Energy & Buildings 225 (2020) 110343

Contents lists available at ScienceDirect

Energy & Buildings

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

Short-term metropolitan-scale electric load forecasting based on load decomposition and ensemble algorithms

Yiyi Chu[a], Peng Xu[b,*], Mengxi Li[c], Zhe Chen[b], Zhibo Chen[b], Yongbao Chen[b], Weilin Li[d]

[a] College of Civil Engineering, Iowa State University, 701 Morill Road, Ames, IA 50011, USA
[b] College of Mechanical and Environmental Engineering, Tongji University, Cao'an Road No. 4800, Shanghai 201804, China
[c] IB SCHOLZ GmbH & Co.KG, Galgenbergstraße 15, 93053 Regensburg, Germany
[d] College of Mechanical Engineering, Zhengzhou University, Kexue Road No. 100, Zhengzhou 450001, China

ARTICLE INFO

Article history:
Received 9 October 2019
Revised 3 March 2020
Accepted 24 July 2020
Available online 28 July 2020

Keywords:
Power load characteristic
Seasonal attribute
Load decomposition
Ensemble algorithm

ABSTRACT

This paper presents an ensemble algorithm based on a new load decomposition method to forecast short-term metropolitan-scale electric load. In this method, a decision tree for hourly seasonal attributes and a weighted average method for daily seasonal attributes are first applied to divide seasons into a completely different way. Then, the load of transition seasons is chosen as a basic component according to power load characteristics, and the differences between total load and the basic component are extracted as the weather-sensitive component. Finally, a time-series method is selected to forecast the basic component and SVM (Support Vector Machine) to the weather-sensitive component. This paper takes the annual electricity load of Shanghai as a case study to verify this ensemble method. The results show that compared with the traditional model based on overall daily load and other load decomposition methods—EMD (Empirical Mode Decomposition) and WT (Wavelet Transform), this ensemble model reduces the error from 3 to 5% to lower than 2% when forecasting the power load of workdays, and for non-work days, the error is decreased from 4 to 5% to lower than 4%.

1. Introduction

The close tracking of electricity generation in response to the load requirements is an important aspect of the operation of a power system. Accurate load forecasting is a necessity of most utilities for energy purchasing, transmission and distribution planning, operations and maintenance, demand side management, etc. Load forecasting has different requirements for the lead time ranging from the short term (a few minutes, hours or days ahead) to the long term (up to 30 years ahead) [1]. It is important for Demand Response, which is defined as "changes in power consumption by demand-side resources from their normal consumption patterns in response to changes in electricity price or to incentive payments designed to reduce electricity use during peak load periods" [2]. By reducing energy consumption from on-peak periods to valley period, DR could improve the efficiency of power stations and ensure grid security. And one of the most important DR strategies is the forecasting of a fair and accurate baseline, especially during peak hours. That's why short-term load forecasting is important. In addition, a forecast error, no matter under-prediction or over-prediction, could result in increased operating costs. For example, it was estimated that an increase of 1% in the forecasting error was related to 10 million pounds increase in operating costs in the British thermal power system [3]. With the deregulation of the competitive electricity markets, the short-term load forecasting (STLF) has become increasingly important to be more accurate and faster.

A significant amount of research on electric load forecasting has been conducted over the past 60 years to improve forecasting accuracy. Methods of short-term electric load forecasting can be roughly categorized into three groups: statistical approaches, artificial intelligence (AI)-based approaches, and hybrid approaches. Statistical approaches attain a mathematical model to build the relationship between the electric load and input parameters, such as multiple regression [4], exponential smoothing [5], ARIMA (Autoregressive Integrated Moving-average) [6], etc. However, for the models mentioned above, the electric load has typically been divided into basic and weather-dependent components based on assumptions of linearity, which is not very effective because of the distinctly nonlinear functions of exogenous variables. Therefore, artificial intelligence (AI)-based approaches have been developed to predict the electricity load, the most commonly used are fuzzy logic [7], artificial neural networks [8], and support vector

[*] Corresponding author.
E-mail address: xupeng@tongji.edu.cn (P. Xu).

https://doi.org/10.1016/j.enbuild.2020.110343
0378-7788/© 2020 Elsevier B.V. All rights reserved.

regression [9]. The hybrid approaches are to combine more than one forecasting method together to overcome the shortcoming of each single method, which includes sequential and parallel hybrid method [10]. For example, Fatemeh Chahkoutahi and Mehdi Khashei [11] used multiplayer perceptrons neural network, Adaptive Network-based Fuzzy Inference System, and Seasonal Autoregression Integrated Moving Average methods to predict electric load separately, then a direct optimum parallel hybrid model was proposed to attain the relative weight of each model. The final forecast was calculated by combing the weighted average of predicted values of aforementioned methods. Although the predicting accuracy was improved, it is difficult to generalize the weightage calculating method of each model. Thus, the other way to accomplish parallel modeling has been proposed, of which the electricity load was decomposed into several components and different components were predicted by different models. And the final forecasting result was the sum of each model's forecasted value.

For the load decomposition method, it is important to select proper inputs to infer the electricity load and use proper method to divide the electricity load into several sub-components based on the load characteristics. The electric load characteristics are influenced by four major types of factors: economic factors, time factors, weather factors, and random effects, summarized by Gross and Galiana [4]. Weather plays an important role in load variations, and a variety of weather variables and treatment methods have been reported in literature. Liu, et al. [12] investigated the effects of meteorological parameters on building energy consumption based on the sensitivity analysis in China. Results show that temperature has the strongest effect on heating and cooling, while wind speed has the least impact on air conditioning energy, and solar radiation is not an important parameter affecting building energy consumption in both winter and summer. Therefore, among meteorological parameters, temperature (dry bulb temperature) is the most important because of its effects on cooling load and electric heating. Hong [13] summarized various ways to treat temperature information in different models, such as current hourly temperature, previous hourly temperatures, the difference between the last hourly temperature and the current one, and the maximum, minimum, or average temperatures over several hours, for example. The most used temperature parameter used as STFC model input is daily average temperature. However, this approach is not always appropriate. When the temperature difference between day and night is large, cooling or heating requirement mainly appear only during a certain time of a day, either daytime or nighttime. In this case, comparing the daily mean temperature to a threshold is not justifiable. Cooling would not be needed for a whole day, and, conversely, heating requirements would sometimes last for a short period as well. Therefore, cooling or heating requirements cannot be evaluated correctly based on the above rule. To address this problem, this paper presents a new rule to re-divide seasons into heating, cooling and transition seasons carefully based on HVAC (Heating, Ventilation, and air conditioning) fundamentals, and meteorological parameters are evaluated comprehensively, including temperature, relative humidity, wind speed, and air enthalpy. In addition, to reflect the lagging effect caused by the thermal inertia of building structures, average meteorological parameters of the forecast time and the previous three hours are used as inputs [14].

To date, researchers have presented a number of methods for signal decomposition and transformation, such as Fourier analysis, which is less effective for capturing short-duration transient variations [15]. Wavelet transform has been shown to overcome the difficulties encountered using Fourier methods for non-stationary signal presentation. It also presents an excellent local performance in analyzing a signal in both the time and frequency domains. Although wavelet analysis has been applied successfully in signal processing, pattern identification, image processing, and other fields, the selection of the mother wavelet and scaling function is largely based on experience and is therefore non-adaptive. Consequently, different decomposition results would be obtained with the same signal with the selection of different wavelet basis [16]. To improve the performance of decomposition, Empirical mode decomposition (EMD) was presented by Huang in 1998 [17], which is a processing method suitable for non-linear and non-stationary series analysis. EMD overcomes the difficulty of selecting the optimal wavelet basis in wavelet transform as a type of self-adaptive signal decomposition method, which has been effectively applied by many researchers [18 19 20]. Fan et al. [21] used the differential empirical mode decomposition method to decompose the electricity load into several detail parts and an approximate part, and then SVR and auto regression were used for prediction. And Zhang et al. [22] have proposed a hybrid model based on improved empirical mode decomposition, autoregressive integrated moving average and wavelet neural network optimized by fruit fly optimization algorithm.

