

A multi-information fusion model for short term load forecasting of an architectural complex considering spatio-temporal characteristics

Jiangjian Xie^{a,b}, Yujie Zhong^a, Tong Xiao^a, Zheng Wang^c, Junguo Zhang^{a,*}, Tuowai Wang^a, Björn W. Schuller^{b,d}

^aSchool of Technology, Beijing Forestry University, Beijing, PR China

^bChair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany

^cSchool of Electrical and Electronic Engineering, North China Electric Power University, Beijing, PR China

^dGLAM, Imperial College London, UK

1. Introduction

In 2019, the energy consumption and carbon emission of China's buildings in operation are 24.5 % and 21.5 % of the total for the country, respectively [1]. Energy savings in buildings have emerged as a critical area for carbon neutrality. Realizing energy conservation and carbon emission reduction of buildings is of great significance to reduce global energy consumption and carbon emission. In recent years, with the rapid development of higher education, the number of universities and students is increasing year by year. The number of campus buildings is growing, and so is the amount of energy required for their development. Therefore, the university campus offers a lot of potential for energy savings as a typical regional architectural complex. Electric energy is an important part of energy consumption in university buildings.

Accurate short-term load forecasting is the basis of reasonable energy dispatching and optimal operation of buildings [2].

There have been several load forecasting methods presented in recent decades. Generally, these methods are mainly divided into conventional statistical methods and machine learning methods. Conventional statistical models are represented by time series analysis and regression analysis [3]. Autoregressive Integrated Moving Average (ARIMA) [4] is a commonly used time series analysis method, which builds the mathematical model to describe the change of load based on the historical load data and forecast the future load. Regression analysis seeks to build a regressive equation based on historical data to determine the link between load and influencing factors. Theoretically, statistical models have difficulty capturing the abrupt variation among time series due to their linear definition.

Machine learning methods can learn the nonlinear relationship between the inputs and outputs, which makes them achieve great success in load forecasting. Support Vector Regression (SVR) [5],

* Corresponding author.

E-mail address: zhangjunguo@bjfu.edu.cn (J. Zhang).

regression trees, Artificial Neural Networks (ANNs), etc., have been introduced for load forecasting. According to different complicated scenes, ensemble models like Bagged Regression Trees (BRT) [6] and Random Forests (RF) [7] were utilized to improve the effectiveness. An ANN uses multiple neurons to simulate the human brain, which enables them to provide excellent learning ability of non-linear features and leads to good generalization capability. There have been several applications of ANN for load forecasting, including Multi-layer Perceptron (MLP) [8,9], Backpropagation Neural Networks (BPNN) [10], Echo State Networks (ESN) [11], Evolutionary Neural Networks (ENN) [12], and Generalized Regression Neural Networks (GRNN) [13–15]. Compared with the ANN, deep learning models boost nonlinear fitting ability by expanding hidden layers or stacking some particularly built structures, which demonstrate strong performance in load forecasting and are growing frequently employed in present. The most popular used deep learning model in load forecasting is the Recurrent Neural Network (RNN), which is a type of neural network well-suited to model time series data. To overcome the difficulties in learning of long-term dependencies, two kinds of improved RNN, the long-short term memory (LSTM) network [16] and Gated Recurrent Units (GRU) network [17,18] have been more popular used in load forecasting.

There are different advantages for different forecasting methods. Hence, model fusion is introduced to improve the forecasting performance, such as combining SVM, RF, and LSTM [19], integrating a Convolutional Neural Network (CNN) and LSTM [20] or GRU [21], cascading a multi-channel convolutional neural network (MCNN), and LSTM [22], fusing a CNN and a Temporal Convolutional Network (TCN) [23] and introducing ensemble learning to fuse several sub-learners [24].

Although several approaches exist for extracting the non-linear time-series features of the load well enough to make reasonable forecasts, they can only be used to forecast the specific building load at a time. People flowing between multiple buildings in a university architectural complex result in loads of these buildings having certain spatial correlation features. That is to say, the university architectural complex loads exhibit considerable dynamics in both the spatial and temporal dimensions. The aforementioned load forecasting methods neglect the spatial features. It is thus more advantageous to present a novel model that fully utilizes both spatial and temporal features, with the goal of forecasting loads for all buildings in an architectural complex at the same time. However, how to extract the inherent spatio-temporal characteristics of non-linear and complicated load data and achieve an accurate building load forecasting remains a challenge.

