Energy Reports 8 (2022) 12442–12455

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

An office building energy consumption forecasting model with dynamically combined residual error correction based on the optimal model

Zengxi Feng [a,b], Maoqiang Zhang [a], Na Wei [c,*], Jintong Zhao [a], Tianlun Zhang [a], Xin He [a]

[a] School of Building Services Science and Engineering, Xi'an University of Architecture and Technology, Xi'an, China
[b] Anhui key laboratory of intelligent building and building energy conservation, Anhui Jianzhu University, Hefei, China
[c] School of art, Xi'an University of Architecture and Technology, Xi'an, China

ARTICLE INFO

Article history:
Received 22 April 2022
Received in revised form 3 September 2022
Accepted 7 September 2022
Available online xxxx

Keywords:
Building energy forecasting
Combination forecasting
Residual error correction
Support vector regression
Optimal combination model

ABSTRACT

Accurate forecasting of energy consumption in office buildings is of great importance for optimal management of energy consumption and reduction of building energy consumption. A variety of combination forecasting models (FMs) have become the current research hotspots in the field of building energy consumption forecasting. For the problems of large systematic errors and poor generalization ability of existing combination FMs, this paper proposes a dynamic combination residual forecasting model (FM) with the optimal combination approach. Firstly, support vector regression (SVR) is selected as the basic FM, and the SVR residual errors are forecasted by the dynamic combination FM based on the weights, and the SVR forecast value is finally corrected. Further, the basis for the selection of the single FM in the combination model and the optimal number of combination terms are given by mathematical proof in this paper. A case study in Xi'an shows that the dynamic combined residual errors correction FM with the optimal number of terms proposed in this paper can reduce the mean absolute error (MAE) of the basic model from 1918.59 kW to 349.37 kW, the mean absolute percentage error (MAPE) from 15.80% to 2.96%, and the root mean square error (RMSE) from 2278.74 to 471.44.

1. Introduction

In recent years, China's building energy consumption has been on the rise. The total energy consumption of buildings nationwide has reached 899 million tons of standard coal, with public buildings accounting for 38.53% of the energy consumption (Chen et al., 2022a). Energy saving and emission reduction in public buildings are mainly achieved through energy-saving renovation and optimized building management (Shaikh et al., 2014; Singh and Dwivedi, 2019). Building energy consumption prediction is an important part of building management, which provides a reference for matching energy supply and demand and intelligent control of building energy systems (Parvin et al., 2021). The analysis of various influencing factors enables the forecasting of energy consumption in large buildings, which helps to reduce building energy consumption and CO2 emissions (Wang et al., 2022b).

Current building energy FMs fall into two main categories, one is a physical model based on thermodynamic calculations and the other is a data-driven model based on machine learning (Li et al., 2020). Physical models rely on software tools to simulate building energy consumption, and these models need to integrate the impact of building physical parameters and internal system operating parameters on building energy consumption (Chen et al., 2022b). The data-driven model is based on historical data and uses mathematical methods to derive hidden relationships between output and input variables (Yang et al., 2022). Compared with physical models based on simulation calculations, data-driven models have the advantage of high forecasting accuracy and fast calculation speed (Zhang, 2021). Data-driven models are mostly used when there is sufficient historical data.

In the field of building energy consumption forecasting, commonly used FM include autoregressive integrated moving average (ARIMA), multiple linear regression (MLR), random forest regression (RFR), grey model (GM), back propagation neural network (BPNN), and support vector regression (SVR) (Amasyali and El-Gohary, 2018).

The single FM proposed above performs well in different forecasting domains, but no single FM can be applied to all cases (Bourdeau et al., 2019). In order to further improve the forecasting accuracy of building energy consumption, Combination FMs that combine the advantages of various single FM

* Corresponding author.
E-mail address: weina9232022@163.com (N. Wei).

https://doi.org/10.1016/j.egyr.2022.09.022

Nomenclature

FM	forecasting model
FMs	forecasting models
SVR	support vector regression
MAE	mean absolute error
MAPE	mean absolute percentage error
RMSE	root mean square error
ARIMA	autoregressive integrated moving average
MLR	multiple linear regression
RFR	random forest regression
GM	grey model
BPNN	back propagation neural network
IEHO	improved elephant herding optimization
GA	genetic algorithm
LSTM	long short-term memory neural network
FNN	feedforward neural network
IGOW	improved grey wolf algorithm
RF	random forest
XGBoost	extreme gradient boosting
KNN	K nearest neighbours
SVR-ARIMA	The residuals are forecasted by ARIMA and corrected for SVR.
SVR-BPNN	The residuals are forecasted by BPNN and corrected for SVR.
SVR-RFR	The residuals are forecasted by RFR and corrected for SVR.
SVR-GM	The residuals are forecasted by GM and corrected for SVR.
SVR-MLR	The residuals are forecasted by MLR and corrected for SVR
SVR-4 Single FMs	Dynamically select 4 single FM combination with the smallest relative error to forecast the residuals and correct the SVR.
SVR-3 Single FMs	Dynamically select 3 single FM combination with the smallest relative error to forecast the residuals and correct the SVR.
SVR-2 Single FMs	Dynamically select 2 single FM combination with the smallest relative error to forecast the residuals and correct the SVR.
BPNN-ARIMA	The residuals are forecasted by ARIMA and corrected for BPNN
BPNN-SVR	The residuals are forecasted by SVR and corrected for BPNN
BPNN-RFR	The residuals are forecasted by RFR and corrected for BPNN
BPNN-GM	The residuals are forecasted by GM and corrected for BPNN
BPNN-MLR	The residuals are forecasted by MLR and corrected for BPNN
BPNN-4 Single FMs	Dynamically select 4 single FM combination with the smallest relative error to forecast the residuals and correct the BPNN.
BPNN-3 Single FMs	Dynamically select 3 single FM combination with the smallest relative error to forecast the residuals and correct the NPNN.
BPNN-2 Single FMs	Dynamically select 2 single FM combination with the smallest relative error to forecast the residuals and correct the SVR.

are proposed. Combination FMs can combine several single FM, which combine techniques such as data pre-processing, feature selection, and optimization algorithms to give full play to the advantages of different techniques and achieve better building energy forecasting performance (Kourentzes et al., 2019). The following three types of Combination FMs are commonly used: (1) Combination models combining optimization algorithms and a single FM. (2) Combined models based on weights. (3) Combined models based on residual error correction.

When machine learning algorithms such as SVR and ANN are used for building energy consumption forecasting, the parameters are generally selected artificially, and the selection of these parameters greatly affects the forecasting accuracy of the model. Scholars solve for the best parameters by intelligent optimization algorithms to get better forecasting performance, which is the combined model with intelligent optimization algorithm. Wang et al. (2022a) used the improved elephant herding optimization (IEHO) to optimize the weights and thresholds of BPNN, and proposed an IEHO-BPNN FM for cold load and heat load. Experimental results show that the forecasting results of this method are more accurate and have less oscillation than BPNN. Luo and Oyedele (2021) used genetic algorithm (GA) to select the optimal architecture for long short-term memory (LSTM) neural network to improve its forecasting accuracy and robustness, and tested the performance of the proposed forecasting model by two educational buildings. GA-LSTM neural network outperforms feedforward neural network (FNN) and LSTM neural network FMs in terms of forecasting performance. Cheng et al. (2022) proposed a short-term hybrid FM with improved grey wolf algorithm optimized support vector regression (IGOW-SVR) and applied the model to ice storage air conditioning load forecasting. Experimental results show that the proposed model has higher forecasting accuracy, shorter running time, and stronger robustness than neural networks in the case of small samples.

Combination models based on weight are an important research branch in the field of forecasting. The theory of combination forecasting was first proposed by Bates and Granger in 1969 (Bates and Granger, 1969). In a combination FM, single FMs usually contain only partial information of the forecasting object. Using multiple single FMs to forecast the same object and combining them by assigning different weights based on the forecasting results can improve the forecasting accuracy of the system by including more comprehensive information. Fan et al. (2014) used eight popular FMs to forecast building energy consumption and optimized the weights of the eight forecasting models using GA to finally construct a combined FM. This method was used to forecast a large public building in Hong Kong. The results showed that the mean absolute percentage error (MAPE) of the combined model was 2.32%, which was lower than that of every single FM. Wang et al. (2020) proposed a novel combined FM for building energy consumption. In their study, random forest (RF), extreme gradient boosting (XGBoost), SVR, and K nearest neighbours (KNN) models were selected as single FM for the combination. The forecasting results indicated that the combined model outperformed the single FM in terms of forecasting accuracy, generalization ability, and robustness. Chou and Bui (2014) used five single models to form an equal-weight combination FM for forecasting heat and cooling loads in the building design phase. The results demonstrate the validity, and accuracy of the combined model for the forecasting of cold and thermal loads.

In addition to the above mentioned combination FMs, there is another combination FM based on residual error correction in the literature, whose structure takes into account the residual values obtained from the basic FM. Assuming that the residual value can be accurately forecasted, the accuracy of the forecasting model is improved by correcting the forecast value of the basic model with this residual value. The importance of this technique is that it can reduce to some extent the negative impact on the forecasting accuracy of systematic errors, which imply that the forecast values are too high or too low most of the time and can be encountered in all FMs. Karthika et al. (2017) used a combined ARIMA-SVM model to forecast the hourly electricity load demand. The demand was first forecasted using the ARIMA model, and then its residual was forecasted by SVR and the forecasting results of the ARIMA model were revised. Using the historical load data of a public building in the south from 2014–2015 as the research object can be obtained, the MAPE of ARIMA was significantly reduced after the residual correction. Li and Li (2017) used GM as the basic forecasting model to forecast energy data in Shandong province in China. Then, the GM residuals were forecasted using the ARIMA model, and finally, the forecasting accuracy of the original forecast was further improved after residual correction. By comparing the average relative errors of different models, the results show that the combined GM-ARIMA model has higher accuracy than individual models. Feng et al. (2021) proposed an EVCS load forecasting method based on a combination of multivariate residual correction GM and LSTM network. The final forecasted load of EVCS is obtained by summing the forecasted values of GM and the forecasted residuals of LSTM. The effectiveness of the proposed method in the literature is verified through experiments and simulations.