Although EMD can provide better decomposition results than wavelet transform method, these methods depend on the reliability of the data. Typically, high-quality and widely ranging data can make load decomposition more effective and adaptive. The sample data set is often limited, so the load decomposition method cannot be generalized. Moreover, application of signal processing methods to decompose load is complex and cannot reflect the physical laws behind load changes. Therefore, a new method for load decomposition and separating out basic load and weather-dependent load based on realistic conditions and physical laws is proposed in this paper, thus the characteristics of electricity load could be extracted more accurately and effectively. Then, a time-series method was applied to forecast basic load and SVM for weather-sensitive load according to the features of each component. The forecasting results were also compared with the traditional model based on overall daily load and other load decomposition methods—EMD (Empirical Mode Decomposition) and WT (Wavelet Transform). Results show that the prediction accuracy was significantly improved.

The remaining parts of the paper are organized as follows. Section 2 provides data analysis on the electricity load series characteristics to find the relationship between weather data and electricity load. Section 3 overview the proposed forecasting method, consisting of seasonal attributes to redefine the seasons (Section 4) and load decomposition method to divide the load into basic load and weather-based load based on physical laws (Section 5). Section 6 describes the forecasting results. And Section 7 concludes the paper.

2. Power load characteristics

The case study used in this article was the annual electricity load of Shanghai in 2014, which were sampled at 15-minute intervals. In 2014, the total electricity consumption in Shanghai was 135,834.37 GWh; the peak load was 26,790.6 MW in summer and 22,202.1 MW in winter. The data set covers the total electricity consumption of Shanghai for the entire year, including the primary, secondary, and tertiary industrial sectors, as well as the residential power consumption of urban and rural residents. Different load demands had disparate power characteristics; for example, office buildings consumed the most power from 7:00 to 18:00 on weekdays, whereas the power consumption of commercial buildings was concentrated mainly within the period of 8:00–21:00. For secondary industries, the load change tended to be mild and varied with different techniques and schedules.

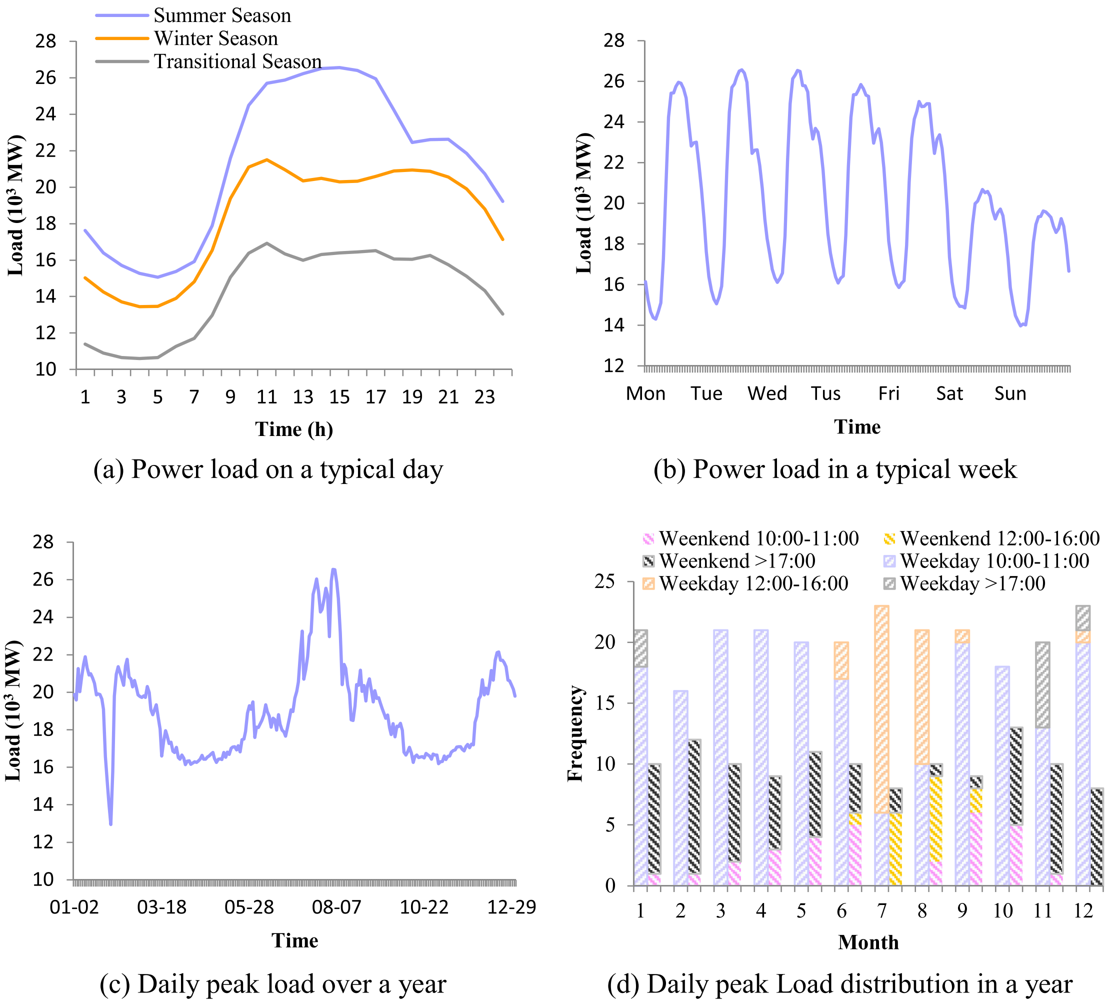
The electric load can be further divided based on different types of end-use, such as air conditioning load, lighting load, and equipment load. The air conditioning load is primarily influenced by weather, and the lighting load is mainly affected by the timing, duration, and intensity of sunlight. Therefore, the characteristics of electric load are the result of the combined effects of these influential factors based on the total energy consumption of these different items. For this reason, power load characteristics are first analyzed to establish the basis of the load decomposition method and ensemble forecasting.

The meteorological data used in this study was obtained from the Shanghai Weather Station and Weather Underground website, coming from the two meteorogical station in Pudong and Hongqiao Airport in Shanghai. These weather profiles have been compared to the typical weather .epw file used in Energyplus [23] – the most popular building energy simulation software, results show that the weather patterns are similar, indicating the weather profile could be representative for the entire Shanghai district. These data were sampled every 30 min, and include dry bulb temperature, dew point temperature, relative humidity, moisture content, wind speed, and air enthalpy. To combine the power load data and weather data, the weather data were filled by applying the cubic spline interpolation method to each pair of adjacent data. Fig. 1 indicates the electricity load profiles of Shanghai in this study.

Fig. 1 (a) shows load curves in typical days of the summer, transition season and winter of 2014. There are two peaks and two valleys on a typical summer day, with large differences between morning and evening peak loads. Winter has similar characteristics, but the differences between the two peaks not as large as that in summer, and the winter evening peak occurred relatively early. The daily load in the transition season was significantly lower than those in winter and summer, and the load curve was relatively flat in the daytime. Because Shanghai is in a weather zone with hot summers and cold winters, the change of weather conditions gave rise to differences in load characteristics. As shown in Fig. 1 (b), the typical week load curve fluctuated periodically; the daily load on weekends was markedly lower than that of weekdays because of the social patterns in production and behavior.