The graph neural network (GNN) extends the neural network paradigm to deal with the data represented in the graph. It can learn and represent unstructured data well, which is also commonly employed in structured scenarios [25]. In the field of power prediction, GNNs are mainly applied to renewable energy generation forecasting [26–29], but there are few applications of GNNs for load forecasting. A graph convolutional network (GCN) is a kind of Graph Neural Network (GNN) that has been frequently utilized to convey spatial correlation of graph-based data, and it can extract the spatial characteristics by aggregating the neighborhood information of nodes through graph convolution [30].

Aiming to make full use of the spatial and temporal features of a university architectural complex load, we introduce a GCN to build a novel deep learning model: a multi-information fusion model combining spatio-temporal attention with GCN and LSTM (MF-STAGL) to realize simultaneous and high-accuracy load forecasting of each building in a university architectural complex. This model can process the load data directly on the original graph-based network and effectively capture the dynamic spatio-temporal features. To the best of our knowledge, this is the first time GCN has been introduced to forecast the architectural complex loads.

The main contributions of this paper are summarized as follows:

- (1) Considering the spatio-temporal characteristics of an architectural complex loads, we propose a multi-information fusion model combining spatio-temporal attention with GCN and LSTM (MF-STAGL). Specifically, GCN is introduced to extract spatial characteristics from the relationships between buildings in an architectural complex, whereas LSTMs are used to explain the temporal characteristics of building loads.
- (2) The spatio-temporal attention modules are incorporated to extract dynamic spatio-temporal correlations of load data. Spatial attention is employed to represent the complicated spatial relationships between different buildings. Temporal attention is utilized to extract dynamic temporal correlations between different timesteps. The adoption of the attention mechanism aids in maintaining considerable forecasting stability as the forecast period extended.
- (3) Extensive experiments using actual load data of a university architectural complex confirm that our model achieves the best forecasting performances when compared to the existing models.

The rest of the paper is organized as follows: In [Section 2](#), the key factors are decided by Person Correlation Analysis; In [Section 3](#), we propose the short-term load forecasting model based on the GCN-LSTM; In [Section 4](#), experimental data processing is described in detail; In [Section 5](#), experiments are performed using actual load data of a university architectural complex, and the results & discussions are presented. The conclusions are provided in [Section 6](#).

2. Key influencing factor analysis of load

2.1. Selection and qualification of influencing factors

Many factors are affecting the building load. In addition to the general influencing factors, there are certain changes in the selection of influencing factors for various geographical regions, scopes, and types of loads.

The loads of university buildings are not affected by the step tariff, which is different from that of the residential and commercial buildings. To build multi-information fusion and increase the accuracy of short-term load forecasting, it is required to select the key factors impacting load fluctuation from numerous alternative factors using correlation analysis. We take the load forecasting of a public university in Beijing as an example, five representative buildings: Second Canteen, Second Teaching Building, Experimental Building, Dormitory Building 7, and Main Building are selected for our study. Considering the building operating conditions, the following four types of influencing factors are chosen.

(1) Historical load type

Since the load data is essentially a time series, the load data before the forecasting time has a great impact on the current load. To account for its variance, the maximum load and minimum load of the day before are selected.

(2) Temperature type

The influence of temperature on the load is obvious. Temperature alters cooling and heating demands, resulting in loads generated by the usage of air conditioners, electric heaters, and other electrical equipment. As shown in [Fig. 1](#), the maximum load and

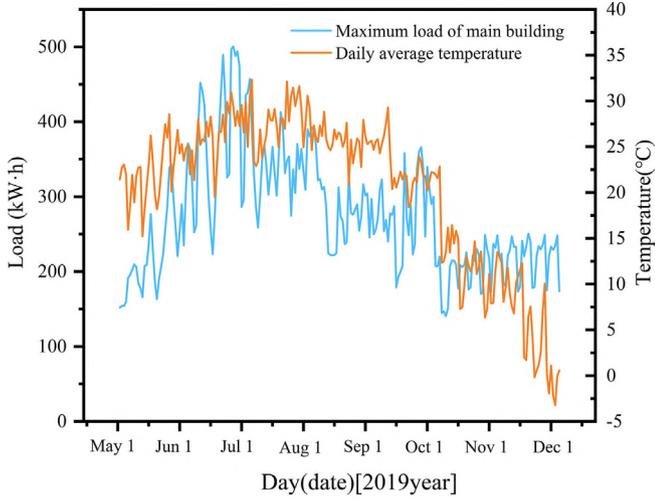


Fig. 1. Relationship between daily average temperature and daily maximum load of the Main Building from May to December.

daily average temperature trends in the main building are almost comparable. In Beijing, the highest temperatures are common in June and July, when the load is at its peak. Because air conditioner is no longer required in the fall, the maximum load decreases as the temperature drops. In the winter, coal-fired heating is used to compensate for temperature fluctuations, and the load remains low and relatively stable. Therefore, hourly temperature and daily average temperature are selected.