These three combination models have different combination principles and model frameworks, but their forecasting accuracy is better than that of the single forecasting models. However, these combined models also have some Shortcomings: (1) The combined model combining optimization algorithm and single FM has high forecasting accuracy, but its generalization ability is poor. Their forecasting accuracy may be less satisfactory if a different study object is used; (2) For the combination model based on weight, the existing literature focuses on how to solve the weights of every single FM. There is no solution to the problems of how to select FM and how many single FM to select for combination; (3) The combined model based on residual correction can reduce the impact of systematic error on the forecasting accuracy, but the residual sequence has strong randomness, and it is still a difficult problem to forecast the residual accurately;

In summary, the forecasting performance of the combined forecasting model still has much space for improvement. How to improve the generalization ability and forecasting accuracy of the combinatorial model is an urgent problem to be solved.

In view of the above problems, this paper combines the advantages of the combined FM and the residual correction FM, and proposed an FM based on the dynamic combined residual correction. First, an FM is selected as the basic FM to forecast building energy consumption, the residual series is obtained by subtracting the actual values from the forecasted values. In view of the strong randomness of the residual series, a combined FM is introduced to improve the accuracy of the residual errors forecasting. Finally, the forecasted value of the basic model is corrected by residual errors. Further, in the combined residual error FM, the two problems of how to select a single model and how many single models to select for combination are mathematically derived, and the basis for selecting a single model and the optimal number of single models are obtained. The method proposed in this paper can effectively improve the forecasting accuracy and generalization ability through an example of energy consumption forecasting in a large office building in Xi'an.

The rest of the paper is organized as follows. Section 2 contains the method. Section 3 contains the case analysis. Section 4 contains the result and discussion. Section 5 contains the conclusions

2. Methods

2.1. Model framework

Fig. 1 shows the model framework. First, an FM is selected as the basic model to forecast building energy consumption. The residual errors of the basic model are forecasted by the combined residual errors FM, and the combined residual errors forecasted value is obtained. By adding the combined residual errors forecasted value with the basic model forecasted value, the final corrected forecast value of the basic model is closer to the actual value. The accuracy of the forecasting of the basic model will be improved. In the combined residual errors forecasting, the selection basis of the single FM and the optimal number of single FM are proved in Section 2.4.

2.2. The single forecasting model

Different forecasting models have different characteristics and applications. For example, the GM directly capitalizes on accumulated original data to identify the rules of a system and then builds exponential models without consideration of the system structure. Moreover, different GMs can be constructed according to the features of the original data. MLR is built using regression analysis to determine the relations between the forecasting object and the relevant influencing factors (Dhaval and Deshpande, 2020). The ARIMA model with a simple structure and unitary data is used to forecast the next value of a series (Bento et al., 2021). RFR makes it easy to get nonlinear relationships in a dataset. However, the forecasting accuracy is low and not suitable for small data sets (Dong et al., 2021). Despite the substantial adaptability and strong learning capacity, the forecasting accuracy of BPNN models still needs to be improved because of a large number of training samples and the difficulty in finding an optimal network structure (Singh and Dwivedi, 2019). Compared with BPNN, SVR is a supervised machine learning method based on statistical theory with the ability to transform nonlinear relationships into linear relationship (Zhang et al., 2016). Since the law of energy consumption change has strong non-linearity and uncertainty, using SVR to forecasting building energy consumption can better solve this problem (Ma et al., 2019). SVR has become a research hotspot in the field of building energy consumption forecasting because of its good fitting ability to the nonlinear characteristics of building energy consumption and the small amount of operations (Moradzadeh et al., 2020; Panahi et al., 2020). Therefore, SVR is used as the basic model for building energy consumption forecasting in this paper where ARIMA, RFR, MLR, GM, BPNN are the single FMs in the combined residual FM.

The principles of the single FMs used in this paper are briefly described in this section, and the details of the application are described in Section 3.

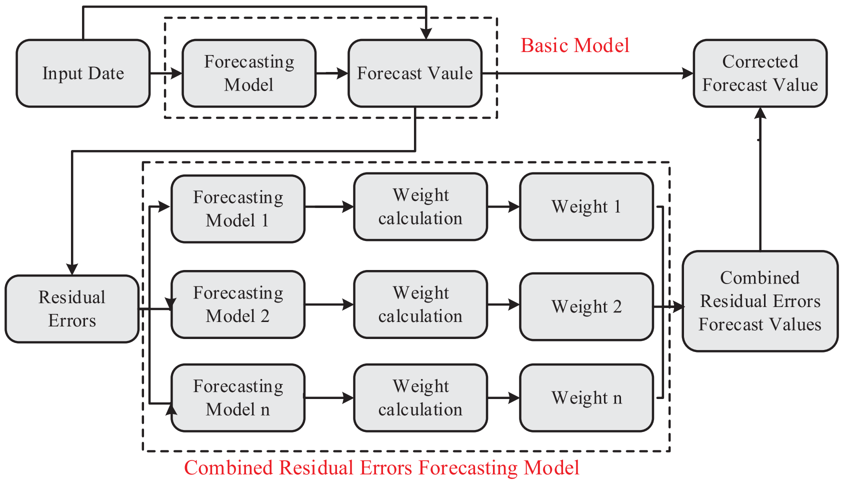

Fig. 1. Combined residual errors correction forecasting model framework.

2.2.1. SVR

Support vector machine (SVM) is a powerful supervised machine learning method that can effectively deal with nonlinear classification and regression problems. It can efficiently handle high-dimensional spatial problems. When using the SVM algorithm for regression problems, it is called SVR. Based on the training set data, the relationship function between the approximating variables is:

$$f\left(x\right)=\left\langle w,\varphi\left(x\right)\right\rangle+b \tag{1}$$

where x is the input vector; w is the weight coefficient; $\varphi(x)$ is the mapping function; b is the bias; where $f(x)$ denotes the forecasting outputs. The SVR algorithm ensures the smoothness of the functional relationship by minimizing the sum of squared weight coefficients in the process of approximating the functional relationship. An error less than ε is tolerated to improve the generalization performance of the model. Therefore, w and b are determined by solving the following quadratic convex programming problem.

$$\min_{w,b,\xi_i,\xi_i^*}\frac{1}{2}\left\|w\right\|^2+C\sum_{i=1}^{n}\left(\xi_i+\xi_i^*\right) \tag{2}$$

$$s.t\begin{cases}y_i-\left\langle w,\varphi\left(x\right)\right\rangle-b\le\varepsilon+\xi_i\\ \left\langle w,\varphi\left(x\right)\right\rangle+b-y_i\le\varepsilon+\xi_i^*\\ \xi_i,\xi_i^*\ge 0, i=1,\ldots,n\end{cases} \tag{3}$$

where: ξ_i, ξ_i^* are the Slack variables; C is the penalty factor, $C > 0$; y_i is the output corresponding to the ith sample; ε is the tolerance error. The linear separable plane is constructed and solved by introducing an implicit kernel function instead of $\varphi(x)$, which maps the nonlinear problem to a higher dimensional space. To solve the nonlinear correlation between building energy consumption and impact factors linear correlation, the kernel function commonly used to deal with nonlinear problems is chosen as Radial Basis Function (RBF).

$$k_{RBF}\left(a',a\right)=e^{\frac{\left\|a'-a\right\|^2}{2\sigma^2}} \tag{4}$$

where: a', a are two low-dimensional vectors. $\frac{1}{2\sigma^2}$, also known as the γ parameter, reflects the degree of separation of the mapping. The C, ε, and γ parameters are called model hyperparameters, which are constant during the model training process, and adjusting the model hyperparameters can change the model performance.

Chen and Tan (2017) proposed a new SVR model that selects the ambient temperature in the next 2 h as the actual input variable for short-term electricity load forecasting. This innovation improves the prediction accuracy by reducing the lagging effect of thermal inertia inside buildings on weather-sensitive loads. Houchati et al. (2022) used indoor temperature, humidity, and solar radiation intensity as input variables and building energy consumption as output variables to train the SVR model. The forecasting results show that the model can effectively improve the forecasting accuracy. In this paper, outdoor dry bulb temperature, relative humidity, wind speed and solar radiation intensity are used as input variables and building energy consumption is used as output variable to train SVR.

2.2.2. ARIMA

The basic idea of the ARIMA model is to describe the development of a time series using its own lag series and random disturbance terms and their lag series. ARIMA(p, d, q) contains moving average process (MA), autoregressive process (AR), autoregressive moving average process (ARMA) and an autoregressive integrated moving average model (ARIMA). The general form of the model is:

$$y_i=\mu+\sum_{i=1}^{p}\gamma_i y_{t-i}+\sum_{i=1}^{q}\theta_i\epsilon_{t-i}+\epsilon_t \tag{5}$$

Where: μ is the constant coefficient. p is the autoregressive order. q is the moving average order. ϵ_t is the random error, which is usually a white noise series and conforms to the normal distribution. γ_i and θ_i are the parameters to be solved. For the non-stationary series, the d-order difference is first applied to obtain the stationary series and then modelled. Ozturk and Ozturk (2018) used coal, oil, gas, renewable energy and total energy consumption data from 1970–2015 to forecast Turkey's energy consumption for the next 25 years through ARIMA.

2.2.3. MLR

In the study of the linear relationship between the independent variables $x_1, x_1, \ldots, x_n$ and the dependent variable y, the multiple linear regression model established is:

$$y=\beta_0+\beta_1x_1+\cdots+\beta_nx_n+\varepsilon \tag{6}$$

β_0 is the constant, $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficient; ε is the random error term, which indicates the part not determined by the independent variable. Substitute the historical sample data into the above equation and use the least squares method to find the estimated value ($\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_n$) of ($\beta_0, \beta_1, \ldots, \beta_n$). The regression equation was obtained as:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_n x_n \tag{7}$$

where $\hat{y}$ is the output value. One drawback of the multiple linear regression model is overfitting, which leads to poor generalization of the model. Catalina et al. (2013) proposed an MLR model to forecast heating energy demand, where the input variables are the building heat loss coefficient, equivalent surface area, and the difference between indoor and outdoor temperatures. Experimental results show that the model has a high accuracy.

2.2.4. RFR

The random forest model is a combinatorial model based on decision trees. If the dependent variable is a categorical variable, a categorical discriminant model is built. If the dependent variable is a continuous variable, a multiple nonlinear regression model is built. In this paper, the response variable is a continuous variable, so a regression forecasting model should be constructed. The RFR model is built as follows: firstly, N samples are randomly selected by using the idea of Bagging with release. Then, m variables are randomly selected at each node ($m < N$, m is the total number of independent variables in the training set). Then, a single decision tree is constructed by using them as the candidate variables to split the node. Repeat the above steps to generate a mass regression decision tree. The final forecasting result of the model is the average of the forecasting results of the mass regression decision tree.