Fig. 1 (c) illustrates the daily peak load on weekdays throughout the year. Electricity consumption was concentrated in July and August (the cooling season), when the daily load was much higher than in other months and reached a maximum of 26,560 MW on August 6. January, February and December (the heating season) follow July and August in electricity consumption, and the lowest daily load appears during February because this time represents the first week after the Spring Festival when production and behavior have not yet returned to normal levels. The daily load fluctuations of other months were relatively mild.

Fig. 1 (d) indicates peak load distribution throughout a year. The distribution frequency during different periods varied from month to month because different weather conditions over the year led to differences in power consumption behaviors. The distribution was



(a) Power load on a typical day

(b) Power load in a typical week

(c) Daily peak load over a year

(d) Daily peak Load distribution in a year

Fig. 1. Load profiles of Shanghai in 2014.

also different between weekdays and weekends. On weekdays, the peak load was concentrated from 10:00 to 11:00, except for July and August, and some peak loads appeared after 17:00 from November to January. For weekends, peak load often occurred after 17:00, mainly because household electricity consumption comprised large proportions in the non-working period. As the cooling season began, peaks became focused within the period of 12:00–16:00 when the temperature was very high.

Based on the above analysis, outdoor weather has a significant impact on the power load. Outdoor temperature influences the heat transfer of the building envelop and changes human behaviors, thus affecting the electricity consumption. The relationship between weather and electricity consumption shows complex nonlinear characteristics, which is difficult to be expressed by mathematical models. In the current study of load forecasting, only dry bulb temperature is treated as the main influencing factor.

Due to the heat transfer characteristics of buildings with surrounding air, the variation of cooling and heating load is delayed in comparison with outdoor weather parameters because of the thermal inertia of building envelopes. There exists attenuation and delay in the process of transforming the heat gained from the envelope into indoor cooling load. And the degree of the attenuation and the delay time are related to the thermal properties of building materials. The heat gain of the building envelope at τ moment in the Fourier series form is shown in the following equation [24]:

$$Q_\tau = hF\left[\overline{t_Z} - t_N + \frac{\alpha_N}{K}\sum_{n=1}^{m}\frac{\Delta t_{Z\cdot n}}{v_n}\cos(w_n\tau - \varphi_n - \varepsilon_n)\right] \quad (1)$$

Where, h is the heat transfer coefficient of the building envelop, $\mathrm{W}/m^2 \cdot K$; F is the area of the wall or roof, m^2; α_N is the heat emission coefficient of the building envelop; $\overline{t_Z}$ is the outdoor average temperature; t_N is the indoor temperature; $\Delta t_{Z\cdot n}$ is the hourly change of outdoor solar-air temperature, °C; w_n is the frequency of the outdoor solar-air temperature changes at order n; φ_n is the epoch angle of outdoor solar-air temperature changes at order n, deg or rad; v_n is the disturbance attenuation of solar-air temperature at order n; ε_n is the phase delay of the solar-air temperature at order n.

In harmonic response method that calculates cooling load, the relative lagging of the inner surface temperature wave to the external temperature is defined as the heat transfer delay time of the wall, denoted by ε_1, and the relative lagging of indoor temperature wave to the inner surface temperature is defined as the heat transfer delay time of the room, denoted by ε_2. The heat release characteristics of different types of rooms and their enclosure are summarized based on engineering experience, as shown in Table 1.

Table 1 shows that the delay time of light structure is 1.6 h and is 4.4 h $(\varepsilon_1 + \varepsilon_2)$ for heavy structure. So when analyzing the relationship between power load and weather factors, we selected the average weather parameter at the forecast time and the previous three hours to reflect the effect of delay more accurately. In addition, this method is further used to implement correlation coefficient analysis to verify its reasonability.

We already know that weather variation was the dominant factor driving load change. Therefore, the Pearson correlation coefficient (CCP) and Grey relational analysis (GRA) methods were implemented in this study to investigate the relationships between electric load and meteorological parameters. PPC describes the degree and direction of relativity between two variables based on calculating the correlation coefficient, which simplifies the correlation of the two variables into a linear relationship but is not comprehensive. Therefore, when the degree of correlation was low, GRA was used instead to demonstrate the associations of the variables.

July was chosen as a representative month of the cooling season, January for the heating season and April for the transition season to calculate correlation coefficients and perform GRA between daily power load and meteorological parameters. The median daily correlation coefficient for each month was then selected to represent the typical value for that month, as summarized in Table 2. Table 2 shows that the correlation coefficients of the cooling and heating seasons were high, whereas that of the transition season was relatively low, which further verified the rationality of selecting the average weather parameter at the forecast time and the previous three hours as the meteorological parameters.

3. Methodology

3.1. Seasonal attributes

In this paper, the seasons were redefined first because the hourly meteorological parameters were chosen as inputs instead of daily mean values to better represent the actual heating and cooling demands in buildings. Seasonal attribute means each day is judged to be belonging to one of the three categories: cooling, heating, and transition season by using certain criterion, according to the hourly meteorological parameters. Fig. 2 shows the flow chart of the determination of the daily seasonal attributes, as well as load decomposition. The hourly seasonal attributes were determined by the decision tree first, and then the daily seasonal attribute was determined based on the weighted mean seasonal attributes of different periods. The Effective Temperature [25]was chosen as the indicator in the decision tree, which considers the effect of relative humidity, wind speed and air enthalpy. Finally, the load was decomposed into basic load and weather-sensitive load based on daily seasonal attributes, followed by the load prediction for each load component.

3.1.1. Hourly seasonal attributes

In this study, hourly seasonal attributes were determined using a decision tree. Compared to other methods, the significant advantage of the decision tree is that it uses a white box model, so the decision tree output results could be easily understood. The classic algorithm ID3 (Iterative Dichotomiser 3 algorithm) [26] was implemented, which calculates the information gain [27] of each attribute and select the attribute with maximum information gain as the root node. Then the branch of the root node will be examined to capture the best attribute, and the similar procedure will be repeated till all attributes have been examined to formulate the decision tree. The advantage of this method is that the statistical nature of all training cases can be used to make decisions to resist noise.

Table 1
Different types of rooms and envelop characteristics [24]

Item	Light structure			Medium structure			Heavy structure		
	Floor	Ceiling	Wall	Floor	Ceiling	Wall	Floor	Ceiling	Wall
ε_1	1	0.8	1.5	2.2	0.6	2.8	3	1.8	2.9
ε_2	0.6	0.5	1.2	1.3	0.3	1.6	1.5	1.3	1.4

Table 2
Correlation coefficients between power load and weather parameters for representative months.

Method	Month	Temperature	Wind speed	Relative humidity
PPC	July	0.835	−0.362	−0.402
	January	0.711	0.566	−0.005
	April	0.572	0.362	−0.389
GRA	July	0.614	0.645	0.622
	January	0.605	0.683	0.514
	April	0.554	0.490	0.310

