]. The combined method has become a mainstream method in forecasting and is used by more and more scholars [38]–[42]. It can obtain higher accuracy and more reliable forecasting results. The main reasons can be summed up in two aspects [43]: one is that different methods can capture different effective information of power load data. Second, different methods can complement each other. It is worth mentioning that the accuracy of the combined forecasting is not always higher than that of the single model, but the result of combined forecasting is often more reliable than that of the single model [44]. Lahouar and Ben [28] construct 24 random forest models (one for each hour) to forecast the load on the day ahead, and the results are satisfactory. However, one drawback is that they combine multiple identical models. Researchers have further developed combined methods. To learn from each other, Moon et al. [45] present a combined model of load forecasting based on RF and multilayer perceptron (MLP). They use load data collected from a university campus in South Korea for six years as an empirical study. On the basis of ARIMA and XGBoost models, Li et al. [46] propose two new methods to predict the security level of energy supply in China. Zhang et al. [47] design a combined model to forecast the electricity consumption of the mineral company by combining deep learning and machine learning, and good results are obtained. Although all of the above studies produced better results than a single model, they used the same weights throughout the forecast period.

The key to the combined method is to solve the problem of weight coefficient. If the weight coefficient is appropriate, the combined model can get better prediction results. Conversely, the opposite result may occur. To the author's knowledge, there are few studies on the determination method of variable weight coefficient. Therefore, our goal is to propose a variable weight model that dynamically adjusts the weight coefficient over time based on the time similarity. Power load is a time series, which is affected by calendar attributes. Therefore, simple average weighted combination can not well combine the advantages of each single model. To this end, we propose a novel method for determining variable weight coefficients. The method calculates the weight of each model at a certain timestamp according to the previous prediction error, and then uses the calendar attribute of the timestamp as the index and uses the weight coefficient as the data to create a weight table, which covers the weight information of 336 timestamps in a week. When forecasting future load, weight information can be obtained from the weight table based on calendar attribute, thus realizing the variable weight prediction.

According to the above analysis, we aim to combine the advantages of various types of regression methods, such as the excellent linear fitting ability of traditional methods, the outstanding non-linear learning ability of artificial intelligence methods, and the decent robustness of ensemble methods. Therefore, our study combines SVR, linear regression and random forest to propose a combined model. Fig. 1 clearly shows the graphical abstract of this paper. The main contributions of this paper are summarized in the following aspects:

1) This paper proposes a novel variable weight combined model, which effectively combines the superiority of each single model and improves the forecasting accuracy.

2) Based on time similarity, we propose a weight table, which is used to save the weight information of 336 timestamps in a week.

3) Swarm intelligence algorithm is employed to optimize the model's hyperparameters, which gives full play to the characteristics of each model.

4) The comparison between the proposed variable weight combined model and the combined based on average weight shows that our proposed model is more effective.

This paper is organized as follows: Section II introduces the basic principles of the prediction method used in this paper; Section III introduces the combined prediction model in detail and proposes a variable weight coefficient based on time similarity; Section IV demonstrates the experiment preparations and two contrast experiments, as well as the forecasting results; Section V concludes the research and notes some recommendations for future work.

II. Methodological Approach

A. Support Vector Regression (SVR)

The function of SVR can be defined as:

$$f(x) = \boldsymbol{\omega}\varphi(x) + b \tag{1}$$

where, $\boldsymbol{\omega}$ is a weight vector; b is a constant. The objective function is:

$$\frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{n}\mathcal{L}(f(x_i) - y_i) \tag{2}$$

$$\mathcal{L}(z) = \begin{cases} 0 & \text{if } |z| \le \varepsilon \\ |z| - \varepsilon & \text{otherwise} \end{cases} \tag{3}$$

where, C is the penalty factor; $\mathcal{L}$ is the ε-insensitive loss function. To determine the coefficients $\boldsymbol{\omega}$ and b, the slack-factor ξ_i and ξ_i^* are introduced, so the objective function can be converted to:

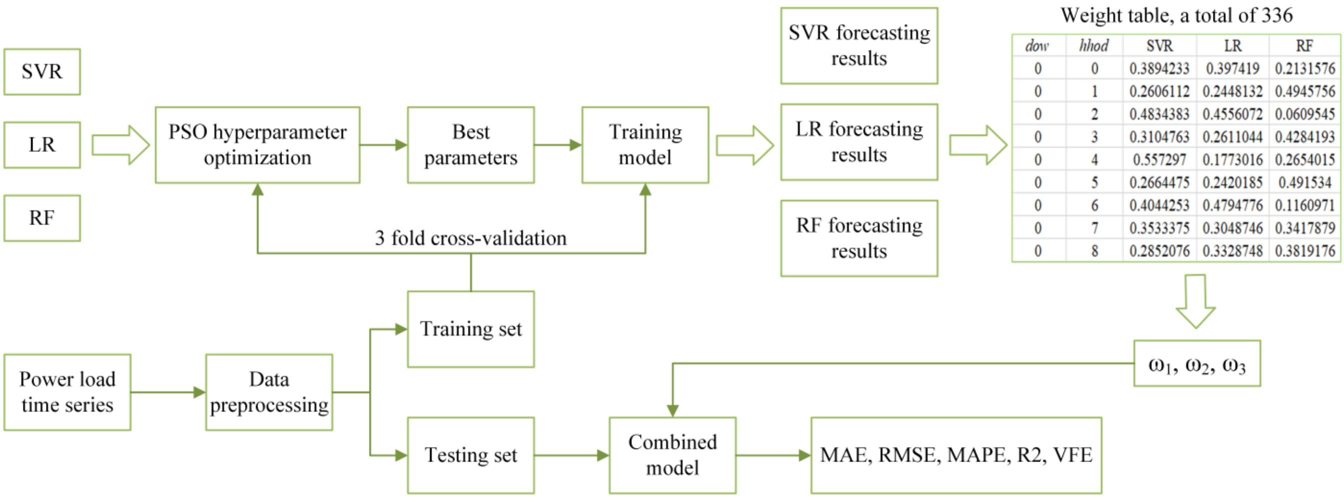


Fig. 1. Graphical abstract.

$$\frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$
$$\text{s.t.} \begin{cases} f(x_i) - y_i \le \varepsilon + \xi_i \\ y_i - f(x_i) \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, i = 1, 2, \cdots, n \end{cases} \tag{4}$$

The above is a constrained programming problem. Finally, the solution of SVR is:

$$f(x) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)K(x_i, x_j) + b \tag{5}$$

where, α_i^* and α_i are Lagrange multiplier; $K(x_i, x_j)$ is the kernel function. The radial basis function (RBF) kernel $K(x_i, x_j)=e^{-\gamma\|x_i - x_j\|^2}$ is used in this study.