In the model building process, the principle of calculating the variable selection at the nodes of the decision tree is the minimum of the mean squared deviation. That is, for an arbitrary division of variable A, the corresponding arbitrary division points s are divided on both sides of the data set D_1 and D_2. The corresponding variables and variable values division points are found such that the mean squared deviation of the respective sets of D_1 and D_2 is minimized, and the sum of the mean squared deviations of D_1 and D_2 is minimized. The expression is:

$$\underbrace{min}_{A,s}\left[\underbrace{min}_{c_1}\sum_{x_i\in D_1(A,s)}(y_i-c_1)^2+\underbrace{min}_{c_2}\sum_{x_i\in D_{2(A,s)}}(y_i-c_2)^2\right] \tag{8}$$

where x_i is the feature factor; y_i is the sample true value; c_1 is the sample output mean of D_1; c_2 is the sample output mean of D_2.

2.2.5. GM

Grey system theory is a system science theory that studies small samples and poor information uncertainty problems. By mining part of the known information, valuable information is extracted to achieve the description and monitoring of the evolutionary law of the system. Grey prediction is an important part of grey system theory and is mainly used for the prediction of grey uncertainty problems. The grey GM (1, 1) model is the basic model of grey prediction technique, and its model principle is as follows.

The original feature data sequence is $x^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\right\}$, A single accumulation is performed to generate a new data sequence: $x^{(1)} = \left\{x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\right\}$.

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1, 2, \ldots, n \tag{9}$$

$$z^{(1)}(k) = \frac{1}{2}\left(x^{(1)}(k) + x^{(1)}(k-1)\right), k = 2, 3, \ldots, n \tag{10}$$

The sequence $z^{(1)} = \left\{z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(n)\right\}$ is obtained by the above equation.

The mean value of the GM (1, 1) model is $x^{(1)}(0) + az^{(1)}(k) = b$. A first-order single-variable differential equation was fitted to generate the series, and the whitened differential equation was obtained as:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{11}$$

where the parameters a and b are the development coefficient and the amount of ash action of the GM (1, 1) model, respectively. The development coefficient reflects the trend of the predicted value, and the amount of grey action reveals the intrinsic changes of the original data. The parameters a and b can be obtained by using the least squares method.

The data prediction model in whitened form is represented as:

$$\hat{x}^{(1)}(k) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-a(k-1)} + \frac{b}{a}, k = 1, 2, \ldots, n \tag{12}$$

The resulting prediction for the serial data is:

$$\hat{x}^{(0)}(k) = \left(1 - e^{a}\right)\{x^{(0)}(1) - \frac{b}{a}\}e^{-a(k-1)} \tag{13}$$

Where, $\hat{x}^{(0)}(k)$ is the reduced value sequence of the original data. When $k > n$, $\hat{x}^{(0)}(k)$ is the predicted value of the model.

2.2.6. BPNN

BPNN was proposed in 1986 and is one of the more widely used neural network models at this stage. The number of nodes in the input and output of BPNN is determined according to the data input category and the expected output category, respectively. The number of nodes in the hidden layer can be determined according to the empirical formula, and each layer is connected by weights, and each node in the hidden layer and output layer has a threshold value. The forward transfer process of the BPNN model is performed by the following equation.

$$\begin{cases} S_j = \sum_{j=1}^{L} w_{ij} + b_j \\ a_j = f\left(s_j\right) \\ P_k = \sum_{k=1}^{N} w_{ij}a_j + b_k \\ t_k = f\left(P_k\right) \end{cases} \tag{14}$$

Where x is the BPNN input value; w_{ij} is the weight value from the input layer to the hidden layer; b_j and b_k are the thresholds of the hidden layer and the output layer, respectively.s_j and a_j are the input and output values of the hidden layer respectively; w_{jk} is the weight value between the hidden layer and the output layer; p_k and t_k are the input and output values of the output layer; $f(x)$ is the neural network transfer function. The formulas of the error function for error back propagation, the weights and threshold correction factors are:

$$\begin{cases} f_i = \sum_{k=1}^{N}(y_k - \overline{y}_k)^2 \\ \Delta w_{jk} = -\eta\frac{\partial E_p}{\partial w_{jk}} \\ \Delta w_{ij} = -\eta\frac{\partial E_p}{\partial w_{ij}} \\ \Delta b_k = -\eta\frac{\partial E_p}{\partial b_k} \\ \Delta b_j = -\eta\frac{\partial E_p}{\partial b_j} \end{cases} \tag{15}$$

f_i is the error function equation; N is the number of sample groups; y_k is the true value; $\bar{y}_k$ is the BPNN output value; η is the learning rate; Δw_{ij} and Δw_{jk} are the correction coefficients; Δb_k and Δb_j are the threshold correction coefficients; E_p is the node error.

2.3. The combined FM for residual error forecasting

SVR is used as the basic model for building energy consumption forecasting. Assuming that the forecasted value of a basic model at time point τ is Q_{basic} and the actual value is Q_{act}, the residual error Q_e, can be written as:

$$Q_e = Q_{basic} - Q_{act} \tag{16}$$

According to Eq. (16), if the Q_e, at τ can be forecast by the residual error FM, then Q_{basic} can be revised. Thus, the Q_e at τ is a forecasting problem (Yao et al., 2006).

Many single FMs can be used to construct a residual error FM. However, every single FM has its own character and complex practice, which makes the forecasting accuracy of Q_e unsteady. To improve the forecasting accuracy of Q_e, two FMs, a and b, are used to establish a residual error model. These models are simultaneously employed to establish a combined FM and correct the residual error.

For a certain forecasting problem, assume the residual error in period t is $Q_{e,t}$ $(t = 1, 2, \ldots, n)$ and the forecasted values of models a and model b are $Q_{e,a}$ and $Q_{e,b}$, respectively. Suppose the weight vector is $W = [\omega_a, \omega_b]^T$; then, the residual error combination FM can be expressed as follows:

$$Q_{e,t,combine} = \omega_a Q_{e,a} + \omega_b Q_{e,b} \tag{17}$$

Where $Q_{e,a}$ and $Q_{e,b}$ are the forecasting values of the residual error from model a and model b, respectively, ω_a and ω_b are the combination weights of the two FMs. Meanwhile, the following constraint equations exist: $\omega_a + \omega_b = 1.0$, $\omega_a \geq 0$, and $\omega_b \geq 0$.

Various methods can be used to calculate the combination weights. To dynamically optimize the combination weights according to historical data, the least error square sum (LESS) is used to determine the weights in the combined FM. The equivalent equations can be expressed as:

$$\begin{cases} \min Z = \sum_{i=1}^{m} \left(Q_{e,t,combine}(i) - Q_{e,t}(i)\right)^2 \\ s.t \\ \omega_a + \omega_b = 1.0 \\ \omega_a \geq 0, \omega_b \geq 0 \end{cases} \tag{18}$$

Suppose that the forecasting deviations of single models a and b for residual error forecasting are given as follows:

$$\varepsilon_a = Q_{e,a} - Q_{e,t} \tag{19}$$

$$\varepsilon_b = Q_{e,b} - Q_{e,t} \tag{20}$$

The deviation of the combined FM, $\varepsilon_{combine}$, can be obtained as:

$$\varepsilon_{combine} = Q_{e,t,combine} - Q_{e,t} = \varepsilon_a \omega_a + \varepsilon_b \omega_b \tag{21}$$

Thus, Eq. (21) can be written in matrix form:

$$\begin{cases} \min Z = W^T H W \\ s.t. \\ e^T W = 1.0 \\ W^T \geq 0 \end{cases} \tag{22}$$

Where $W = [\omega_a, \omega_b]^T$, $e = [1, 1]^T$, and $H = \begin{bmatrix} \varepsilon^2{}_a & \varepsilon_a \varepsilon_b \\ \varepsilon_a \varepsilon_b & \varepsilon^2{}_b \end{bmatrix}$.

The forecasting error sequences of the residual error FMs of a and b are $Q_{e,a}$ and $Q_{e,b}$, respectively, defined as Eq. (23) and Eq. (24).

$$\varepsilon_a = \left(\varepsilon_{a,1}, \varepsilon_{a,2}, \cdots, \varepsilon_{a,n}\right) \tag{23}$$

$$\varepsilon_b = \left(\varepsilon_{b,1}, \varepsilon_{b,2}, \cdots, \varepsilon_{b,n}\right) \tag{24}$$

Because $Q_{e,a}$ and $Q_{e,b}$ are independent, the covariance between $Q_{e,a}$ and $Q_{e,b}$ is equal to zero according to statistical theory (Clemen, 1989). That is, Eq. (25) can be obtained as follows:

$$Cov\left(\varepsilon_b, \varepsilon_a\right) = E\left\{\left[\varepsilon_b - E\left(\varepsilon_b\right)\right]\left[\varepsilon_a - E\left(\varepsilon_a\right)\right]\right\} = 0 \tag{25}$$

For stationary series, their deviation series is white noise, and the mathematical expectation is equal to zero (Zhang, 1991). Therefore, Eq. (26) can be obtained as follows:

$$E\left(\varepsilon_b\right) - E\left(\varepsilon_a\right) = 0 \tag{26}$$

By combining Eqs. (25) and (26), the following results can be obtained:

$$E\left(\varepsilon_b \varepsilon_a\right) = Cov\left(\varepsilon_b, \varepsilon_a\right) = \frac{\left(\sum_{i=1}^{n} \varepsilon_{b,i} \varepsilon_{a,i}\right)}{n} = 0 \tag{27}$$

Therefore, means

$$\sum_{i=1}^{n} \varepsilon_{b,i} \varepsilon_{a,i} = 0 \tag{28}$$

The Lagrange function is introduced to solve Equation (22), as shown in Eq. (29):

$$L = W^T H W + \lambda\left(e^T W - 1\right) \tag{29}$$

Under the Kuhn–Tucker condition, Eq. (30) can be obtained by solving Equation (29):

$$\begin{cases} \dfrac{\partial L}{\partial W} = 2HW + \lambda e = 0 \\ \dfrac{\partial L}{\partial \lambda} = e^T W - 1 = 0 \end{cases} \tag{30}$$

The weights ω_a and ω_b in the combined FM can be calculated as:

$$W = [\omega_a, \omega_b]^T = \frac{\left(H^{-1} e\right)}{\left(e^T H^{-1} e\right)} \tag{31}$$

According to Eqs. (17) and (31), the combined model can be built to revise the forecasting result of the basic SVR model.