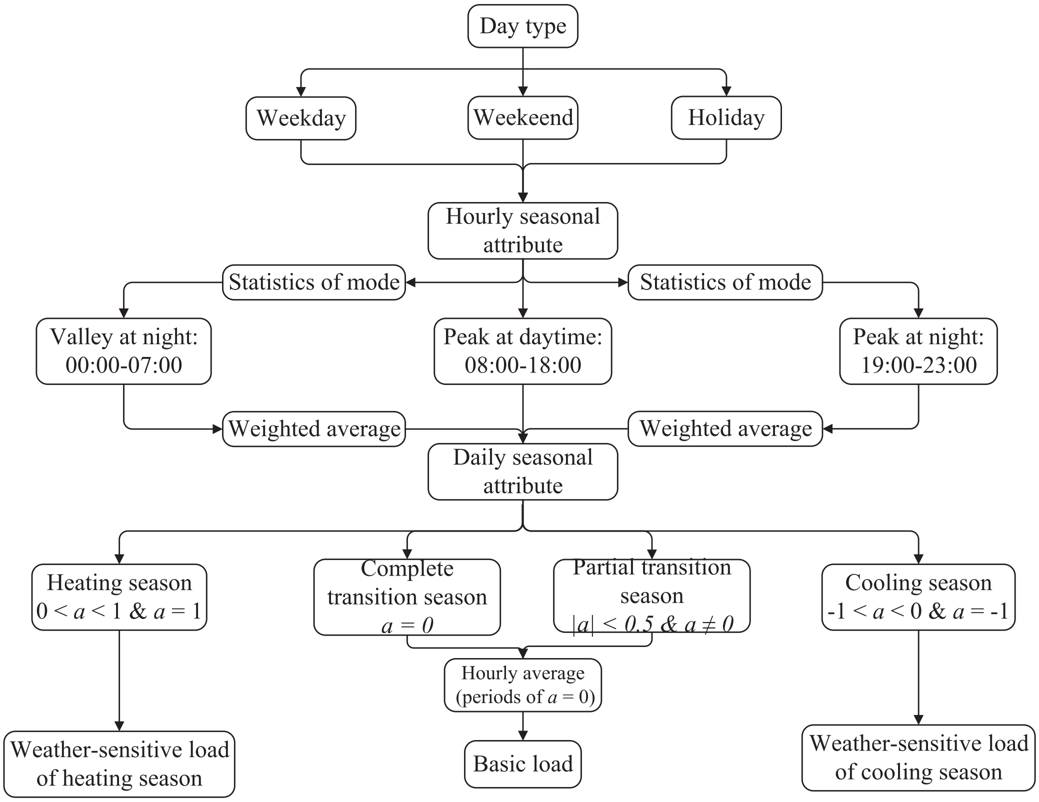

Day type
Weekday
Weekeend
Holiday
Hourly seasonal attribute
Statistics of mode
Statistics of mode
Valley at night: 00:00-07:00
Peak at daytime: 08:00-18:00
Peak at night: 19:00-23:00
Weighted average
Weighted average
Daily seasonal attribute
Heating season 0 < a < 1 & a = 1
Complete transition season a = 0
Partial transition season |a| < 0.5 & a ≠ 0
Cooling season -1 < a < 0 & a = -1
Hourly average (periods of a = 0)
Weather-sensitive load of heating season
Basic load
Weather-sensitive load of cooling season

Fig. 2. Seasonal attribute judgment and load decomposition methods.

Since only the maximum information gain of a single attribute is taken into consideration at each node, the effective temperature (ET) was chosen as input instead of dry bulb temperature to consider the combined effects of wind speed and relative humidity on temperature. Other input parameters were air enthalpy, humidity and dew point temperature. To reflect the lagging effect of weather parameters on load changes, average parameters for the forecast time and the previous three hours were chosen. The hourly seasonal attribute was then decided by using the resulting decision tree, which was defined as a variable a ($-1 \leq a \leq 1$), where it represent heating season if a is positive, cooling season if a is negative and $a = 0$ for transition season. If $a = 1$, it represents complete heating season, $a = 1$ means complete cooling season. And $0 < |a| < 0.5$ represents a partial transition season, while $0.5 < |a| < 1$ indicates a partial heating/cooling season.

3.1.2. Daily seasonal attributes

As shown in Fig. 2, each day was decomposed into three periods of time: a peak in the daytime within the period of 8:00–18:00, a peak at night within the period of 19:00–23:00, and a valley at night within the period of 0:00–7:00, according to the power load profiles in Section 2. Then the largest mode was chosen as the seasonal attribute for that period, followed by the weighted average seasonal attribute of each period, which was used as the daily seasonal attribute. The weight coefficients were determined based on the potential to use HVAC facilities as 0.45 for peaks and 0.1 for valleys during different periods based on engineering experience. If the weighted average $0 < |a| < 0.5$, the daily seasonal attribute was identified as a partial transition season, while partial cooling or heating season was defined if $0.5 < |a| < 1$. The traditional transition season was then subdivided into four parts: the complete transition season, partial transition season, partial cooling season and partial heating season based on weighted average daily seasonal attributes.

3.2. Load decomposition

As discussed in Section 2, the weight of the influence of meteorological factors on the power load varied in different seasons. The degree of correlation was relatively low during the transition seasons and high during the cooling and heating seasons. Therefore, the total load in the transition season was inferred to be insensitive to weather changes and dominated by the time schedule. This load was defined as the basic load. Then the total load in the cooling and heating seasons was divided into the basic loads and the weather-sensitive loads. Therefore, the average electric load of days with

a = 0 during the complete transition season was taken as the basic load, and the periods with a = 0 during the partial transition season was also added to the sample. Periods that did not have a = 0 during the partial transition season are divided into the basic and weather-sensitive components. For cooling/heating seasons, or the cooling/heating attributes during partial transition seasons, the difference between the total power load and the basic load component was evaluated to determine the weather-sensitive load. In this study, the weather-sensitive load is mainly decided based on the electricity consumption for cooling and heating since the influence of weather on the other types of load is not nontrivial, such as lighting load, hot water load, etc. Based on the Building America House Simulation Protocols [28], the normalized hourly lighting profiles for different months of a year are quite similar, indicating that the influence of the weather on the interior lighting is very small, and the lighting load is stable, especially for the inner zone which will not be largely affected by the outdoor environment. The same as the hot water load, there are certain hourly profile for each end use, including clothes washer, common laundry, dishwasher, shower, bath, sink, etc., as well as the combined profile.

Other traditional load decomposition methods were also implemented to be compared with the proposed load decomposition method, including EMD (Empirical Mode Decomposition) and WT (Wavelet Transform), in order to compare the performance of the proposed load decomposition method with that of the traditional methods.

3.3. Load prediction

The power load was divided into the basic and weather-sensitive components based on seasonal attribute analysis. The basic load was noticeably periodic and had consistent and stable characteristics for days of the same type. Therefore, a time-series algorithm was chosen to forecast the basic component, while SVM [2] was applied to forecast the weather-sensitive load component. For training dataset, the input parameters include effective temperature of the forecast time and the previous three hours, and the load at the corresponding time.

4. Seasonal attributes

4.1. Hourly seasonal attributes

In Shanghai, the heating season traditionally begins in January; the cooling season in July, and the transition season begins in both April and October. Therefore, the seasonal attribute of January was assigned as a = 1, and a = -1 for July. According to [2930], the Effective Temperature falling within the range of 15–23 °C ensures human thermal comfort; therefore, the seasonal attribute of periods when the ET falls within this range was assigned as 0 in April and October.

The hourly data for January and July were selected as the training data set, along with data for April and October when the Effective Temperature falls between 15 °C and 23 °C. For the remaining periods, the decision tree was used for further judgment. The rules for determining hourly seasonal attributes using the decision tree are shown in Fig. 3.

In Fig. 3, mET is the moving average Effective Temperature for four hours (°C), mEnth is the mean air enthalpy (kJ/kg), and md is mean humidity (g/kg). The precision of the decision tree classification model was 92.5%. The classification error was smallest for cooling season, which is 4%, however, it was 10% for transition season, and the error between the transition season and heating season was 7%. The possible reasons for this difference are that there are a lot of overlapping between weather parameters of the transition and heating seasons, and the number of branches on the left of the decision tree is relatively small. However, the precision level cannot be adjusted too high to avoid over-fitting.