B. Linear Regression (LR)

LR is a traditional method, compared with artificial intelligence methods, such as neural networks and SVR, its training time has obvious advantages. At the same time, for data with strong periodicity and smoothness, LR can get more precise result than neural networks and SVR. Its effect is similar to that of SVR using linear kernel function, but it does not require tedious iterative training process and parameter adjustment. Therefore, for power load data with strong periodicity and smoothness, the use of linear regression is a more appropriate choice than other methods.

LR is a traditional mathematical statistical method, whose mathematical expression is:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \varepsilon \tag{6}$$

The least square method is applied to solve the regression parameters, and find the regression function, that is, the prediction model. The formula is:

$$\boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{7}$$

C. Random Forest (RF)

Random forest is based on decision tree as learner to construct bagging integration, and further introduce random feature selection during the training of decision tree. The dataset M is sampled t times by bootstrap to get t sample subsets. Each decision tree is trained from one sample subset, and random feature selection is introduced during the training process, that is, k feature subsets are randomly selected from the feature set D of the current node, and the optimal feature is selected from the feature subset for divide. For regression problems, the predicted value is the average of the predicted results of all trees, as shown in following formula.

$$Y = \frac{1}{t}\sum_{i=1}^{t}Y_t \tag{8}$$

The structure of the random forest is shown in Fig. 2.

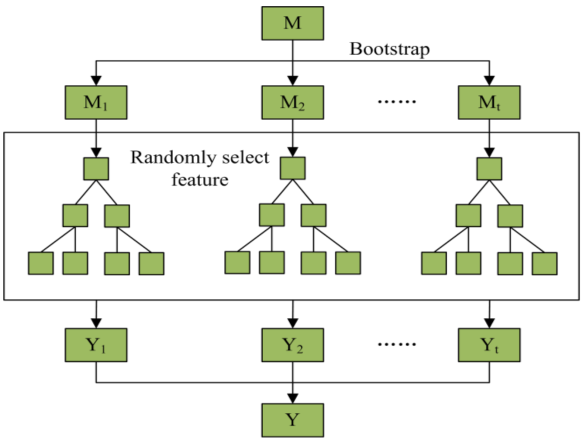


Fig. 2. Schematic diagram of random forest.

D. Particle Swarm Optimization (PSO)

PSO is a stochastic searching method based on swarm cooperation inspired by the predation behavior of birds. Its basic idea is: by simulating the predation process of a flock of birds, each bird is regarded as a particle in the PSO algorithm, which is the feasible solution of the problem. The specific process can be summarized as follows: First, a number of particles are randomly initialized in a population space, all of which have the following two properties: velocity v_i and position x_i. Then, the velocity and position are updated by the following formula:

$$v_{i+1} = \omega \times v_i + c1 \times \mathrm{rand}() \times \left(pbest_i - x_i \right) + c2 \times \mathrm{rand}() \times \left(gbest_i - x_i \right) \quad (9)$$

$$x_i = x_i + v_{i+1} \quad (10)$$

where, ω is the inertia weight; $c1$ and $c2$ are the acceleration constant; rand() is a uniformly distributed random number on the interval [0,1].

III. The Variable Weight Combined Model

A single model often has its limitations, which cannot fully reflect the variation trend of future data, so it is difficult to achieve the ideal forecasting effect. Combined model can make full use of the useful information in each model to enhance the forecasting accuracy. The prediction process of the combined model is shown in Fig. 3.

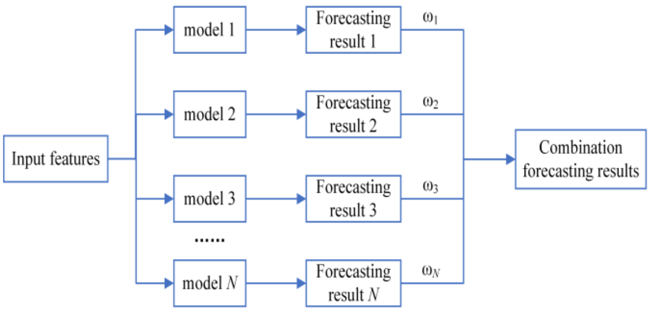


Fig. 3. Basic framework of combined model.

A. Determination of Variable Weight Coefficient

Considering the system information and variable factors provided by various forecasting models, a more systematic and comprehensive combined model is achieved by giving a certain weight to a single model. The weight coefficient determines the accuracy and reliability of the combined model. For time series, the forecasting accuracy of single model is inconsistent at different times, so the weight coefficient of single model needs to be adjusted with time. If the weight coefficient of each model is fixed, it will inevitably lead to the combined forecasting model is not scientific enough, so it is necessary to adopt variable weight combination in time series prediction. The variable weight combination can avoid the interference of a single model by random factors, and can combine the superiority of each model into a new model. Therefore, the variable weight combined model can obtain high forecasting accuracy and good reliability.

This paper uses a simple and practical method to determine the variable weight coefficient, that is, the distribution method of the reciprocal of the forecasting error. Specifically, a single model with large error is assigned a small weight coefficient; otherwise, a large weight coefficient is assigned.