2.4. Forecasting error of the corrected basic model

Here, proofs of the accuracy improvement of the combination of two different forecasting models are given. Denote the relative forecasting error of model a as δE_a and the relative forecasting error of model b as δE_b, The combined relative forecasting error, δE_z, can be expressed as:

$$\delta E_z = \omega_a \delta E_a + \omega_b \delta E_b \tag{32}$$

Denote the forecasted value of the basic model at time τ as $Q_{basic,t}$, and the actual value as Q_{act}. The relative forecasting error of the basic model $\delta E_{basic,act}$, can be written as:

$$\delta E_{basic,act} = \frac{\left|\Delta E_{basic,\mathrm{act}}\right|}{Q_{act}} \tag{33}$$

Where $\Delta E_{basic,act}$ is the actual residual error of the basic model at time τ. $\Delta E_{basic,act} = Q_{basic,f} - Q_{act}$

When the combined FM is used to forecast the actual residual error, $\Delta E_{basic,act}$, the relative forecasting error δE_z, can be written as:

$$\delta E_z = \frac{\left|\Delta E_z - \Delta E_{basic,act}\right|}{\left|\Delta E_{basic,act}\right|} \tag{34}$$

Where ΔE_z is the forecast residual error of the basic model for the combined FM.

When ΔE_z is used to correct the forecasted value of the basic model, the ultimate residual error of the basic model can be expressed as:

$$\Delta E_{cor} = \Delta E_{basic,act} - \Delta E_z \tag{35}$$

Thus, the relative forecasting error of basic model with combined model correction δE^*_{basic}, can be written as:

$$\delta E^*_{basic} = \frac{\left|\Delta E_{basic,act} - \Delta E_z\right|}{Q_{act}} \tag{36}$$

According to Eqs. (33), (34) and (36), the following equation can be obtained:

$$\delta E^*_{basic} = \frac{\delta E_z \left|\Delta E_{basic,act}\right|}{Q_{act}} = \delta E_z \delta E_{basic} \tag{37}$$

In accordance with combination theory, Eq. (23) exists:

$$\max\left(\delta E_a, \delta E_b\right) \geq \delta E_z \geq \min\left(\delta E_a, \delta E_b\right) \tag{38}$$

Eqs. (37) and (38) define the range of δE^*_{basic} as shown in Eq. (39):

$$\max\left(\delta E_a, \delta E_b\right) \delta E_{basic} \geq \delta E^*_{basic} \geq \min\left(\delta E_a, \delta E_b\right) \delta E_{basic} \tag{39}$$

Assume the following conditions: $\delta E_a < 1$, $\delta E_b < 1$

Then, Eq. (40) can be deduced from Eqs. (37) and (38):

$$\delta E^*_{basic} < \delta E_{basic} \tag{40}$$

Eq. (40) shows that the forecasting accuracy of the basic model is improved by the residual error correction in the combined FMs (Andrawis et al., 2011; Chan et al., 2010; Martins and Werner, 2012). Eqs. (37), (38) and (40) indicate that the smaller the relative error of the individual FM, the smaller the relative error of the combined FM. The forecasting accuracy of the basic model corrected by the combined residual error FM was improved.

2.5. Dynamical adjustment of single FM for combined residual forecasting model

According to the above derivation, the forecasting accuracy can be improved after the modification of the residual error combination forecasting. Different FMs, however, have diverse characteristics and application scopes. In addition, when making forecasting with different types of FMs, no model is likely to reduce the forecasting error to zero, and the forecasting accuracy will be unstable due to the uncertainty and randomness of the various factors affecting the system. Models of different types often provide a variety of useful information. If a model is abandoned because of its poor forecasting performance, some useful information will be lost. Therefore, in the case of using the same number of single FMs for combination forecasting, how to select the single FMs to improve the forecasting accuracy will be demonstrated below.

2.5.1. Same number of options for combination forecasting

We take two single FMs as an example to investigate residual error combination FMs with the same number of single FMs.

Four types of FMs are assumed to exist within some time period, represented by 1, 2, 3 and 4. Their relative forecasting errors are δE_1, δE_2, δE_3 and δE_4, respectively, which satisfy the following conditions: $\delta E_1 \leq \delta E_3$, $\delta E_1 \leq \delta E_4$, $\delta E_2 \leq \delta E_3$ and $\delta E_2 \leq \delta E_4$.

$$\delta E_1 \leq \delta E_3 \tag{41}$$

$$\delta E_1 \leq \delta E_4 \tag{42}$$

$$\delta E_2 \leq \delta E_3 \tag{43}$$

$$\delta E_2 \leq \delta E_4 \tag{44}$$

Assume

$$\omega_1 + \omega_2 = 1 \tag{45}$$

Where $0 \leq \omega_1 \leq 1$, $0 \leq \omega_2 \leq 1$;

$$\omega_3 + \omega_4 = 1 \tag{46}$$

Where $0 \leq \omega_3 \leq 1$, $0 \leq \omega_4 \leq 1$;

According to Eqs. (41), (42), (43), (44), (45) and (46), the following result can be obtained:

$$\begin{aligned}&\omega_1\delta E_1 + \omega_2\delta E_2 \leq \omega_1\delta E_2 + \omega_2\delta E_2 \\ &\quad = \delta E_2 \leq \delta E_3 = \omega_3\delta E_3 + \omega_4\delta E_3 \leq \omega_3\delta E_3 + \omega_4\delta E_4\end{aligned} \tag{47}$$

That is:

$$\omega_1\delta E_1 + \omega_2\delta E_2 \leq \omega_3\delta E_3 + \omega_4\delta E_4 \tag{48}$$

Where ω_1, ω_2, ω_3, and ω_4 are the weights of FM 1, FM 2, FM 3 and FM 4, respectively.

According to Eq. (32), Eqs. (49) and (50) can be obtained:

$$\delta E_{Z,12} = \omega_1\delta E_1 + \omega_2\delta E_2 \tag{49}$$

$$\delta E_{Z,34} = \omega_3\delta E_3 + \omega_4\delta E_4 \tag{50}$$

Combining Eqs. (48), (49) and (50), the following conditions are given:

$$\delta E_{Z,12} = \omega_1\delta E_1 + \omega_2\delta E_2 \leq \omega_3\delta E_3 + \omega_4\delta E_4 = \delta E_{Z,34} \tag{51}$$

where $\delta E_{Z,12}$ is the combined relative forecasting error based on FM 1 and FM 2 and $\delta E_{Z,34}$ is the combined relative forecasting error based on FM 3 and FM 4.

Eqs. (52) and (53) can be obtained according to Eq. (37):

$$\delta E^*_{basic,12} = \frac{\delta E_{z,12}\left|\Delta E_{basic,act}\right|}{Q_{act}} = \delta E_{z,12}\delta E_{basic} \tag{52}$$

$$\delta E^*_{basic,34} = \frac{\delta E_{z,34}\left|\Delta E_{basic,act}\right|}{Q_{act}} = \delta E_{z,34}\delta E_{basic} \tag{53}$$

Combining Eqs. (51), (52) and (53) yields the following results:

$$\delta E^*_{basic,12} \leq \delta E^*_{basic,34} \tag{54}$$

After the residual error correction by different combinations, the range of relative forecasting errors δE^*_{basic} , is given:

$$\min\left(\delta E_1, \delta E_2\right)\delta E_{basic} \leq \delta E^*_{basic,12} \leq \max\left(\delta E_1, \delta E_2\right)\delta E_{basic} \tag{55}$$

$$\min\left(\delta E_3, \delta E_4\right)\delta E_{basic} \leq \delta E^*_{basic,34} \leq \max\left(\delta E_3, \delta E_4\right)\delta E_{basic} \tag{56}$$

Eq. (42) is obtained by combining Eqs. (39), (40) and (41):

$$\begin{aligned}&\min\left(\delta E_1, \delta E_2\right)\delta E_{basic} \leq \delta E^*_{basic,12} \leq \max\left(\delta E_1, \delta E_2\right)\delta E_{basic} \leq \\ &\min\left(\delta E_3, \delta E_4\right)\delta E_{basic} \leq \delta E^*_{basic,34} \leq \max\left(\delta E_3, \delta E_4\right)\delta E_{basic}\end{aligned} \tag{57}$$

The above derivation indicates that when a residual error combination forecasting model composed of the same number of single FMs is used for forecasting, the final error can be reduced by combining the single FMs with smallest errors, thereby improving the forecasting accuracy.

2.5.2. Different number of options for combination forecasting

Assume that there are five types of FMs in some time period, represented by 1, 2, 3, 4 and 5, respectively, whose relative forecasting errors are δE_1, δE_2, δE_3, δE_4 and δE_5.

$$\delta E_1 = \delta E_2 = \delta E_3 = \delta E_4 \neq \delta E_5 \tag{58}$$

Suppose that there are two combined residual error correction FMs: the first model consists of single models 1 and 2 and the second consists of single models 3, 4 and 5.

Assume

$$\omega_1 + \omega_2 = 1 \tag{59}$$

Where $0 \leq \omega_1 \leq 1$, $0 \leq \omega_2 \leq 1$ and

$$\omega_3 + \omega_4 + \omega_5 = 1 \tag{60}$$

Where $0 \leq \omega_3 \leq 1$, $0 \leq \omega_4 \leq 1$, $0 \leq \omega_5 \leq 1$

According to Eq. (32), Eqs. (61) and (62) can be obtained:

$$\delta E_{Z,12} = \omega_1 \delta E_1 + \omega_2 \delta E_2 \tag{61}$$

$$\delta E_{Z,345} = \omega_3 \delta E_3 + \omega_4 \delta E_4 + \omega_5 \delta E_5 \tag{62}$$

where $\delta E_{Z,12}$ is the combined relative forecasting error based on FM 1 and FM 2, and $\delta E_{Z,345}$ is the combined relative forecasting error based on FMs 3, 4 and 5.

Assume $E_1 = \delta E_2 = \delta E_3 = \delta E_4 \leq \delta E_5$, the following result can then be obtained:

$$\begin{aligned} \delta E_{Z,345} &= \omega_3 \delta E_3 + \omega_4 \delta E_4 + \omega_5 \delta E_5 \geq \omega_3 \delta E_3 + \omega_4 \delta E_3 + \omega_5 \delta E_3 \\ &= \delta E_3 = \delta E_2 = \omega_1 \delta E_2 + \omega_2 \delta E_2 = \omega_1 \delta E_1 + \omega_2 \delta E_2 = \delta E_{Z,12} \end{aligned} \tag{63}$$

Eq. (64) can be obtained by from Eq. (37):

$$\delta E^*_{basic,345} = \delta E_{z,345} \delta E_{basic} \geq \delta E_{z,12} \delta E_{basic} = \delta E^*_{basic,12} \tag{64}$$

Clearly, if $\delta E_1 \leq \delta E_2 \leq \delta E_3 \leq \delta E_4 \leq \delta E_5$, Eq. (53) is obtained:

$$\delta E^*_{basic,12} \leq \delta E^*_{basic,345} \tag{65}$$

The above proof shows that when two single FMs and three single FMs are used to form a combined FM to correct the basic model, the two single FMs with small relative errors are chosen to have higher forecasting accuracy than the combined FM composed of three single FMs with large relative errors.