(3) Meteorology type

Meteorology conditions may affect the load. On cloudy and rainy days, for instance, students will go out less, and the dim light will increase the electricity consumption of indoor lighting equipment. As a result, on average, the load on cloudy and rainy days is greater than that on sunny days. Furthermore, in recent years, the air quality has gradually attracted the attention of residents. When air pollution is serious, the load of air purification equipment increases, leading to the increase of load. Finally, the weather type, season, wind speed, and air quality index are selected.

The quantification criteria attempt to depict the growing and decreasing features of the load under various types. We analyzed the load variation under various types, and because the quantified values must be normalized, we simply used discrete natural numbers to represent the various types. The quantified value of the corresponding type increases as the load increases. The detailed weather types and corresponding quantified values are listed in Table 1. Similarly, Table 2 presents the seasonal types.

(4) Date type

The influence of date type on load varies between workdays and weekends. Classifying the historical load data by date type will aid in improving the accuracy of short-term load forecasting. In each week, it is found that the loads on weekdays are higher than those on weekends. In addition, holidays have a significant impact on the decline of the load. Most students return home during the winter

Table 1
Quantified values of weather types.

Weather	Sunny	cloudy	light rain	moderate rain	heavy rain	snow
Value	1	2	3	4	5	6

Table 2
Quantified values of seasonal types.

Season	spring	summer	autumn	winter
Value	2	4	3	1

and summer vacations, the load naturally lowers. Hence, the winter and summer vacations are treated specially.

In addition, according to the load curve analysis, the rest day load for the Teaching & Office Building is much lower than the weekday load, the Monday load is impacted by the Sunday load, and the Friday load differs from the general weekday load (Tuesday to Thursday) owing to the arrival of Saturday. The load in the Dormitory Building is notably different from that in the Teaching and Office Building, with the total daily load level being lower than the rest day. Hence, the university buildings are divided into two categories: Teaching & Office building and Dormitory Building. Different quantification criteria are utilized for different categories. To summarize, the quantified values are proportional to the load. The detailed date types and corresponding quantified values are listed in Table 3.

2.2. Selection of key factors based on Pearson correlation analysis

We choose 30 weeks of data from May to November 2019 to do the Pearson correlation analysis between the selected factors and the load. Further, the Student's *t*-test is utilized to measure the representativeness of the sample to the overall correlation degree. The results are listed in Table 4.

According to Table 4, considering both the magnitudes of the Pearson correlation coefficients (larger than 0.2) and the significances (smaller than 0.01), five factors, i.e., the maximum and minimum load of the day before, hourly temperature, weather type, and daily type are selected as the key influencing factors, which are combined with the historical load to form the multi-information.

3. Multi-information fusion short-term load forecasting model

3.1. Spatio-temporal feature of a university architectural complex loads

We explore the spatio-temporal correlation using the loads of five selected buildings on June 12th and 13th, 2019. Fig. 2 presents

Table 3
Quantified values of date types.

Date type	Teaching & office building	Dormitory Building
Monday	0.8	0.4
Tuesday	0.7	0.4
Wednesday	0.7	0.4
Thursday	0.7	0.4
Friday	0.7	0.4
Saturday	0.5	0.7
Sunday	0.5	0.7
Mini-break	0.4	0.8
Winter and summer vacation	0.2	0.2

Table 4
Correlation analysis results.

Type	Factor	Pearson correlation coefficient	Significance
Historical load	Maximum load of the day before	0.651**	0
	Minimum load of the day before	0.642**	0
Temperature	Hourly temperature	0.732**	0
	Daily average temperature	0.220	0.090
Meteorology	Weather type	0.511**	0
	Season	0.326	0.060
	Wind speed	0.023	0.844
	Air quality index	0.032	0.887
Date	Date type	-0.210*	0.008

* indicates a significant correlation at 0.05 level (bilateral); ** indicates a significant correlation at the 0.01 level (bilateral). The selected key factors are marked in Bold.