4.2. Daily seasonal attributes

Table 3 shows the results of seasonal attributes from the proposed weighted mean method, as well as the comparison with the traditional degree-days method. Based on this new rule, the seasonal attributes of the traditional cooling, transition and heating seasons were determined more reasonably. For example, the daily mean air temperature on April 9 was 17.04 °C. According to the degree days method, this day falls into the heating season because the mean air temperature was lower than 18 °C [31]. In effect, the mean temperature in daytime was 17–24 °C, so that heating was only required at night. The seasonal attribute was assigned as a = 0.1 due to the weighted mean method, indicating partial transition season which is more accurate. Similarly, the cooling season can be separated from the transition season more specifically. For instance, the daily mean air temperature on June 6 was lower than 26 °C [31] and would be ascribed to the transition season based on the conventional degree days method, however, it is likely that cooling was required during the peak time. The seasonal attribute was a = $-$ 0.9 that indicate partial cooling season based on the new seasonal attribute's method. For other months, such as November, December, January, February, and early March, the seasonal attributes were attributed to the heating season, and July, August and early September were ascribed to the cooling season, which are identical to the results of the traditional method.

5. Load decomposition

5.1. Load data preparation

The existence of abnormal load data, including missing values and outliers, will influence the accuracy of predictions. Therefore, data should be processed before load decomposition to improve data quality.

According to the electric load analysis, the typical load curve was smooth and had no abrupt changes over time, and adjacent days with the same weather conditions should have similar load patterns. Accordingly, values with hourly jumps or distinct variations between similar days were considered outliers. Outliers do not necessarily indicate poor data. Whether a value was considered an outlier depended on different points of view. For example, if the load curve at a certain moment was smooth, then it could be considered a non-outlier from a local point; however, from a global point of view, the same data point may be abnormal compared with adjacent similar days and should therefore be considered an outlier. In this paper, load data were divided into two groups, workday data and non-workday data, which were detected from both local and global points of view. The three-sigma rule was used to detect the global outliers, which means the data is considered to be outliers if it lies in the region of values of the normal distribution of a random variable at a distance from its mathematical expectation of more than three times the standard deviation [32].

In addition to outliers, there are also missing values that were filled using the linear interpolation of adjacent values. After checking the whole data set, it was very likely that all values for that day (96 values in total) were global outliers. As shown in Table 4, the total number of workdays was 243 in 2014, and there were eight global outliers, which consisted of three days before the Spring Festival, two days after the Spring Festival, and one day before and two days after National Day. On these days, power load was lower

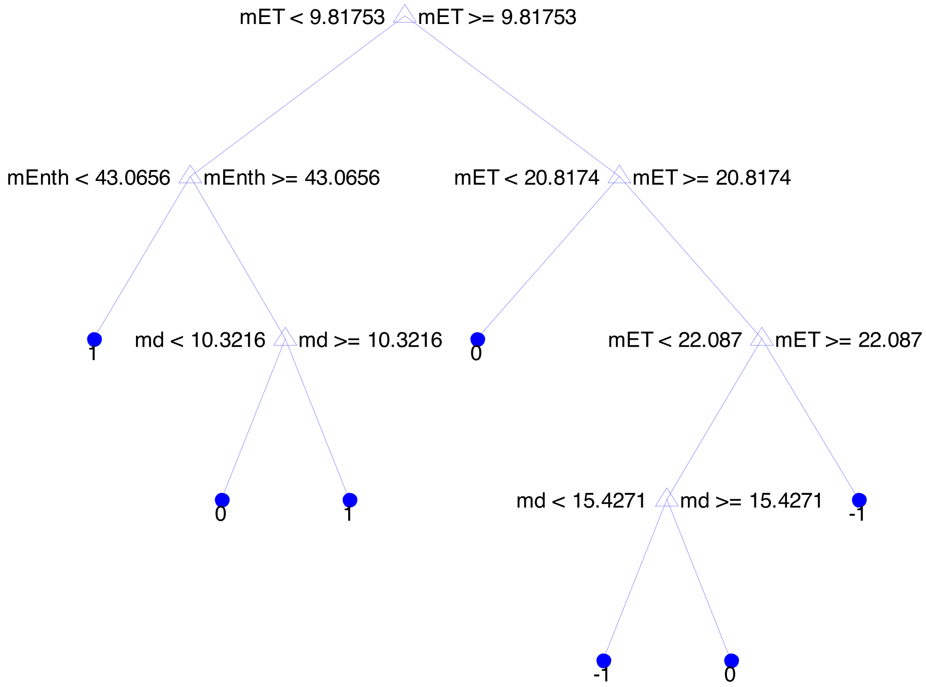


Fig. 3. Decision tree for hourly seasonal attributes.

Table 3
Daily seasonal attribute contrast of two methods.

Date	Daily mean temperature	Degree-days method	Weighted mean method
April 9	17.04 °C	1	0.1
May 9	18.28 °C	0	0.55
June 6	25.64 °C	0	−0.9
September 4	24.70 °C	0	−0.45
October 28	17.67 °C	1	0.1

Table 4
Global outliers.

Type of day	Number of days	Number of outlier days	Proportion
Workday	243	8	3.29%
Non-workday	122	3	2.45%
Total	365	11	3.01%

than on adjacent workdays. In addition, there were three global outliers out of the 123 non-workdays from January 31 to February 2 during the Spring Festival. Thus, we could draw a conclusion that holidays had a great influence on the characteristics of the power load, and that the load during holidays, as well as several days before and after holidays should be considered separately.

5.2. Power load decomposition

After processing the abnormal load data, the total power load could be decomposed into the basic and seasonal weather-sensitive components due to the seasonal attribute results, as shown in Fig. 4.

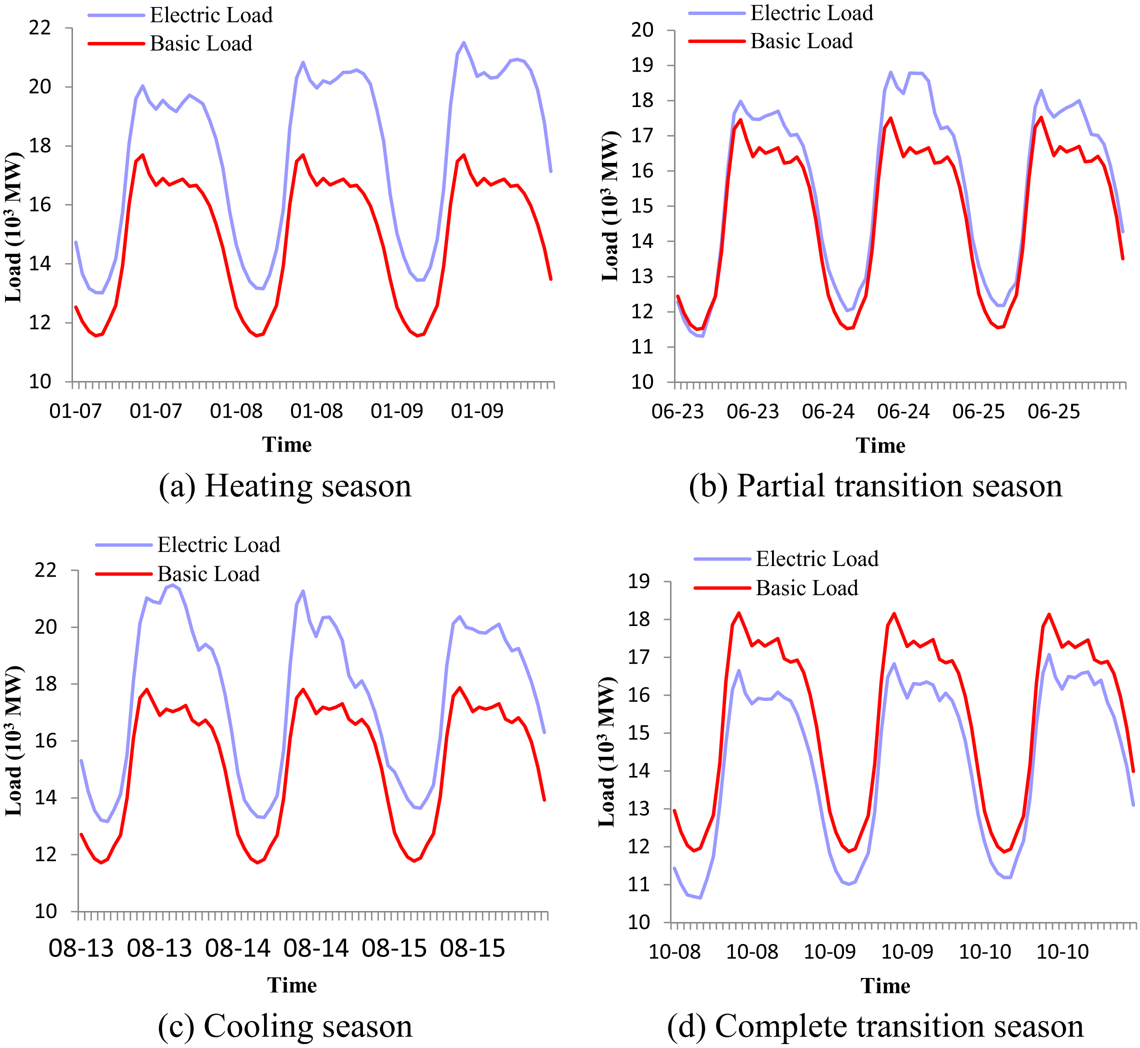
Taking the heating season (January, February and March) of 2014 for example, its basic load should be calculated based on the weighted average of the late transition season (September and October) in 2013. However, data from September and October 2014 were used instead because of a lack of data for 2013. For the cooling season of 2014 (July, August and early September), the basic component was calculated based on the weighted average of the near transition season (April and May) in 2014. For the heating season of November and December, the weighted average of the former transition season (April, May, September and October) was used. The weather-sensitive component was the difference between the total load and the basic component.