2.5.3. The method for dynamically adjusting single FMs for combination forecasting

As mentioned earlier, due to the different features of single FMs and the randomness and uncertainty of some of the factors affecting the forecasting accuracy, the performance of single FMs is unstable. Therefore, in order to minimize those effects and get a more accurate forecasting, we pose the method of dynamical adjustment of single FM for combined residual error forecasting model. That is, the two single FMs with smallest relative forecasting errors in the last forecasting period are chose to form the combined residual error forecasting model in every current forecasting period. So far, the FM with dynamically combined residual error correction based on the optimal model combination is built.

2.6. Evaluation index

To evaluate the performance of forecasting models, four statistical evaluation indicators were introduced include relative error, MAE, MAPE and root mean square error (RMSE). The statistical evaluation index reflects the degree of fit between the predicted value and the actual value. MAE could avoid the problem of cancelling errors, and thus can accurately reflect the size of the actual forecast error. MAPE has more denominator y_i than MAE. MAPE equal to 0 indicates a perfect model, more than 1 means inferior model (Yu et al., 2021). Moreover, compared with other indicators, it can identify the impact of errors caused by outliers on the accuracy of the model. RMSE is the quadratic root of the ratio of the square of the deviation between the predicted value and the actual value. RMSE is easier to identify large errors and can describe the degree of dispersion of the predicted values. If the maximum deviation is large, the RMSE will be enlarged.

$$Relative\ Error = \frac{\left| y_i^* - y_i \right|}{y_i} \tag{66}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i^* - y_i \right| \tag{67}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i^* - y_i}{y_i} \right| \times 100\% \tag{68}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i^* - y_i \right)^2} \tag{69}$$

where y_i represents the actual value, y_i^* represents the predicted value.

3. Case analysis

3.1. Project introduction

The case study is from an office building in Xi'an, a western city. The office building has 44 floors above ground and 3 floors underground, with a total construction area of 300000 m^2 and land area of 30000 m^2. The wall structure of the building adopts a concrete shear wall with shape coefficient of 0.8. The building has deployed the environmental detection sensors and energy consumption collection system, which could collect outdoor humidity, temperature and energy consumption data in real time (Yu et al., 2021).

3.2. Data sources

The building energy consumption data collected from April 1, 2021 to July 31, 2021, including hourly data of air conditioning power, motive power, special power and lighting socket power consumption. The data have been collected from 8 am to 22 pm every day at per hour granularity. Fig. 2 shows the four types of electricity loads for the forecast period from July 29 to July 31 from 8 a.m. to 22 p.m.

Meteorological data were included outdoor dry bulb temperature, relative humidity, wind speed and intensity of solar radiation. The data have been collected from 8 am to 22 pm every day at per hour granularity. Fig. 3 shows the outdoor meteorological information for the forecast period of July 29–31, and the outdoor temperature change ranged from 20.8 °C to 34.6 °C. On the other hand, relative humidity, wind speed and intensity of solar radiation are affected by weather conditions and uncertain factors. These meteorological factors are used as inputs to the SVR model, and the real building energy consumption is used as an output.

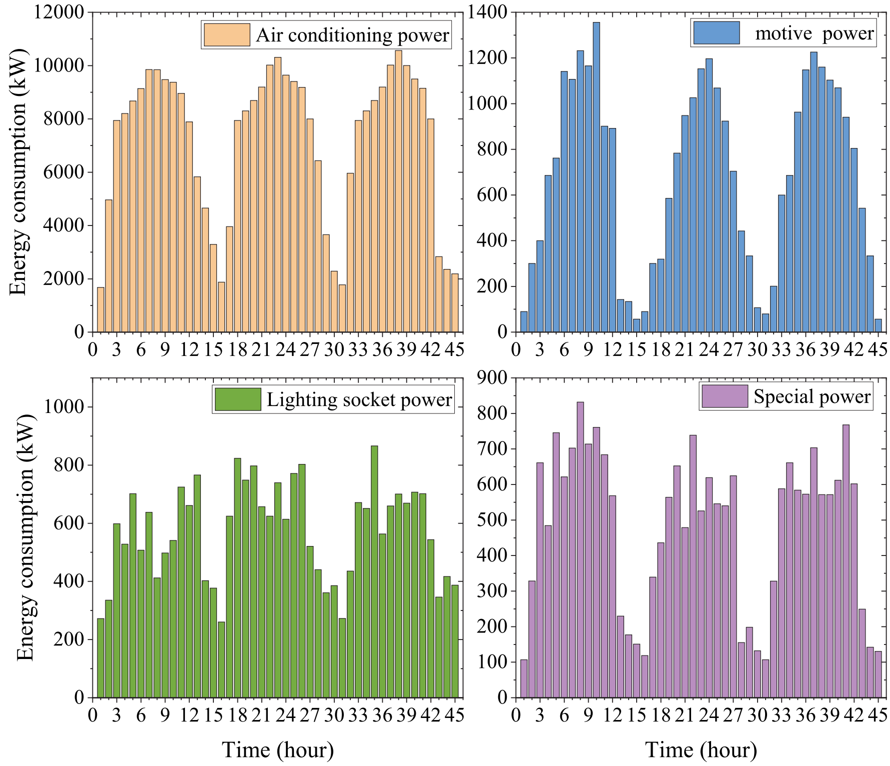

Fig. 2. Four types of electricity load for the forecast period from July 29 to 31.

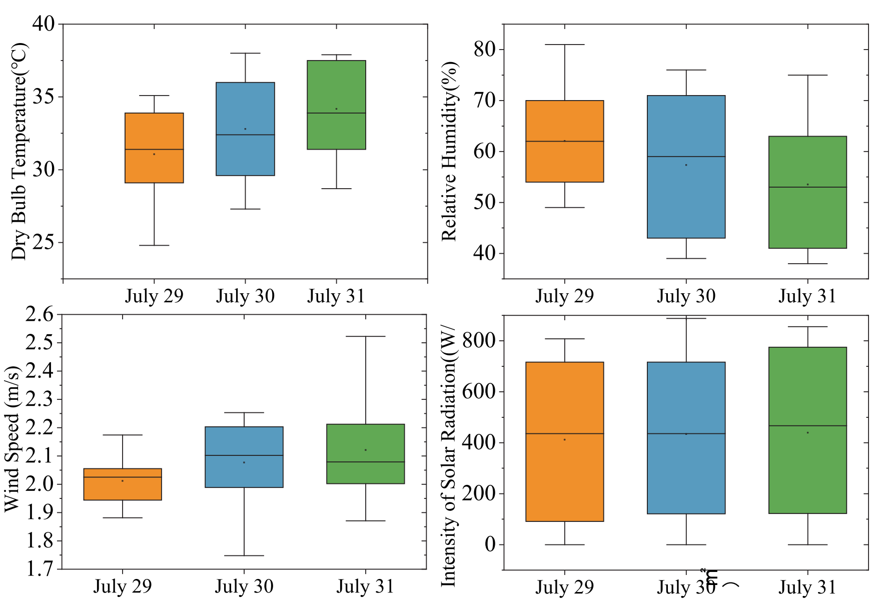

Fig. 3. The outdoor meteorological information for the forecast period of July 29 to July 31.

The forecasting of energy consumption in an office building is taken as an example to verify the accuracy of the FM with dynamically combined residual error correction based on the optimal model combination. The six FMs adopted in this paper, the SVR is regarded as the basic FM, ARIMA, BPNN, RFR, GM and MLR are used to construct the residual error combination forecasting models.

Data are collected at hourly intervals, and the collection time is from 8:00 am to 22:00 pm every day. The model construction is divided into two parts: the first part is the construction of the SVR model. A total of 1360 sets of weather and building energy data from April 1 to July 1 were used to train and validate the SVR model. The SVR model was tested with data from July 1 to July 31, and 465 sets of SVR residual errors were obtained. The second part is the forecasting of the residual values. A total of 420 data residual errors from July 1 to July 28 were used as the training set, and five single FMs were used to forecasting the residual errors from July 29 to 31, respectively, and a combined residual error FM was further constructed to correct the SVR model.

Table 1
The errors of the five single FMs forecast for the residual error of SVR.

TIME	Real	ARIMA		BPNN		RFR		GM		MLR	
	residual error of SVR (kW)	Forecast residual error (kW)	Relative error	Forecast residual error (kW)	Relative error	Forecast residual error (kW)	Relative error	Forecast residual error (kW)	Relative error	Forecast residual error (kW)	Relative error
7.29–8:00	−363.17	−148.73	0.59	−474.40	0.31	−474.40	0.30	−271.36	0.25	−296.85	0.18
7.29–9:00	−2248.08	−1280.73	0.43	−2063.45	0.29	−2063.45	0.08	−1844.60	0.17	−2718.08	0.20
7.29–10:00	−1285.80	−786.60	0.38	−1078.72	0.24	−1078.72	0.16	−1036.36	0.19	−950.58	0.26
7.29–11:00	−2800.33	−3517.65	0.25	−1667.22	0.40	−2192.13	0.21	−4006.75	0.43	−3517.65	0.25
7.29–12:00	−2665.31	−1697.63	0.36	−1772.96	0.33	−1718.67	0.35	−1184.18	0.55	−3804.71	0.42
7.29–13:00	2310.65	3119.48	0.35	3077.86	0.33	1906.27	0.17	2786.94	0.20	1609.76	0.30
...	...	...	...	...	...	...	...	...	...	...	...
7.31–17:00	3906.09	4987.56	0.27	2761.85	0.29	5204.89	0.33	3368.21	0.13	5125.28	0.31
7.31–18:00	5960.33	3103.43	0.47	7818.70	0.31	4176.86	0.29	3281.19	0.44	4183.18	0.29
7.31–19:00	2992,98	1460.89	0.51	3584.42	0.19	2161.64	0.27	3973.70	0.32	3832.01	0.28
7.31–20:00	970.43	1443.44	0.48	896.18	0.076	711.14	0.26	775.54	0.20	671.75	0.30
7.31–21:00	263.39	363.58	0.38	205.96	0.21	318.90	0.21	161.09	0.38	299.55	0.13
7.31–22:00	279.21	218.66	0.21	331.47	0.18	227.16	0.18	244.51	0.12	404.69	0.44

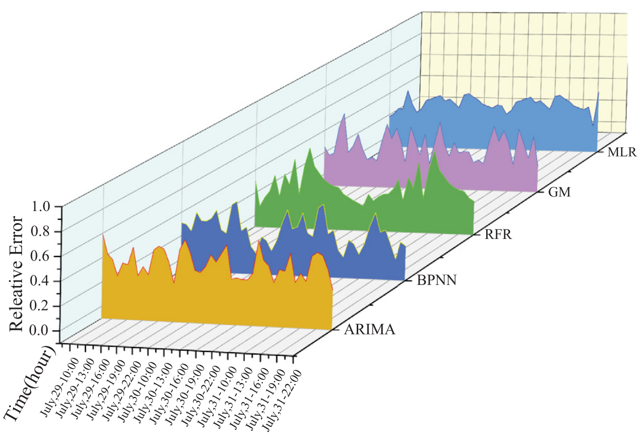

Fig. 4. The relative errors of the five single FMs forecast for the residual error of SVR.