Suppose the same forecasting problem, there are N single models and n timestamps (training samples), $Y(t)$ is the actual value at time t ($t = 1,2,\cdots,n$), $\hat{Y}_i(t)$ is the predicted value of the *i-th* model at time t, $\omega_i(t)$ is the weight coefficient of the *i-th* model at time t, and satisfy:

$$\sum_{i=1}^{N} \omega_i(t) = 1 \quad (t = 1,2,\cdots,n) \quad (11)$$

where, $\omega_i(t) \geq 0 (i = 1,2,\cdots,N)$. Then the variable weight combined model can be expressed as:

$$\hat{Y}(t) = \sum_{i=1}^{N} \omega_i(t) \hat{Y}_i(t) \quad (12)$$

where, $\hat{Y}(t)$ is the predicted value at time t. $\varepsilon_{it} = \left| \hat{Y}_i(t) - Y(t) \right|$ is the prediction error of the *i-th* model at time t. The weight coefficient of the *i-th* model at time t is shown in the following formula.

$$\omega_i(t) = \frac{\varepsilon_{it}^{-1}}{\sum_{i=1}^{N} \varepsilon_{it}^{-1}} \quad (13)$$

Obviously, $\sum_{i=1}^{N} \omega_i(t) = 1, \omega_i(t) \geq 0$, which satisfies the requirement of non-negativity of the weight coefficient. The expansion of $\omega_i(t)$ can be represented by a matrix.

$$\omega_i(t) = \begin{pmatrix} \omega_1(1) & \omega_2(1) & \cdots & \omega_N(1) \\ \omega_1(2) & \omega_2(2) & \cdots & \omega_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \omega_1(n) & \omega_2(n) & \cdots & \omega_N(n) \end{pmatrix} \quad (14)$$

B. Create Weight Table Based on Time Similarity

Through 3-fold cross-validation on a training set, the predicted value corresponding to different training samples is obtained for each model, and then the weight coefficient of each model on the different sample is obtained by Eq. 13. Because the samples are sampled over time, it is equivalent to obtaining the weight coefficient for each model on different timestamps. For time series with significant periodicity, we can use the calendar attribute of time series to make each single model that has different weight coefficient at different timestamp, thereby establishing a weight table, in which the weight is dynamically adjusts with the calendar attribute.

For the power load data used in this study, it is obvious that the load data shows a periodic change (see Fig. 5). Therefore, based on timestamp of load data (which day of the week, which half-hour of the day), we can establish a weight table that uses the calendar attribute as the index and the weight coefficient as the data. The weight table covers the weights of 336 time points in a week (sampling interval of 30 minutes). The algorithm description for establishing the weight table is shown in Algorithm 1. Table I shows the 5 records of the generated weight table.

When we need to forecast the power load data of a certain time point in the future, we can get the corresponding model weights from the weight table according to the *dow* and *hhod* of the time point. For example, to forecast the load at 12 o'clock on Monday, that is, *dow*=0, *hhod*=24, then the weights corresponding to *dow*=0 and *hhod*=24 in the weight table are taken out as the weight of each single model at the time point. Get the weight from the weight table as shown in Fig. 4.

Algorithm 1. Establishment of weight table.

Input: Training set $D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_n,y_n)\}$;

Single models $\mathcal{L}_1,\mathcal{L}_2,\cdots,\mathcal{L}_N$.

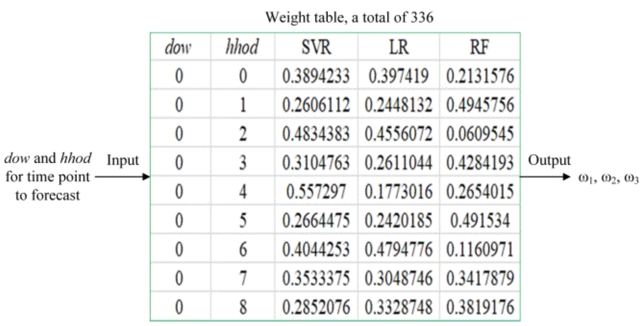
Process:

for $i=1,2,\cdots,N$ **do**
  for $t=1,2,\cdots,n$ **do**
  $\hat{Y}_i(t)=\mathcal{L}_i(D)$
  end for
end for
for $i=1,2,\cdots,N$ **do**
  for $t=1,2,\cdots,n$ **do**
  $\varepsilon_{it}=\left|\hat{Y}_i(t)\text{-}Y(t)\right|$
  end for
end for
for $i=1,2,\cdots,N$ **do**
  for $t=1,2,\cdots,n$ **do**
  $\omega_i(t)=\dfrac{\varepsilon_{it}^{-1}}{\sum_{i=1}^{N}\varepsilon_{it}^{-1}}$
  end for
end for

Output: the weight table

TABLE I
WEIGHT TABLE

dow	*hhod*	SVR	LR	RF
0	0	0.3894	0.3974	0.2132
0	1	0.2606	0.2448	0.4946
		……		
2	12	0.7749	0.0747	0.1504
2	13	0.2840	0.4363	0.2797
		……		
6	47	0.5701	0.1371	0.2928



dow	hhod	SVR	LR	RF
0	0	0.3894233	0.397419	0.2131576
0	1	0.2606112	0.2448132	0.4945756
0	2	0.4834383	0.4556072	0.0609545
0	3	0.3104763	0.2611044	0.4284193
0	4	0.557297	0.1773016	0.2654015
0	5	0.2664475	0.2420185	0.491534
0	6	0.4044253	0.4794776	0.1160971
0	7	0.3533375	0.3048746	0.3417879
0	8	0.2852076	0.3328748	0.3819176

Fig. 4. Schematic diagram of getting weight.

IV. CASE STUDY

In this section, two experiments for New South Wales data are designed to prove the predictive capacity of our proposed variable weight combined model. Before this, necessary preparations such as data exploration, data processing and evaluation metrics are also presented as the research basis.

A. Load Data Exploration

To verify the performance of the variable weight combined model based on time similarity and particle swarm optimization, the half-hour load data of May 2007 in New South Wales is used as a study case. The statistical information of the data set used in the case is listed in Table II.

The load curve is shown in Fig. 5(a). The load data has both daily and weekly cycles. Except for weekend, the daily load curve is roughly the same. Also, for a week, the load on one day of the week is more or less the same as the load on the same day of previous week and next week. These cycles are caused by daily human activities. Therefore, the power demand of New South Wales has an important relationship with the calendar attribute.