The above basic loads of the heating and cooling seasons were fixed for similar days in different months because the value was dependent on identical transition seasons. However, the basic component of the partial transition season ($0 < |a| < 0.5$) varied slightly with increasing number of samples, as illustrated in Fig. 4 (b).

Fig. 4 shows load decomposition on typical days with different seasonal attributes: heating, cooling and transition season. Red curves represent the basic load, and blue curves represent the total electricity load. The gap between the two curves is the weather-sensitive component. As illustrated in Fig. 4 (a) and Fig. 4 (c), the basic component remained unchanged between two similar days in the heating or cooling season.

In Fig. 4 (b), the basic load during the partial transition season varied slightly from day to day. The day peak from June 23 to 25 was classified as cooling season, whereas the nights of those days were classified as transition season. The basic component at night generally coincided with the total power load, and the weather-sensitive component appeared almost exclusively in the daytime. The basic load of June 23 was calculated based on the mean value of complete transition days just before that day. For June 24, hourly data with $a = 0$ on June 23 was also added to the samples, and for transition days approaching the heating season, the basic component during daytime coincided with the total load and the weather-dependent component, which appeared at night.

Fig. 4 (d) shows the load profile from October 8 to 10. These were days after National Day, and the overall electricity load was significantly lower than the ordinary level. Therefore, certain days before and after special holidays should be considered separately for load decomposition and load forecasting. The influential number of days could be decided by comparing the daily load before



(a) Heating season (b) Partial transition season (c) Cooling season (d) Complete transition season

Fig. 4. Load decomposition on typical days.

and after the holidays with previous workday load. If the absolute value of the changing rate is more than 4%, that day will be influenced by holidays, and the load will be scaled up or down to be close to the real load profile.

The difference between the basic load and total load is defined as the weather-sensitive component. According to the calculations, this component accounted for 6–15% of total electric load during the heating season and 10–26% of the total during the cooling season, which approximates the proportions of the HVAC load of building energy consumption; therefore, the decomposition method presented in this paper is reasonable. Load decomposition on non-workdays of different seasons was the same as that on workdays; therefore, there is no need to provide additional details.

5.3. Basic load adjustment of special days

Based on the above analysis, the global outliers of long holidays and days before and after these holidays need to be treated separately. Otherwise, abnormal situations like that shown in Fig. 4 (d) may occur.

To take the Spring Festival as an example, the daily mean load decreased linearly day by day during the festival and increased after the festival. In terms of the average load, the mean changing rate of the daily mean load on adjacent workdays and non-workdays were normally − 0.06% and − 0.28%, respectively. It was found that the mean changing rates of the daily mean load before and after long vacations became higher. For example, the total load on January 27 decreased by 14% from January 24 (the previous workday), and on the first day of the Spring Festival (January 31), the load was 29% lower than that of the previous day. Therefore, when the absolute value of the changing rate is more than 4%, it can be inferred that significant changes of load were

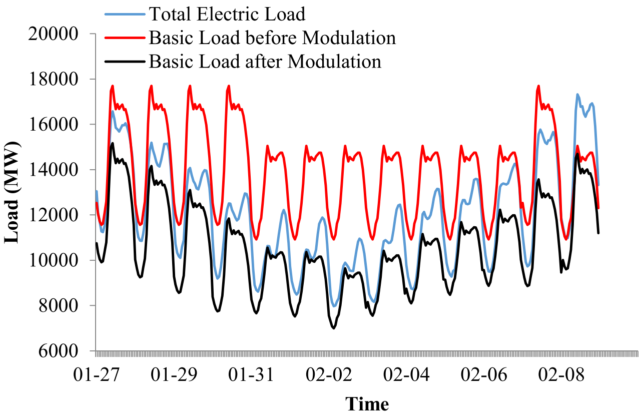


Fig. 5. Basic load adjustment of special days.

caused by holidays. The basic component during these special days would be scaled up or down after load decomposition, as shown in Fig. 5. The basic loads of special days after modulation were lower than the total electric load during the Spring Festival, however, it could still fully reflect the changing trend of the load pattern.

5.4. Characteristics of weather-sensitive load

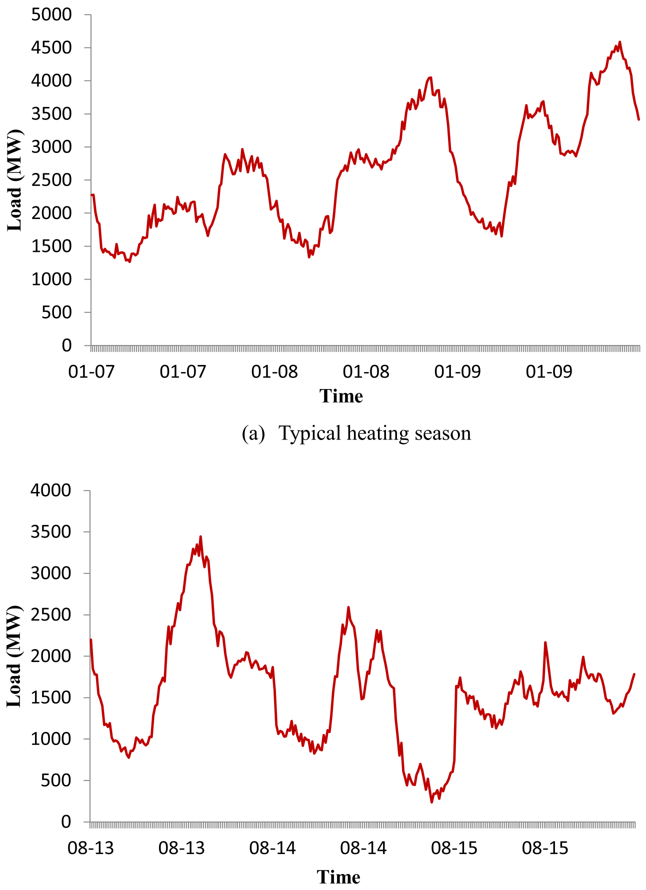
Weather-sensitive load, separated from the total load based on the above method, is shown in Fig. 6.