4. Result and discussion

4.1. Residual error forecasting

Fig. 4 shows the relative errors of the forecasting residual error of the five single FMs. The forecast period runs from 8:00 a.m. to 22:00 p.m. on July 29–31, at hourly intervals, for a total of 45 forecast moments. Table 1 shows the forecasting results of five single FMs for the residuals at some moments.

In the actual forecasting, the residual values of the basic model at the forecast moment are not known. Therefore, it is not possible to dynamically select a single FM based on the relative error magnitude of the single FM. However, it is possible to select the single FM that constitutes the combined forecasting model at the current forecasting moment based on the relative error of the single FM at the previous forecasting moment. The single model with a small relative error at the previous time is selected to form the combined residual error correction forecasting model at the current time. The 2, 3 and 4 single forecasting models with the smallest relative errors are selected to form a dynamic combination of residual error models for SVR correction. (defined in the following as SVR-2 Single FMs, SVR-3 Single FMs, SVR-4 Single FMs) Finally, the building energy consumption value forecast by the SVR were corrected. The model is deployed in Python 3.7 of win10, a 64-bit operating system.

4.2. Building energy consumption forecast results

Fig. 5 shows the forecast building energy consumption after correcting SVR by different models. There are eight combination forecasting models in Fig. 5. Five fixed models forecast residuals and correct SVR.(SVR-ARIMA, SVR-BPNN, SVR-RFR, SVR-GM, SVR-MLR), Three dynamic combination models to forecast residuals correction SVR.

Fig. 6 presents the relative forecasting error of SVR corrected by different FMs. The relative errors of the uncorrected SVR forecasts ranged from 6%–45%. The forecasting accuracy after correcting the SVR by different single models varied, with the mean relative error ranging from 5%–7%. The forecasting accuracy was significantly improved after correcting the residuals of SVR by the single model. The mean relative error of the dynamic combined model corrected SVR is about 3%, which is higher than the forecasting accuracy of the fixed model corrected SVR.

Fig. 7 shows the weights of the two single FMs in Fig. 6 that constitute the dynamic combined residual error FM. Fig. 8 shows the weights of the three single FMs in Fig. 6 that constitute the combined residual error FM. Fig. 9 shows the weights of the four single FMs in Fig. 6 that constitute the dynamic combined residual error FM.

Figs. 7–9 show that the weights of different single FMs are not constant. They also show that the bigger is the relative errors of forecasting model, the less is the corresponding weight.

Fig. 10 shows the curves of the relative forecasting errors of SVR corrected by three types of dynamically combined residual error model. The forecasting accuracy of the dynamic combination of two single FMs to correct SVR is better than that of using three and four single FMs. The relative error is between 1%–7% when correcting SVR using a dynamic combined residual model. The forecasting performance is better than that of correcting SVR using a fixed single FM.

Fig. 11 and Table 2 show that the FMs with dynamically combined residual error correction, the forecasting accuracy decreases with increasing number of single FMs that constitute the residual error combination FM. In this case, when multiple single FMs with known relative forecasting errors are considered, two is the optimal number of models that is used to construct the residual combination FM. Therefore, selecting the two single FMs with the smallest relative error will minimize the final error of the combination model.

Table 2 shows that the accuracy of the SVR forecasting model is significantly improved by residual error correction. The MAE,

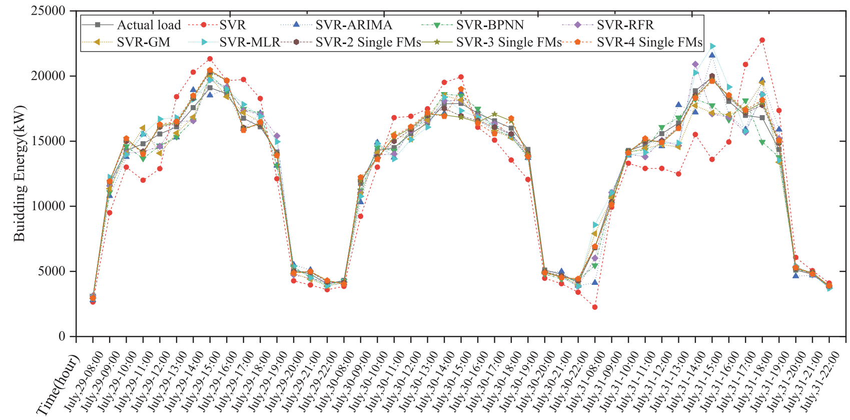

Fig. 5. Forecast building energy consumption after correction of SVR by different kinds of models.

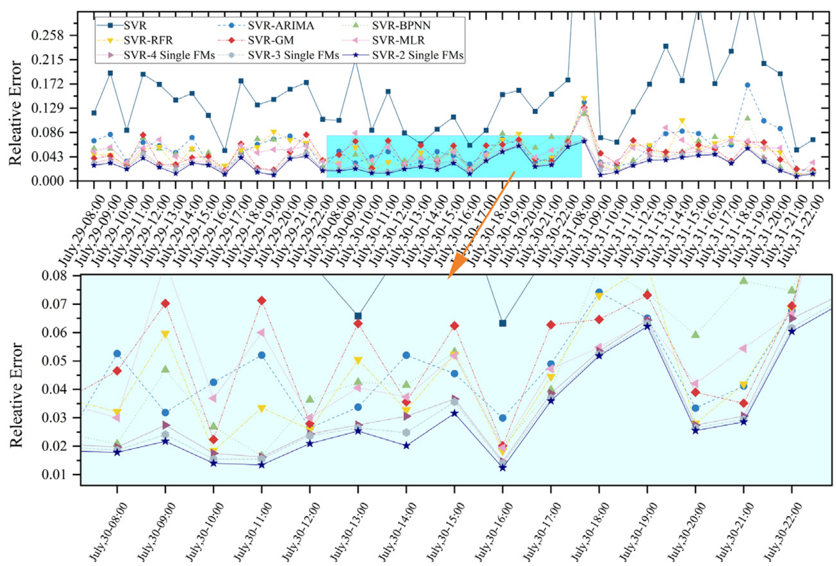

Fig. 6. Relative errors in forecasting building energy consumption after correcting SVR for different kinds of models.

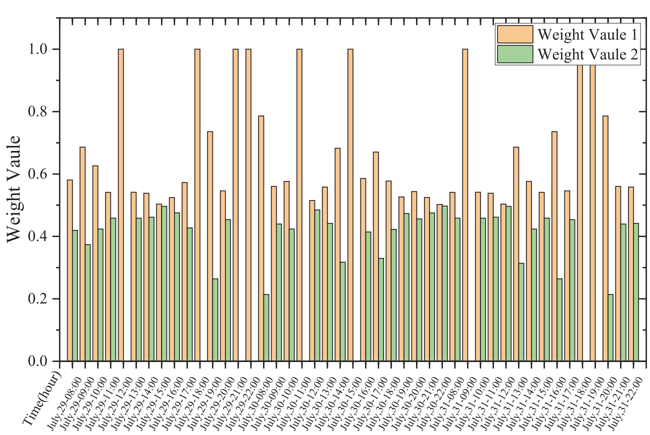

Fig. 7. The weights of the two single FMs constituting the dynamic residual errors combination FM.

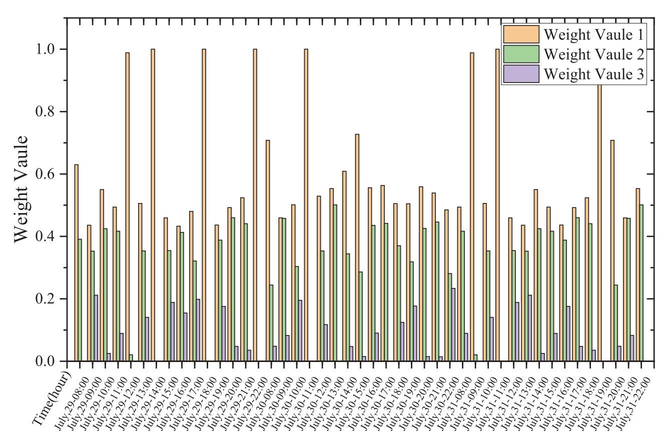

Fig. 8. The weight values of the three single FMs constituting the dynamic residual errors combination FM.

MAPE and RMSE of the dynamic combined RE correction model using two single FMs to correction are 349.37, 0.0296 and 471.44, which is less error than the combined residual error correction model using three and four FMs, and also have more less error than other FMs in Table 2.

To further verify the effectiveness and generalization ability of the model proposed in this paper, BPNN was selected as the basic forecasting model, and the residual values of BPNN were forecasted by the other five single models, and the final forecasting values of BPNN were improved after the residual correction.

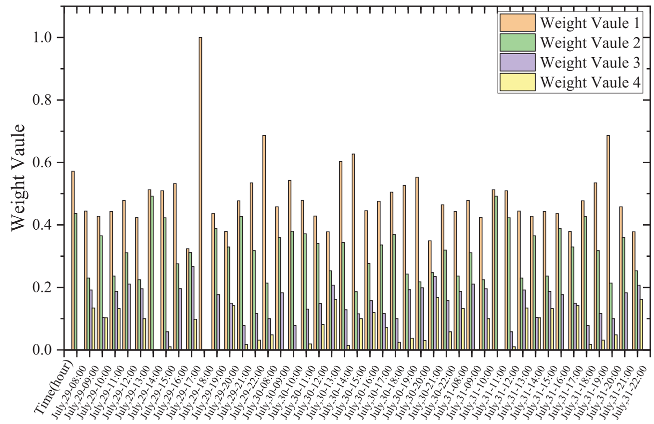

Fig. 9. The weight values of the four single FMs constituting the dynamic residual error combination FM.