Fig. 5(b) plots the load curve of 24 hours per day in the first week of May 2007. There is a significant difference in power demand between day and night. From 20:00 in the evening, power demand begins to fall and reaches a minimum between 4:00 and 5:00 in the morning. After 5 o'clock, the demand for electricity rose rapidly as people began to move. Between 18:00 and 19:00 in the afternoon, electricity consumption reached its peak. Meanwhile, power demand on weekends is usually lower than on weekdays, but the general trend of daily demand is the same. Fig. 5(b) shows that the power demand follows a specific law every day and at some point in the day.

B. Data Preprocessing and Evaluation Metrics

In order to eliminate numerical differences and speed up learning, we need to standardize the input data. The following formulas are used to standardize the data:

$$x^*=\frac{x-\mu}{\sigma} \tag{15}$$

where, μ is the mean of the original data x, σ is the standard deviation, and x^* is the normalized data.

This study uses five metrics to evaluate the performance of the proposed variable weight model from two aspects: accuracy and stability. The mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and coefficient of determination (R2) are used to evaluate the forecasting accuracy.

$$MAE=\frac{1}{n}\sum_{t=1}^{n}\left|\hat{y}_t-y_t\right| \tag{16}$$

$$MAPE=\frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t-\hat{y}_t}{y_t}\right| \tag{17}$$

$$RMSE=\sqrt{\sum_{t=1}^{n}\frac{(y_t-\hat{y}_t)^2}{n}} \tag{18}$$

$$R2=1-\frac{\sum_{t=1}^{n}(y_t-\hat{y}_t)^2}{\sum_{t=1}^{n}(y_t-\bar{y})^2} \tag{19}$$

For stability, the variance of forecasting error (VFE) is used to evaluate the stability.

$$VFE = \frac{\sum_{t=1}^{n}(E_t - \bar{E})^2}{100n} \tag{20}$$

where, n is the number of data, y_t is the actual load at time t, and $\hat{y}_t$ is the forecasting load at time t. $E_t = y_t - \hat{y}_t$, $\bar{E} = \frac{1}{n}\sum_{t=1}^{n} E_t$.

C. Model Hyperparameters Selection

In order to make each model give full play to their characteristics, PSO algorithm is considered to optimize their hyper-parameters. The parameters of PSO are selected as follows: the population size is 10, the maximum iteration is 30, $c1 = 2$, $c2 = 2$ and the inertial weight is set to 1. The results of these models' selection are shown in Table III.

D. Experiment I: Comparative Study of Single Model and Proposed Combined Model

Experiment I is designed to verify the prediction performance of the variable weight model based on time similarity, by comparing the proposed model to its components. The predicted results are shown in Fig. 6 and

Table IV, and the detailed description of predicted results is as follows:

For one day's prediction results, the proposed model obtains almost the best values of all evaluation metrics, such as the smallest MAE and MAPE values, and the largest R2 value, which means that the proposed model obtains better prediction results than the three single models.

On average, our developed model gets the smallest errors and the largest R2: MAE, RMSE, MAPE, and R2 are 93.617, 140.937, 0.974%, and 0.987, respectively. Compared with SVR, LR and RF models, the MAE value of the combined model decreased by 14.256, 25.319 and 38.929, respectively; RMSE decreased by 32.539, 28.989 and 68.124, respectively; MAPE decreased by 0.127%, 0.285% and 0.416%, respectively.

Considering the average results, our proposed variable weight model has optimal performance. The forecasting accuracy is superior to that of its components. Our proposed model with a VFE of 174.684 is smaller than the three single models, which evaluates the forecasting stability.

Remark. Based on the prediction results and the above analysis, it can be inferred that the proposed variable weight combined model has better accuracy and stability than its components. Therefore, our proposed variable weight coefficient based on time similarity is effective.

E. Experiment II: Comparative Study of Combination Methods

In order to reflect the superiority of variable weight based on time similarity, Experiment II compares the proposed model with the combined model based on average weight. Table V and Fig. 7 show the results of five evaluation metrics for two combined models. The details are as follows:

For one day's prediction results, our model gets the best values of all evaluation metrics. On average, the values of MAE, RMSE and MAPE of the proposed model are 93.617, 140.937 and 0.974%, which decrease by 9.506, 8.275 and 0.108% compared with the average combination, respectively. The R2 value of proposed model is 0.987. Meanwhile, our proposed model's variance of forecasting error is also the best. In comparison, the combined model based on average weight also reduces the error, but the effect is not as significant as the variable weight combination. Therefore, the forecasting performance of variable weight combination outperforms the simple average weighted combination. The reason is that the average weighted combination treats its components equally, while the variable weighted combination treats its components differently by dynamically adjusting the weight coefficient of the single model over time.

Remark. The errors of the two combined models are significantly reduced compared to single models, which reflects the effectiveness of combined model in improving prediction accuracy, but our developed combined model performs better.

TABLE II
MAIN DIGITAL FEATURES OF DATASET

Dataset	Time	Number	Mean	Std.	Min.	Median	Max.
Training	May 1—May 29	1392	8628.475	1124.083	6148.720	8791.435	11051.210
Test	May 30—May 31	96	9343.533	1229.770	6671.760	9549.385	11616.780

TABLE III
HYPERPARAMETERS SELECTION RESULTS

Model	Parameter	Description	Value
SVR	$C \in [10^{-5},100]$	Penalty factor	28.47810177
	$\gamma \in [10^{-5},1]$	Kernel coefficient	0.33499965
	$\varepsilon \in [0,1]$	Allowable deviation	0
RF	*n_estimators* ∈ [10,200]	Number of regression trees	78
	max_depth ∈ [1,15]	Maximum tree depth	12
	min_samples_leaf ∈ [2,20]	Minimum sample number of leaf node	2
	min_samples_split ∈ [2,20]	Minimum sample number of node splitting	2