The properties of the weather-sensitive component included several aspects that differed from the total load:

1) The periodicity of change of the weather-sensitive load was weak between adjacent days and appears as a varying curve trend between two adjacent days in Fig. 6. This curve also highlights the influence of unpredictable weather conditions on electric load. The change patterns of power load were similar between similar days, because they were mainly affected by daily life.

2) The pattern of the weather-sensitive component became less smooth and more random, as shown in Fig. 6. The function of load decomposition was like a magnifier, such that the observation can be zoomed in to scale of 10^3 instead of 10^4. Therefore, the initially obscure fluctuation at the scale of 10^4 was manifested at the 10^3 scale.

3) The daily load pattern of weather-sensitive component differed from that of the power load. For example, the peaks appeared at different times. For the total load, the daytime peak was essentially consistent with the nighttime peak during heating seasons. For the weather-sensitive component, Fig. 6 shows that the nighttime peak is noticeably larger than the daytime peak, which agreed with the power load pattern during cooling seasons.



(a) Typical heating season

(b) Typical cooling season

Fig. 6. Weather sensitive load profile during typical heating and cooling season.

6. Load prediction

6.1. Model Evaluation index

After models were established, the prediction performance was evaluated, including the fitting degree and prediction accuracy.

The fitting degree was expressed by coefficient of determination, denoted by R^2. R^2 lies between 0 and 1. The fitting degree becomes higher with the value of R^2 approaching 1. The formula for R^2 calculation is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

RMSE (Root mean square Error) was used to characterize the differences between predicted values and the observed values. The smaller the RMSE value is, the higher the model accuracy is. The formula for calculation is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} \tag{3}$$

MAE (Mean absolute error) was regularly employed to tell the difference between predicted and observed values, which prevented the offset of positive–negative error, which is given by:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|x_i - \hat{x}_i|}{n} \tag{4}$$

MAPE (Mean absolute percentage error) is another indicator of prediction accuracy. We can compare the accuracy of different models based on data with different orders of magnitude. MAPE is given by:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - \hat{x}_i|}{x_i} \tag{5}$$

Where x_i is observed values, $\hat{x}_i$ is predicted values, n is the sample size.

6.2. Model evaluation

As described in previous sections, the electricity load was divided into basic and weather-sensitive components based on seasonal attributes. Different models were compared, and these models are listed in Table 5. In this table, w, w', and s denote weekday, weekend and special day, respectively, and h and c denote the heating and cooling seasons respectively.

Table 5
Models of power load forecasting.

Day type	Seasonal attribute	Basic load	Weather-sensitive load
Weekdays / weekends	Heating season	Weighted mean method	SVR_wh/ w'h
	Cooling season	Weighted mean method	SVR_wc/ w'c
	Transition season	Time-series method	SVR_wh/wc/ w'h/ w'c
Special days	Heating season	Weighted mean method	SVR_sh
	Cooling season	Weighted mean method	SVR_sc
	Transition season	Time-series method	SVR_sh/sc

Table 6
Performance of models of basic load.

Evaluation index	Workday	Non-workday
R^2	0.921	0.953
RMSE	312.450	343.390
MAE	255.180	265.370
MAPE	2.03%	1.75%

First, the model established for basic load based on a time-series method was evaluated, using various indicators. The 96 electric load data points for March 24 as a workday and May 25 as a non-workday were calculated using the time-series method. The calculated results were then compared with the actual load data, as shown in Table 6. Generally, applying the time-series method to forecast the basic load achieved good prediction accuracy with MAPE<2.5%, and the R^2 value was also high. These findings indicate that the imitation degree of the model was good.

The ensemble prediction method was then used to compare with the traditional method, including load forecasting based on total electric load and other load decomposition method EMD and WL, to evaluate the performance of this new load decomposition method.

Forecasting models were built based on both the load decomposition method presented in this paper and the traditional methods. For workdays, taking the heating season and cooling season for examples, the training set for the heating season from January 7 to 16 was used to forecast the power load from January 17 to 20, and the training set for the cooling season from July 24 to August 6 was chosen to forecast the power load of August 7 and 8. For non-work days, models were established for the Spring Festival and an ordinary weekend of the cooling season to evaluate performance. The performance of the different forecasting models is summarized in Tables 7, which covers different day type in each season, including workday during heating season, workday during cooling season, special holiday during heating season, and weekends during cooling season.

As shown in Table 7, for workdays, the RMSE and MAE of the model based on the new load decomposition method were lower than that of the traditional method. Compared to load forecasting method with total electric load, MAPE decreased from 4.08% to 1.91% by using the ensemble model during the heating season, and MAPE of the cooling season dropped from 3.66% to 1.65%. Besides, results show that the ensemble algorithm performs better than the other two decomposition methods EMD and WL, and less training time is required for the ensemble model.

For special holidays and weekends, the new ensemble method based on load decomposition performs superior to the traditional methods in terms of RMSE, MAE, MAPE and training time. However, the basic load on special days after adjustment could not perfectly reflect this special characteristic when peak load at night was much greater than that during daytime. The ensemble model was also unable to reflect this characteristic. The R^2 value of the traditional model based on the power load was higher, and MAPE was only slightly lower by using the ensemble model. The main reason may be because the data samples for long holidays are too few and the load pattern was noticeably different from those of other day types, it would be better to use the traditional model based on total power load to forecast electric load. However, for most other weekends, the ensemble model tends to perform better than others, and MAPE decreased to 3.09%.

Table 7 shows that the proposed load decomposition method and the ensemble algorithm perform better than the other two load decomposition schemes. Although the electric load can be divided into low and high frequency components from the signal field by the two methods, the components cannot exactly be explained by physic laws. It is also seen that the MAPE of EMD_SVR_w'c model is much bigger than that of others, indicating that the EMD_SVR model's prediction accuracy relies much more on strong regularity of the load itself. On the contrary, the new load decomposition method was constructed based on realistic conditions and physical laws behind the power load, which could avoid the excessive reliance on the load data itself. Results show that the forecasting accuracy is significantly improved.

7. Discussion and conclusions

Improving the precision of short-term load forecasting has long been a focus of researchers. In this article, the electricity load characteristics of the Shanghai metropolitan area were first analyzed, the load decomposition approach was put forward, and an ensemble forecasting model was developed. The results of the ensemble model show high accuracy and superior applicability for weekdays, weekends, and holidays in different seasons. The main conclusions are as follows:

1) A new load decomposition method was developed to divide the electric load of the heating and cooling seasons into basic and weather-sensitive components effectively. Unlike the traditional degree-day's method, a decision tree was built to determine hourly seasonal attributes, and daily seasonal attributes were then calculated based on the weighted averages of different periods of time throughout the day. After the seasonal attribute judgment,

Table 7
Performance of the four models on different day types and seasons.