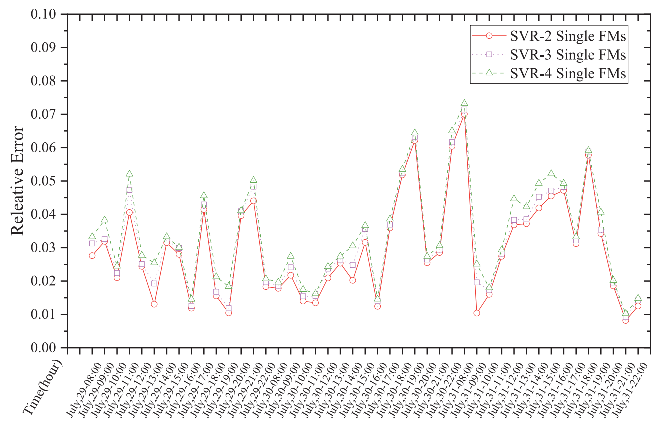

Fig. 10. Error analyses on SVR with dynamical residual errors combined correction based on two, three and four single FMs.

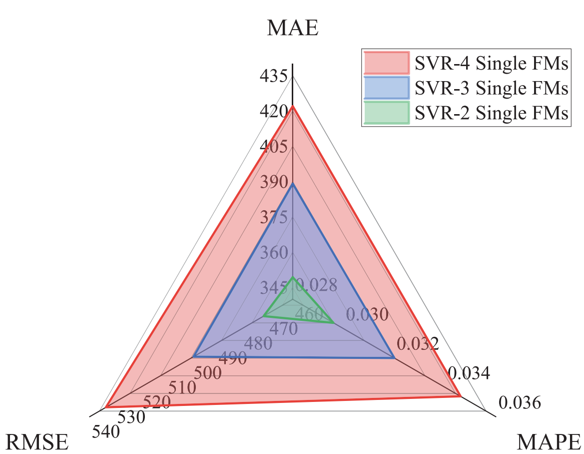

Fig. 11. Performance comparison of dynamically combined 2, 3 and 4 single FMs correction SVR.

Table 3 shows the forecasting results of different correction methods. BPNN-2 Single FMs has the lowest forecasting error, and its MAE, MAPE, and RMSE are 401.33, 0.0332, and 434.45, respectively. From the forecasting results of BPNN-2 Single FMs, BPNN-3 Single FMs, and BPNN-4 Single FMs, it can be obtained that the best forecasting results are obtained by choosing the two single models with the smallest relative errors to form a combined model to correct the BPNN, which is the same as the conclusion proved in Section 2.4.

Table 2
Error analysis between different forecasting models.

Forecasting methods	MAE	MAPE	RMSE
SVR	1918.59	0.1580	2278.74
SVR-ARIMA	790.36	0.0684	898.01
SVR-BPNN	623.97	0.0563	779.43
SVR-MIR	587.35	0.0517	703.23
SVR-GM	558.96	0.0491	677.31
SVR-REF	543.87	0.0488	643.12
SVR-4 Single FMs	422.07	0.0346	533.45
SVR-3 Single FMs	389.12	0.0320	498.98
SVR-2 Single FMs	349.37	0.0296	471.44

Table 3
Error analysis between different forecasting models.

Forecasting methods	MAE	MAPE	RMSE
BPNN	2338.44	0.1673	2566.72
BPNN-ARIMA	860.42	0.0702	992.13
BPNN-SVR	793.32	0.0639	864.24
BPNN-MIR	627.54	0.0677	783.04
BPNN-GM	599.04	0.0571	664.92
BPNN-REF	593.87	0.0588	743.12
BPNN-4 Single FMs	501.30	0.0496	585.65
BPNN-3 Single FMs	444.42	0.0435	508.32
BPNN-2 Single FMs	401.33	0.0332	434.45

4.3. The Kolmogorov–Smirnov predictive accuracy (KSPA) test

The Kolmogorov–Smirnov predictive accuracy (KSPA) test is a supplemental statistical test used to determine the accuracy of two sets of forecasts. The first part of the KSPA test is the two-sample two-side KSPA test, which determines whether the distribution of the two forecast errors is statistically significant. The second part is a two-sample one-sided KSPA test, which determines whether the forecast with the smallest error also has a smaller random error than the competitor's forecast based on the loss function. Thus, the forecasting accuracy of the models can be compared (Fan et al., 2022).

First, a two-sample bilateral KSPA test was used to determine whether there was a statistically significant difference between the two distributions of prediction errors. The original hypothesis is that there is no significant difference between the two statistical forecasts. The original hypothesis is rejected when the bilateral KSPA test observation sample test statistic is less than (usually) 1%, 5%, or 10% (Fan et al., 2021). In this case, a statistically significant difference is inferred between the distributions of the forecasts made by the models used, which indicates that there is a statistically significant difference between the two forecasts based on the bilateral KSPA test. The purpose of the two-sample one-sided KSPA test is to determine whether a model based on the minimum error of the loss function has a smaller random error forecasting model compared to other model.

KSPA test results show that the two sided (p-value) of the SVR-2 Single FMs with other methods are <0.01 *. The value of one sided (p-value) is <0.01 * (* indicates that SVR-2 Single FMs are statistically significant based on a p value of 0.01). First, it confirms the statistically significant differences between the proposed forecasting model and the eight comparison models. Then a one-sided KSPA test is used to determine the proposed model and to compare the low random errors reported by the predictions. The results show that the proposed model has the greatest forecasting performance. The prediction results obtained using SVR-2 Single FMs outperformed the other compared forecasting models. Significant differences were found between the proposed model and the comparative models. The KSPA test error distribution and the empirical cumulative distribution function(c.d.f.) are given in Figs. 12 and 13. The dynamic combination residual correction

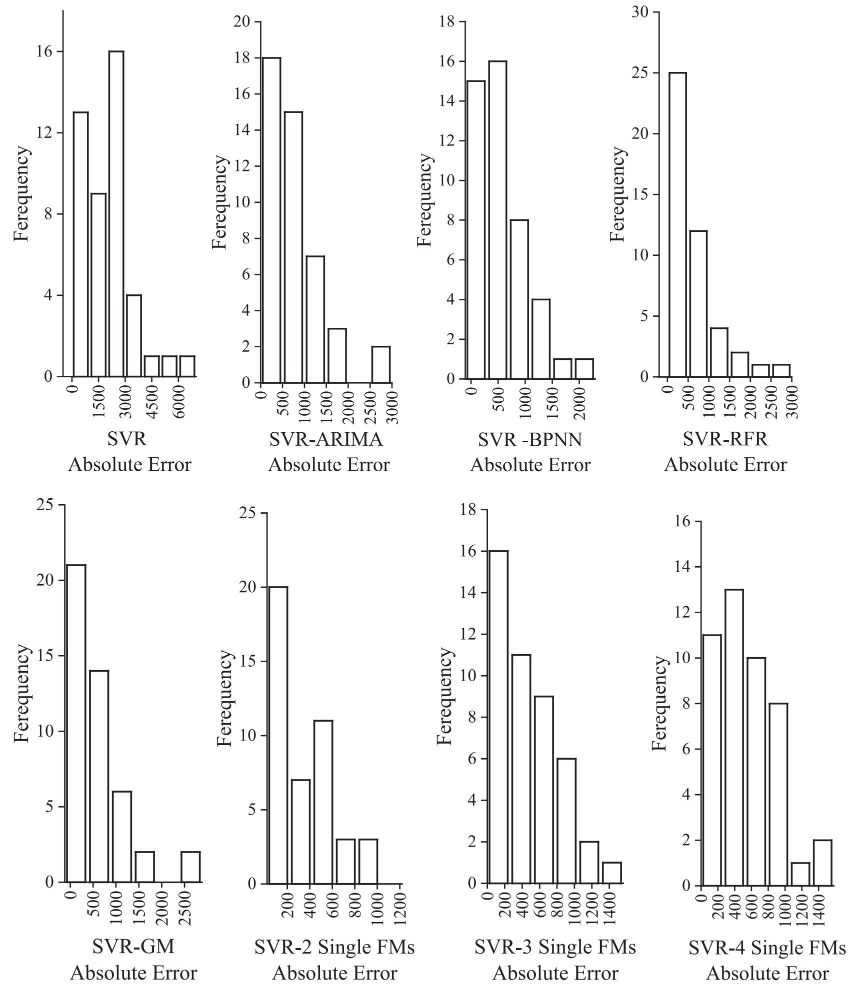

Fig. 12. Distribution of errors.

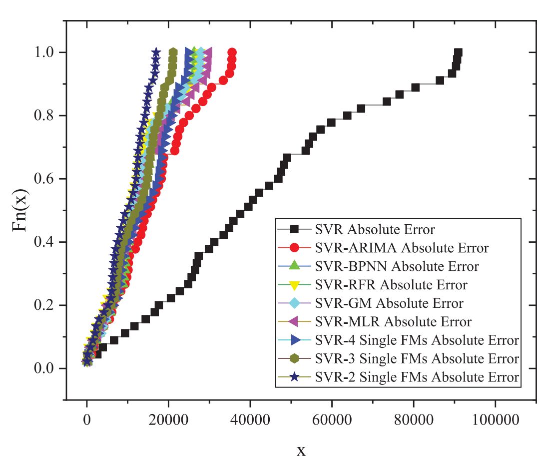

Fig. 13. Empirical Cumulative Distribution Function of Error (c.d.f.).

model based on the optimal combination approach proposed in this paper better describes the random deviation, resulting in smaller errors and higher forecasting accuracy.

5. Conclusions

This paper describes the study of the optimal combined FM for the dynamic combined residual correction model, including the method of selecting the single FM and determining the optimal number of single forecasting model for constructing the combined residual model. The following conclusions are obtained:

(1) Combining the combined residual forecasting model with the residual correction model can effectively improve the basic model forecasting accuracy and generalization ability.

(2) Dynamic combined residual forecasting model corrected SVR performs better than fixed single forecasting model corrected SVR.

(3) The basis for selecting the combined residual FM is obtained through mathematical derivation. It is verified by example that the two single FMs with the smallest relative error can form the optimal combined residual FM to minimize the forecasting error of the basic model. The corrected SVR can effectively improve the accuracy of energy consumption forecasting in office building.