TABLE IV
PERFORMANCE COMPARISON BETWEEN THE PROPOSED MODEL AND ITS COMPONENTS

Date	Model	MAE	RMSE	MAPE	R2	VFE
May 30	SVR	67.448	**92.142**	0.720	0.994	**83.772**
	LR	113.007	167.619	1.234	0.980	275.964
	RF	86.515	115.738	0.969	0.990	128.110
	Proposed	**65.376**	94.571	**0.716**	**0.994**	87.341
May 31	SVR	148.298	227.371	1.483	0.966	401.495
	LR	124.866	**172.203**	1.284	0.980	280.309
	RF	178.578	272.062	1.811	0.951	578.169
	Proposed	**121.859**	175.451	**1.231**	**0.980**	**238.271**
Average	SVR	107.873	173.476	1.101	0.980	266.074
	LR	118.936	169.926	1.259	0.981	278.940
	RF	132.546	209.061	1.390	0.971	379.719
	Proposed	**93.617**	**140.937**	**0.974**	**0.987**	**174.684**

Note: The best results of five evaluation metrics are shown in bold.

TABLE V
PERFORMANCE COMPARISON BETWEEN THE PROPOSED MODEL AND AVERAGE WEIGHT COMBINATION

Date	Indices	MAE	RMSE	MAPE	R2	VFE
May 30	**Variable weight (Proposed)**	**65.376**	**94.571**	**0.716**	**0.994**	**87.341**
	Average weight	75.532	108.074	0.836	0.992	113.169
May 31	**Variable weight (Proposed)**	**121.859**	**175.451**	**1.231**	**0.980**	**238.271**
	Average weight	130.715	181.242	1.328	0.978	244.439
Average	**Variable weight (Proposed)**	**93.617**	**140.937**	**0.974**	**0.987**	**174.684**
	Average weight	103.123	149.212	1.082	0.985	191.989

Note: The best results of five evaluation metrics are shown in bold.

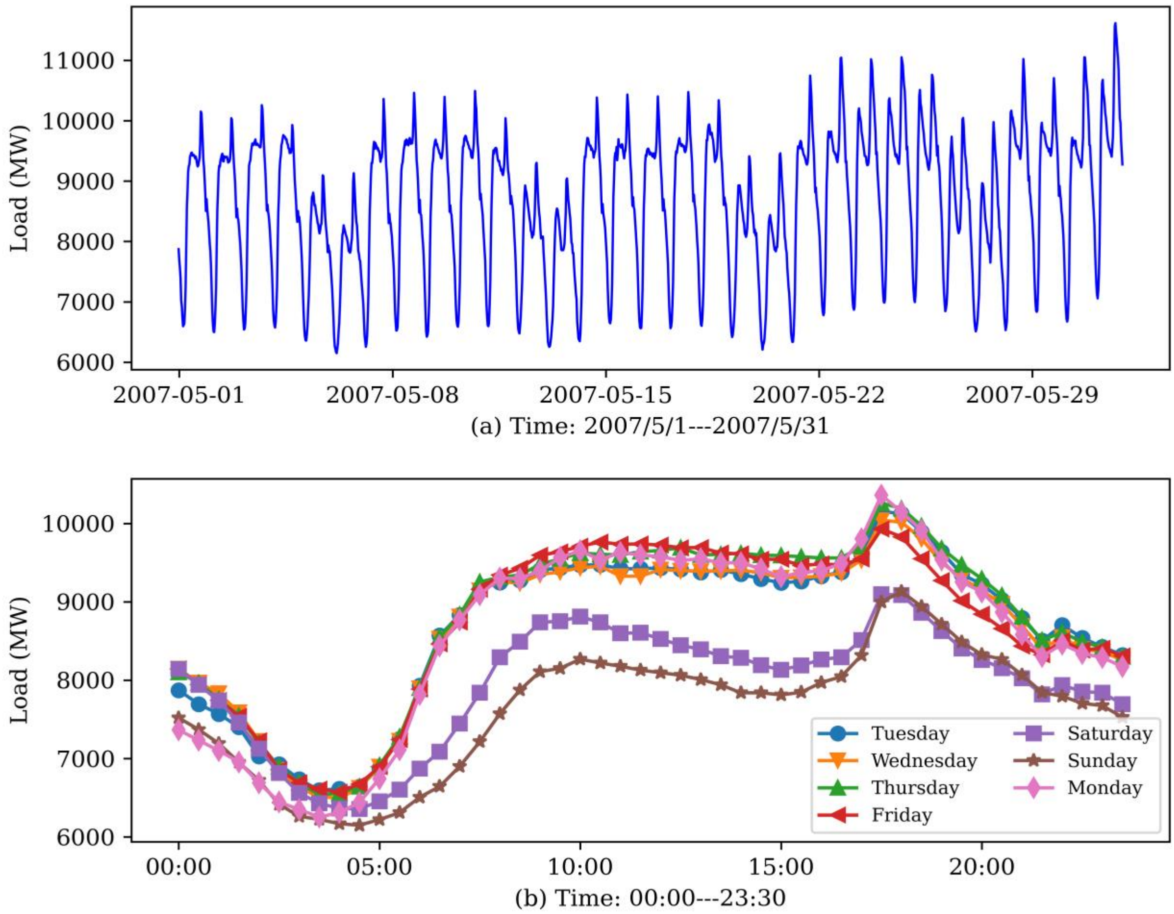


Fig. 5. (a) Power load curve in May 2007; (b) Daily power load curve of one week.