Period	Evaluation index	R^2	RMSE	MAE	MAPE	Training time	
Workday during Heating Season	Traditional method	SVR_wh	0.933	740.83	664.88	4.08%	1 min 48 s
	EMD	EMD_SVR_wh	0.919	714.98	572.31	3.44%	10min20s
	WL	WL_SVR_wh	0.856	955.25	833.1	5.03%	20min22s
	Proposed method	P_SVR_wh	/	408.9	332.54	1.91%	1 min 33 s
Workday during Cooling Season	Traditional method	SVR_wc	0.967	997.24	820.64	3.66%	1 min 48 s
	EMD	EMD_SVR_wc	0.501	2457.84	1781.23	7.84%	11 min
	WL	WL_SVR_wc	0.8863	1173.64	779.64	3.69%	19min34s
	Proposed method	P_SVR_wc	/	464.82	362.31	1.65%	1 min 33 s
Special Holidays on Heating Season	Traditional method	SVR_sh	0.868	587.63	496.61	4.30%	1 min 21 s
	EMD	EMD_SVR_sh	0.835	582.18	428.93	4.03%	6 min
	WL	WL_SVR_sh	0.826	597.12	465.51	4.56%	6 min
	Proposed method	P_SVR_sh	/	337.52	336.85	3.98%	1 min 9 s
Weekends on Cooling Season	Traditional method	SVR_w'c	0.886	923.42	809.26	4.36%	35 s
	EMD	EMD_SVR_w'c	0.359	2016.95	1669.48	10.33%	15min50s
	WL	WL_SVR_w'c	0.938	625.6	570.01	3.30%	9min34s
	Proposed method	P_SVR_w'c	/	558.81	442.54	3.09%	27 s

the load of the transition season was chosen as the basic component based on power load characteristic, and the difference between total load and the basic component was separated out as the weather-sensitive component.

2) The characteristics of the basic load during the transition season tended to be steady and had apparent periodic cycles. Regression models based on the time-series method were built to forecast the basic load. The results indicate that this method ensured high prediction accuracy with MAPE of<2.5%. SVM was employed to forecast the weather-dependent components during the heating season and cooling season because SVM can approximate nonlinear functions with great precision. The predicted total electric load was the sum of the predicted basic load and the weather-sensitive load. Compared with the traditional model based on daily overall load and other decomposition methods (EMD and WL), this ensemble model reduced error from 3 to 5% to lower than 2% when forecasting the power load of workdays. For non-workdays, the error was decreased from 4 to 5% to lower than 4%.

The loads on special event days are still hard to forecast. For example, on Spring Festival and National Day, the load was scaled up or down the trend which cannot reflect the variation pattern of the overall load. In addition, the impact of the sudden change of the weather on power load was not considered in this paper. The load data of similar days with sudden weather changes should be collected over the years to research the relationship between the degree of meteorological parameters and power load. The solar irradiance not only varies with seasons, it is also an important factor influencing human electrical behavior. Due to the lack of sources of solar irradiance, its relationship with power load has not been researched, which can be supplemented in the following studies.

CRediT authorship contribution statement

Yiyi Chu: Conceptualization, Methodology, Software, Writing - review & editing. **Peng Xu:** Conceptualization, Supervision. **Mengxi Li:** Investigation, Formal analysis. **Zhe Chen:** Software, Validation. **Zhibo Chen:** Software, Validation. **Yongbao Chen:** Writing - review & editing. **Weilin Li:** Data curation.

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.

Acknowledgements

The work was supported by the Research Support Scheme of the National Key R&D Program of China and Shanghai Key Laboratory of Smart Grid Demand Response, Grant NO. 2018YFC0705900.

References

[1] Kyriakides E, Polycarpou M. Short term electric load forecasting: A tutorial. Trends in Neural Computation. Springer Berlin Heidelberg, 2007: 391-418.
[2] 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, Appl. Energy 1 (195) (2017 Jun) 659–670.
[3] Bunn D, Farmer E D. Comparative models for electrical load forecasting. 1985.
[4] Alex D. Papalexopoulos, Timothy C. Hesterberg, A regression-based approach to short-term system load forecasting, IEEE Trans. Power Syst. 5 (4) (1990) 1535–1547.
[5] D.G. Infield, D.C. Hill, Optimal smoothing for trend removal in short term electricity demand forecasting, IEEE Trans. Power Syst. 13 (3) (1998) 1115–1120.
[6] E.H. Barakat et al., Short-term peak demand forecasting in fast developing utility with inherit dynamic load characteristics. I. Application of classical time-series methods. II. Improved modelling of system dynamic load characteristics, IEEE Trans. Power Syst. 5 (3) (1990) 813–824.
[7] K.B. Song, Y.S. Baek, D.H. Hong, et al., Short-term load forecasting for the holidays using fuzzy linear regression method, IEEE Trans. Power Syst. 20 (1) (2005) 96–101.
[8] H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural networks for short-term load forecasting: A review and evaluation, IEEE Trans. Power Syst. 16 (1) (2001) 44–55.
[9] N.I. Sapankevych, R. Sankar, Time series prediction using support vector machines: a survey, IEEE Comput. Intell. Mag. 4 (2) (2009).
[10] Ye Ren et al., Random vector functional link network for short-term electricity load demand forecasting, Inf. Sci. 367 (2016) 1078–1093.
[11] Fatemeh Chahkoutahi, Mehdi Khashei, A seasonal direct optimal hybrid model of computational intelligence and soft computing techniques for electricity load forecasting, Energy 140 (2017) 988–1004.
[12] D. Liu, W. Wang, J. Liu, Sensitivity analysis of meteorological parameters on building energy consumption, Energy Procedia 1 (132) (2017 Oct) 634–639.
[13] T. Hong, Short term electric load forecasting, North Carolina State University, 2010.
[14] 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, Appl. Energy 1 (195) (2017 Jun) 659–670.
[15] L. Ghelardoni, A. Ghio, D. Anguita, Energy load forecasting using empirical mode decomposition and support vector regression, IEEE Trans. Smart Grid 4 (1) (2013) 549–556.
[16] Fan X, Zhu Y. The application of empirical mode decomposition and gene expression programming to short-term load forecasting[C]. Natural Computation (ICNC), 2010 Sixth International Conference on. IEEE, 2010, 8: 4331-4334.
[17] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[C]. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society, 1998, 454(1971): 903-995.
[18] Zhu Z, Sun Y, Li H. Hybrid of EMD and SVMs for short-term load forecasting[C]. Control and Automation, 2007. ICCA 2007. IEEE International Conference on. IEEE, 2007: 1044-1047.
[19] G.F. Fan, S. Qing, H. Wang, et al., Support vector regression model based on empirical mode decomposition and auto regression for electric load forecasting, Energies 6 (4) (2013) 1887–1901.
[20] N. An, W. Zhao, J. Wang, et al., Using multi-output feed-forward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting, Energy 49 (2013) 279–288.
[21] G.F. Fan, L.L. Peng, W.C. Hong, et al., Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression, Neurocomputing 173 (2016) 958–970.
[22] Jinliang Zhang et al., Short term electricity load forecasting using a hybrid model, Energy 158 (2018) 774–781.
[23] Energyplus. https://energyplus.net/
[24] Rongyi Zhao, Cunyang Fan, Dianhua Xue, et al. Air Conditioning (The Fourth Edition) . China Architecture & Building Press,2009.
[25] Standard, A.S.H.R.A.E., 2017. Standard 55-2017. Thermal environmental conditions for human occupancy. ASHRAE, Atlanta, GA, 30329-2305.
[26] MargaretH.Dunham et al. Data Mining: Introductory and Advanced Topics. Tsinghua University Press, 2005.
[27] Yuzi Luo, Fu Xinghong, Data Mining ID3 Decision Tree Classification Algorithm and its Improved Algorithms, Comput. Syst. Appl. 22 (10) (2013) 136–139.
[28] Hendron R, Engebrecht C. Building America house simulation protocols. Office of Energy Efficiency and Renewable Energy (EERE), Washington, DC (United States); 2010 Sep 1.
[29] Wei Zhang, Zhi Gao, Wowo Ding, Outdoor thermal comfort indices: a review of recent studies, Environ Health 32 (09) (2015) 836–841.
[30] Da'si He. Study on Several Issues Relating to Meteorological Data in HVAC Field, Tongji University, 2006.
[31] JGJ 134-2012. Design standard for energy efficiency of residential buildings in hot summer and cold winter zone.
[32] Smirnov NV, Dunin-Barkovskii IV. Mathematische Statistik in der Technik (translated from Russian). Verlag Wissenschaft. 1969.