(4) In the combination residual forecasting model, the weights of single FMs change with the forecasting accuracy. The larger the relative error of a single forecasting model, the smaller the corresponding weights, and vice versa, the larger the weights.

CRediT authorship contribution statement

Zengxi Feng: Data curation, Writing – original draft. **Maoqiang Zhang:** Methodology, Writing – review & editing. **Na Wei:** Supervision. **Jintong Zhao:** Software. **Tianlun Zhang:** Validation. **Xin He:** Visualization.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

This research was funded by Shaanxi Provincial Education Department 2022 Key Research Program Project (22JS022), National Key R&D Program Project Topic of China (2017YFC0704104-03), Social Science Foundation of Shaanxi Province (2022J175).

References

Amasyali, K., El-Gohary, N.M., 2018. A review of data-driven building energy consumption prediction studies. Renew. Sustain. Energy Rev. http://dx.doi.org/10.1016/j.rser.2017.04.095.

Andrawis, R.R., Atiya, A.F., El-Shishiny, H., 2011. Combination of long term and short term forecasts, with application to tourism demand forecasting. Int. J. Forecast. 27, http://dx.doi.org/10.1016/j.ijforecast.2010.05.019.

Bates, J.M., Granger, C.W.J., 1969. The combination of forecast. J. Oper. Res. Soc. http://dx.doi.org/10.1057/jors.1969.103.

Bento, P., Pombo, J., Calado, M., Mariano, S., 2021. Stacking ensemble methodology using deep learning and ARIMA models for short-term load forecasting. Energies http://dx.doi.org/10.3390/en14217378.

Bourdeau, M., Zhai, X. qiang, E., Guo, X., Chatellier, P., 2019. Modeling and forecasting building energy consumption: A review of data-driven techniques. Sustain. Cities Soc. http://dx.doi.org/10.1016/j.scs.2019.101533.

Catalina, T., Iordache, V., Caracaleanu, B., 2013. Multiple regression model for fast prediction of the heating energy demand. Energy Build. 57, http://dx.doi.org/10.1016/j.enbuild.2012.11.010.

Chan, C.K., Witt, S.F., Lee, Y.C.E., Song, H., 2010. Tourism forecast combination using the CUSUM technique. Tour. Manag. 31, http://dx.doi.org/10.1016/j.tourman.2009.10.004.

Chen, Y., Guo, M., Chen, Zhisen, Chen, Zhe, Ji, Y., 2022b. Physical energy and data-driven models in building energy prediction: A review. Energy Rep. http://dx.doi.org/10.1016/j.egyr.2022.01.162.

Chen, Y., Tan, H., 2017. Short-term prediction of electric demand in building sector via hybrid support vector regression. Appl. Energy 204, http://dx.doi.org/10.1016/j.apenergy.2017.03.070.

Chen, W., Yang, S., Zhang, X., Jordan, N.D., Huang, J., 2022a. Embodied energy and carbon emissions of building materials in China. Build Environ. 207, http://dx.doi.org/10.1016/j.buildenv.2021.108434.

Cheng, R., Yu, J., Zhang, M., Feng, C., Zhang, W., 2022. Short-term hybrid forecasting model of ice storage air-conditioning based on improved SVR. J. Build. Eng. 50, http://dx.doi.org/10.1016/j.jobe.2022.104194.

Chou, J.S., Bui, D.K., 2014. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energy Build. 82, http://dx.doi.org/10.1016/j.enbuild.2014.07.036.

Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography. Int. J. Forecast. 5, http://dx.doi.org/10.1016/0169-2070(89)90012-5.

Dhaval, B., Deshpande, A., 2020. Short-term load forecasting with using multiple linear regression. International Journal of Electrical and Computer Engineering 3911–3917. http://dx.doi.org/10.11591/ijece.v10i4.

Dong, Z., Liu, J., Liu, B., Li, K., Li, X., 2021. Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification. Energy and Buildings http://dx.doi.org/10.1016/j.enbuild.2021.110929.

Fan, C., Xiao, F., Wang, S., 2014. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Appl. Energy 127, http://dx.doi.org/10.1016/j.apenergy.2014.04.016.

Fan, G.F., Yu, M., Dong, S.Q., Yeh, Y.H., Hong, W.C., 2021. Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. Util. Policy 73, http://dx.doi.org/10.1016/j.jup.2021.101294.

Fan, G.F., Zhang, L.Z., Yu, M., Hong, W.C., Dong, S.Q., 2022. Applications of random forest in multivariable response surface for short-term load forecasting. Int. J. Electr. Power Energy Syst. 139, http://dx.doi.org/10.1016/j.ijepes.2022.108073.

Feng, J., Yang, J., Li, Y., Wang, H., Ji, H., Yang, W., Wang, K., 2021. Load forecasting of electric vehicle charging station based on grey theory and neural network. Energy Rep. 7, http://dx.doi.org/10.1016/j.egyr.2021.08.015.

Houchati, M., Beitelmal, A.M.H., Khraisheh, M., 2022. Predictive modeling for rooftop solar energy throughput: A machine learning-based optimization for building energy demand scheduling. J. Energy Resour. Technol. Trans. ASME 144, http://dx.doi.org/10.1115/1.4050844.

Karthika, S., Margaret, V., Balaraman, K., 2017. Hybrid short term load forecasting using arima-svm. In: 2017 Innovations in Power and Advanced Computing Technologies, i-PACT 2017. http://dx.doi.org/10.1109/IPACT.2017.8245060.

Kourentzes, N., Barrow, D., Petropoulos, F., 2019. Another look at forecast selection and combination: Evidence from forecast pooling. Int. J. Prod. Econ. 209, http://dx.doi.org/10.1016/j.ijpe.2018.05.019.

Li, S., Li, R., 2017. Comparison of forecasting energy consumption in Shandong, China using the ARIMA Model, GM Model, and ARIMA-GM Model. Sustainability (Switzerland) 9, http://dx.doi.org/10.3390/su9071181.

Li, K., Xue, W., Tan, G., Denzer, A.S., 2020. A state of the art review on the prediction of building energy consumption using data-driven technique and evolutionary algorithms. Build. Serv. Eng. Res. Technol. http://dx.doi.org/10.1177/0143624419843647.

Luo, X.J., Oyedele, L.O., 2021. Forecasting building energy consumption: Adaptive long-short term memory neural networks driven by genetic algorithm. Adv. Eng. Inform. 50, http://dx.doi.org/10.1016/j.aei.2021.101357.

Ma, Z., Ye, C., Ma, W., 2019. Support vector regression for predicting building energy consumption in southern China. Energy Procedia http://dx.doi.org/10.1016/j.egypro.2019.01.931.

Martins, V.L.M., Werner, L., 2012. Forecast combination in industrial series: A comparison between individual forecasts and its combinations with and without correlated errors. Expert Syst. Appl. 39, http://dx.doi.org/10.1016/j.eswa.2012.04.007.

Moradzadeh, A., Mansour-Saatloo, A., Mohammadi-Ivatloo, B., Anvari-Moghaddam, A., 2020. Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings. Applied Sciences http://dx.doi.org/10.3390/app10113829.

Ozturk, S., Ozturk, F., 2018. Forecasting energy consumption of Turkey by Arima model. J. Asian Sci. Res. 8, http://dx.doi.org/10.18488/journal.2.2018.82.52.60.

Panahi, M., Sadhasivam, N., Pourghasemi, H.R., Rezaie, F., Lee, S., 2020. Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). Journal of Hydrology http://dx.doi.org/10.1016/j.jhydrol.2020.125033.

Parvin, K., Lipu, M.S.H., Hannan, M.A., Abdullah, M.A., Jern, K.P., Begum, R.A., Mansur, M., Muttaqi, K.M., Mahlia, T.M.I., Dong, Z.Y., 2021. Intelligent controllers and optimization algorithms for building energy management towards achieving sustainable development: Challenges and prospects. IEEE Access 9, http://dx.doi.org/10.1109/ACCESS.2021.3065087.

Shaikh, P.H., Nor, N.B.M., Nallagownden, P., Elamvazuthi, I., Ibrahim, T., 2014. A review on optimized control systems for building energy and comfort management of smart sustainable buildings. Renew. Sustain. Energy Rev. http://dx.doi.org/10.1016/j.rser.2014.03.027.

Singh, P., Dwivedi, P., 2019. A novel hybrid model based on neural network and multi-objective optimization for effective load forecast. Energy 182, http://dx.doi.org/10.1016/j.energy.2019.06.075.

Wang, H.J., Jin, T., Wang, H., Su, D., 2022a. Application of IEHO–BP neural network in forecasting building cooling and heating load. Energy Rep. 8, http://dx.doi.org/10.1016/j.egyr.2022.01.216.

Wang, R., Lu, S., Feng, W., 2020. A novel improved model for building energy consumption prediction based on model integration. Appl. Energy 262, http://dx.doi.org/10.1016/j.apenergy.2020.114561.

Wang, X., Yang, J., Zhou, Q., Liu, M., Bi, J., 2022b. Mapping the exchange between embodied economic benefits and CO2 emissions among belt and road initiative countries. Appl. Energy 307, http://dx.doi.org/10.1016/j.apenergy.2021.118206.

Yang, X., Liu, S., Zou, Y., Ji, W., Zhang, Q., Ahmed, A., Han, X., Shen, Y., Zhang, S., 2022. Energy-saving potential prediction models for large-scale building: A state-of-the-art review. Renew. Sustain. Energy Rev. http://dx.doi.org/10.1016/j.rser.2021.111992.

Yao, Y., Lian, Z., Hou, Z., Liu, W., 2006. An innovative air-conditioning load forecasting model based on RBF neural network and combined residual error correction. Int. J. Refrig. 29, http://dx.doi.org/10.1016/j.ijrefrig.2005.10.008.

Yu, J., Zhang, T., Zhao, A., Xie, Y., 2021. Research on energy consumption prediction of office buildings based on comprehensive similar day and ensemble learning. J. Intell. Fuzzy Systems 40, http://dx.doi.org/10.3233/JIFS-210069.

Zhang, Y.W., 1991. The mathematic method of prediction. National defence of Industry Press.

Zhang, L., 2021. Data-driven building energy modeling with feature selection and active learning for data predictive control. Energy Build 252, http://dx.doi.org/10.1016/j.enbuild.2021.111436.

Zhang, F., Dec, C., Lee, S.E., Yang, J., Shah, K.W., 2016. Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique. Energy and Buildings http://dx.doi.org/10.1016/j.enbuild.2016.05.028.