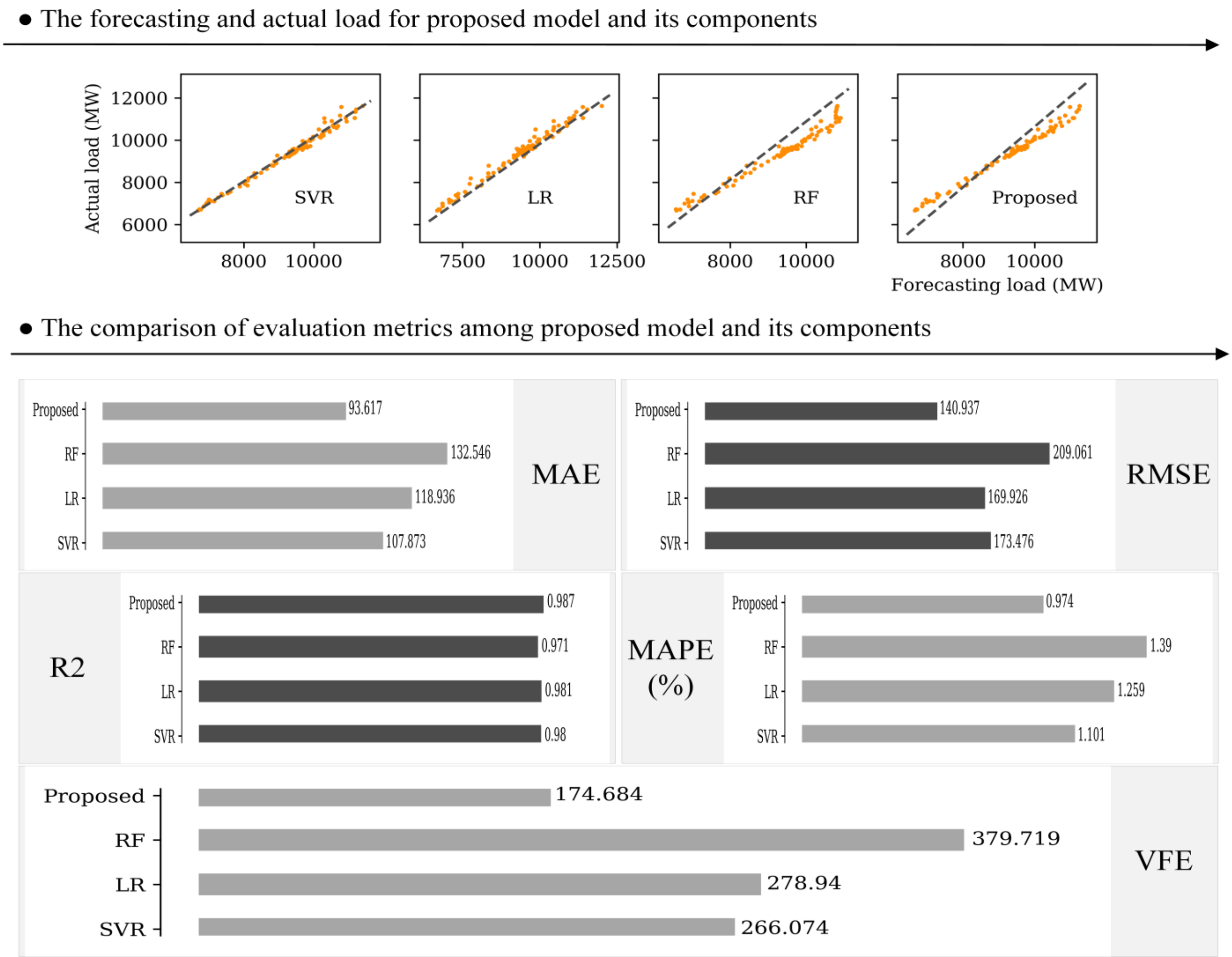


Fig. 6. Comparison between the proposed model and its components.

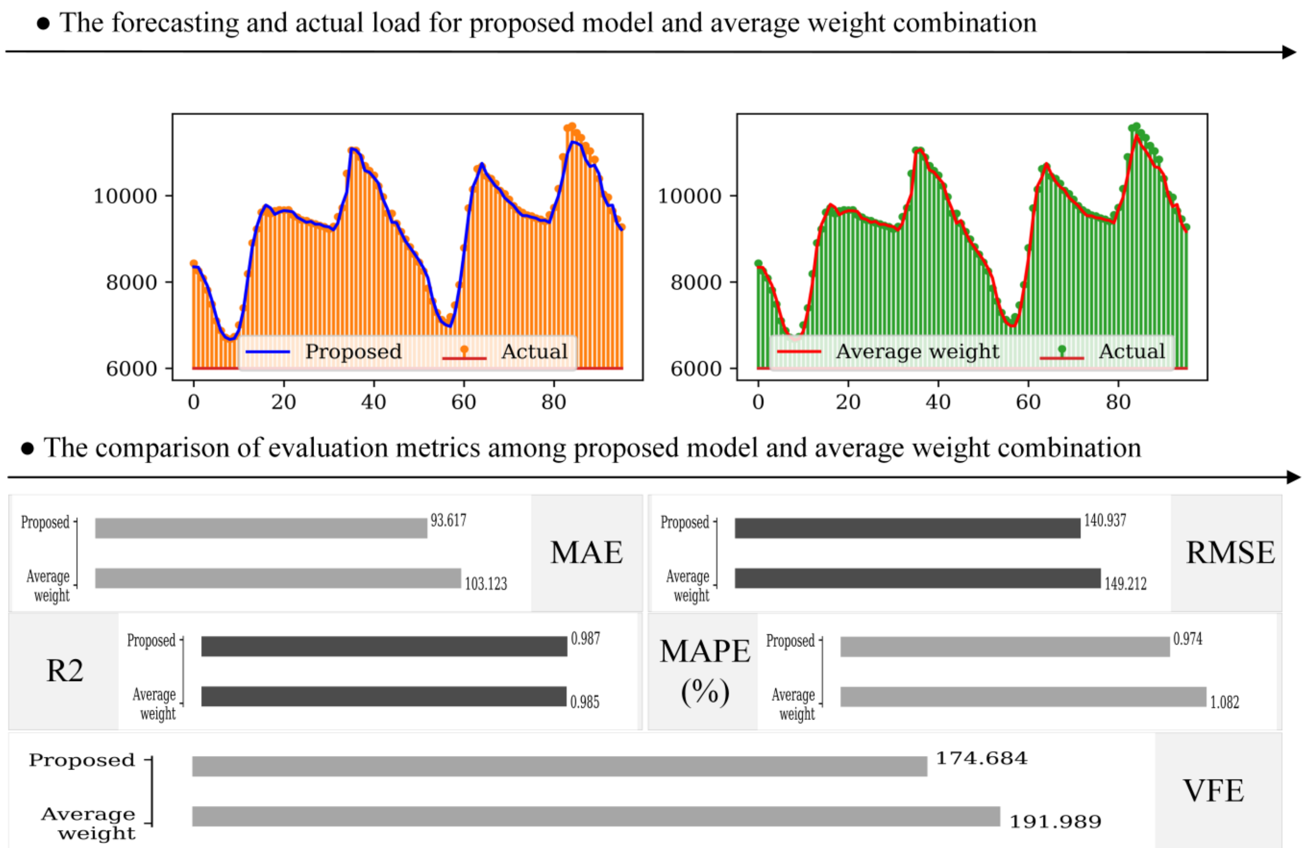


Fig. 7. Comparison between the proposed model and average weight combination.

V. Conclusion and Future Work

For short-term power load forecasting, because it is affected by many factors, a model shows different forecasting results in different time periods. In this case, one approach to enhance performance is to use two or more prediction models. To this end, this paper proposes a novel variable weight combined model that combines three single models of SVR, LR and RF. The key to the combined model is the determination of the weight coefficient. For this reason, this paper also presents a new variable weight coefficient based on time similarity. Moreover, a swarm intelligence algorithm is used to optimize the model's hyperparameters, which gives full play to the characteristics of each model. The performance of the proposed model is verified by using power load data from New South Wales. Experimental analysis shows that the variable weight combined model is superior to its components in all evaluation metrics. Meanwhile, our variable weight combination also outperforms the average weight combination.

All in all, our combined model will be a promising research direction in the future, and it can also be widely used in other prediction fields, such as residential water consumption prediction, rainfall prediction, wind speed prediction, and so on. However, there are still some problems in the combined model that we need to solve. For example, when establishing a combined model, how many forecasting algorithms can make the combined model achieve the best forecast accuracy, or which algorithms can improve the forecast accuracy. This will be the direction of our work in the future.

References

[1] A. Goia, C. May, G. Fusai, Functional clustering and linear regression for peak load forecasting, International Journal of Forecasting, vol. 26, no.4, pp700-711, 2010.
[2] N. Amral, C.S. Ozveren, D. King, Short term load forecasting using Multiple Linear Regression, in: 2007 42nd International Universities Power Engineering Conference, Brighton, pp1192-1198, 2007.
[3] A.Y. Saber, A.K.M.R. Alam, Short term load forecasting using multiple linear regression for big data, in: 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, pp1-6, 2017.
[4] W. Li, Z.G. Zhang, Based on Time Sequence of ARIMA Model in the Application of Short-Term Electricity Load Forecasting, in: 2009 International Conference on Research Challenges in Computer Science, Shanghai, pp11-14, 2009.
[5] L. Marrero, L. García-Santander, D. Carrizo, F. Ulloa, An Application of Load Forecasting Based on ARIMA Models and Particle Swarm Optimization, in: 2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, Romania, pp1-6, 2019.
[6] B. Nepal, M. Yamaha, A. Yokoe, T. Yamaji, Electricity load forecasting using clustering and ARIMA model for energy management in buildings, Japan Architectural Review, vol. 3, no.1, pp62-76, 2019.
[7] Teresita L. Todelo, and Chris Jordan G. Aliac, "Predictability of Earthquake Occurrence Using Auto Regressive Integrated Moving Average (ARIMA) Model," Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2019, 13-15 March, 2019, Hong Kong, pp252-256.
[8] H.M. Al-Hamadi, S.A. Soliman, Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model, Electric Power Systems Research, vol. 68, no.1, pp47-59, 2004.
[9] W.C. Hong, Electric load forecasting by support vector model, Applied Mathematical Modelling, vol. 33, no.5, pp2444-2454, 2009.
[10] P.A. Steenhof, W. Fulton, Factors affecting electricity generation in China: Current situation and prospects, Technological Forecasting and Social Change, vol. 74, no.5, pp663-681, 2007.
[11] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, K. Wang, Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings, Applied Energy, vol. 195, pp659-670, 2017.
[12] S. Maldonado, A. González, S. Crone, Automatic time series analysis for electric load forecasting via support vector regression, Applied Soft Computing, vol. 83, pp105616, 2019.
[13] D. Niu, Y. Wang, D.D. Wu, Power load forecasting using support vector machine and ant colony optimization, Expert Systems with Applications, vol. 37, no.3, pp2531-2539, 2010.
[14] Bagnasco, F. Fresi, M. Saviozzi, F. Silvestro, A. Vinci, Electrical consumption forecasting in hospital facilities: An application case, Energy and Buildings, vol. 103, pp261-270, 2015.
[15] M. Cai, M. Pipattanasomporn, S. Rahman, Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques, Applied Energy, vol. 236, pp1078-1088, 2019.
[16] Ghanbari, A. Naghavi, S.F. Ghaderi, M. Sabaghian, Artificial Neural Networks and regression approaches comparison for forecasting Iran's annual electricity load, in: 2009 International Conference on Power Engineering, Energy and Electrical Drives, Lisbon, pp675-679, 2009.
[17] P.H. Kuo, C.J. Huang, A High Precision Artificial Neural Networks Model for Short-Term Energy Load Forecasting, Energies, vol. 11, no.1, pp213-216, 2018.
[18] Moises Jimenez-Martinez, and Mariel Alfaro-Ponce, "Fatigue Life Prediction of Aluminum Using Artificial Neural Network," Engineering Letters, vol. 29, no.2, pp704-709, 2021.
[19] M. Ghamariadyan, M. A. Imteaz, and F. Mekanik, "Application of Wavelet Transform and Artificial Neural Networks for Predicting Rainfall Time Series with Various Variable Inputs," Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2019, 22-24 October, 2019, San Francisco, USA, pp315-319.
[20] D. Ali, M. Yohanna, M.I. Puwu, B.M. Garkida, Long-term load forecast modelling using a fuzzy logic approach, Pacific Science Review A: Natural Science and Engineering, vol. 18, no.2, pp123-127, 2016.
[21] S.A. Benson, J.K. Ogunjuyigbe, Impact of weather variables on electricity power demand forecast using fuzzy logic technique, Nigerian Journal of Technology, vol. 37, no.2, 2018.
[22] P. Mukhopadhyay, G. Mitra, S. Banerjee, G. Mukherjee, Electricity load forecasting using fuzzy logic: Short term load forecasting factoring weather parameter, in: 2017 7th International Conference on Power Systems (ICPS), Pune, pp812-819, 2017.
[23] S. Rahman, R. Bhatnagar, An expert system based algorithm for short term load forecast, IEEE Transactions on Power Systems, vol. 3, no.2, pp392-399, 1988.
[24] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, NY , USA, 1995.
[25] J. Wang, L. Li, D. Niu, Z. Tan, An annual load forecasting model based on support vector regression with differential evolution algorithm, Applied Energy, vol. 94, pp65-70, 2012.
[26] Y.L. Yang, J.X. Che, C.Z. Deng, L. Li, Sequential grid approach based support vector regression for short-term electric load forecasting, Applied Energy, vol. 238, pp1010-1021, 2019.
[27] N.J. Johannesen, M. Kolhe, M. Goodwin, Relative evaluation of regression tools for urban area electrical energy demand forecasting, Journal of Cleaner Production, vol. 218, pp555-564, 2019.
[28] Lahouar, J. Ben Hadj Slama, Day-ahead load forecast using random forest and expert input selection, Energy Conversion and Management, vol. 103, pp1040-1051, 2015.
[29] C. Li, Y. Tao, W. Ao, S. Yang, Y. Bai, Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition, Energy, vol. 165, pp1220-1227, 2018.
[30] Nan Jiang, Fei Fu, Hua Zuo, Xiuping Zheng, and Qinghe Zheng, "A Municipal PM2.5 Forecasting Method Based on Random Forest and WRF Model," Engineering Letters, vol. 28, no.2, pp312-321, 2020.
[31] G. Barta, G.B.G. Nagy, S. Kazi, T. Henk, GEFCOM 2014-Probabilistic Electricity Price Forecasting, Intelligent Decision Technologies, vol. 39, pp67-76, 2015.
[32] R.A. Abbasi, N. Javaid, M.N.J. Ghuman, Z.A. Khan, S. Ur Rehman, Amanullah, Short Term Load Forecasting Using XGBoost, in: Web, Artificial Intelligence and Network Applications, pp1120-1131, 2019.
[33] X. Liao, N. Cao, M. Li, X. Kang, Research on Short-Term Load Forecasting Using XGBoost Based on Similar Days, in: 2019

International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), pp675-678, 2019.
[34] Y. Liu, H. Luo, B. Zhao, X. Zhao, Z. Han, Short-Term Power Load Forecasting Based on Clustering and XGBoost Method, in: 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, pp536-539, 2018.
[35] L. Breiman, Random Forests, Machine Learning, vol. 45, no.1, pp5-32, 2001.
[36] L. Ren, L. Zhang, H. Wang, Q. Guo, An Extreme Gradient Boosting Algorithm for Short-Term Load Forecasting Using Power Grid Big Data, in: Proceedings of 2018 Chinese Intelligent Systems Conference, pp479-490, 2019.
[37] J.M. Bates, C.W.J. Granger, The Combination of Forecasts, Journal of the Operational Research Society, vol. 20, no.4, pp451-468, 1969.
[38] P. Jiang, Z. Liu, Variable weights combined model based on multi-objective optimization for short-term wind speed forecasting, Applied Soft Computing, vol. 82, pp105587, 2019.
[39] H.J. Sadaei, P.C. de Lima e Silva, F.G. Guimarães, M.H. Lee, Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series, Energy, vol. 175, pp365-377, 2019.
[40] L. Yang, H. Yang, A Combined ARIMA-PPR Model for Short-Term Load Forecasting, in: 2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), Chengdu, China, pp3363-3367, 2019.
[41] C. Yuan, Y. Chi, X. Li, A combined forecasting method for short term load forecasting based on random forest and artificial neural network, IOP Conference Series: Earth and Environmental Science, vol. 252, 2019.
[42] Toni Toharudin, Resa Septiani Pontoh, Rezzy Eko Caraka, Solichatus Zahroh, Antony Akbar, Bens Pardamean, and Rung Ching Chen, "Indonesia in Facing New Normal: An Evidence Hybrid Forecasting of COVID-19 Cases Using MLP, NNAR and ELM," Engineering Letters, vol. 29, no.2, pp749-758, 2021.
[43] J. Zhang, Y. Wei, D. Li, Z. Tan, J. Zhou, Short term electricity load forecasting using a hybrid model, Energy, vol. 158, pp774-781, 2018.
[44] M. Hibon, T. Evgeniou, To combine or not to combine: selecting among forecasts and their combinations, International Journal of Forecasting, vol. 21, no.1, pp15-24, 2005.
[45] J. Moon, Y. Kim, M. Son, E. Hwang, Hybrid Short-Term Load Forecasting Scheme Using Random Forest and Multilayer Perceptron, Energies, vol. 11, no.12, pp3283, 2018.
[46] P. Li, J.-S. Zhang, A New Hybrid Method for China's Energy Supply Security Forecasting Based on ARIMA and XGBoost, Energies, vol. 11, no.7, pp1687, 2018.
[47] Y. Zhang, Q. Li, A Regressive Convolution Neural Network and Support Vector Regression Model for Electricity Consumption Forecasting, in: Advances in Information and Communication, pp33-45, 2020.

Huafeng Xian is received his B. E. degree from Dongguan University of Technology in 2019. He is currently a master student in College of Information Engineering, Nanchang Institute of Technology, China. His main research interest is data analysis theory and application, as well as intelligent optimization algorithms.

Jinxing Che received his B. S. degree from Jiujiang University in 2007 and his M. S. degree in applied mathematics from Lanzhou University in 2010, as well as his Ph. D. degree in mathematical statistics from Xidian University, China in 2019. He is currently an associate professor and Master's Supervisor in College of Science, Nanchang Institute of Technology, China. His main research interest is data analysis theory and application, hydrological information processing as well as prediction theory and method.