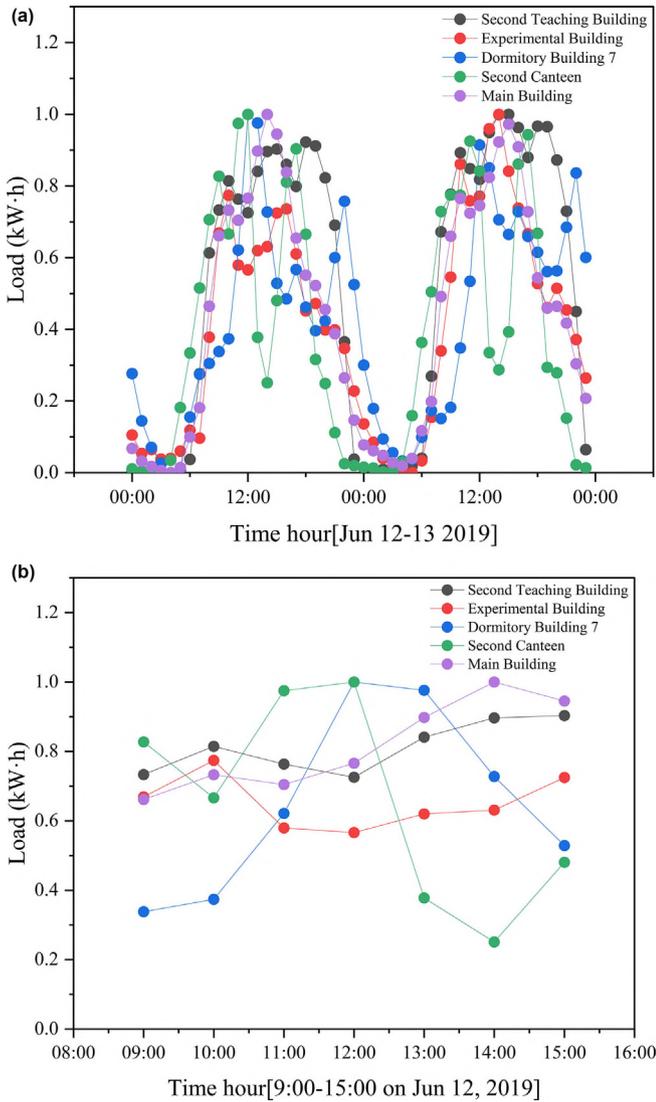


Fig. 2. Normalized loads of five selected buildings.

the normalized loads of these buildings. It is discovered that there is a clear temporal characteristic for the loads of each building (cf. Fig. 2(a)). Furthermore, the loads of different buildings exhibit a specific spatial correlation at the same period. Taking the loads of 12th as example, from 10:00 to 14:00 (cf. Fig. 2(b)), there are

the university's busiest hour for people flow. After class, students eat in the canteen before returning to their dormitories, which is followed by the afternoon's classes. As a result of this circumstance, the loads of the Second Teaching Building and Experimental Building decrease and then increase. Conversely, the loads of the Second Canteen and Dormitory Building 7 fluctuate in the opposite trend, increasing and then decreasing. As for Main Building (an office building for faculty), the load decreases slightly when faculty go to the canteen after work. However, the load will continue to increase throughout the afternoon, which is when most meetings take place. Altogether, the load will alter in response to the people flow, resulting in spatial correlation between the loads of different buildings in the architectural complex.

In summary, for the architectural complex loads, there are not only temporal features for each building, but also certain spatial features among different buildings. For this reason, a multi-information fusion model combined with spatio-temporal attention, GCN, and LSTM (MF-STAGL) is designed to realize the short-term load forecasting of the architectural complex. Fig. 3 shows the overall framework of the MF-STAGL model.

The proposed model consists of a feature fusion layer, two STAGL blocks and a fully-connected (FC) layer. The feature fusion layer is used to fuse the input multi-information. In each STAGL block, there are a spatio-temporal attention module and a GCN-LSTM module. To optimize the training efficiency, we adopted a residual learning framework [31] in each block. Finally, the predicted loads of all buildings in the architectural complex can be simultaneously achieved from the outputs of FC. The main modules are described in detail in Section 3.2 and 3.3.

3.2. Graph convolutional network with spatio-temporal attention

3.2.1. Basic graph convolutional network

In this study, we define the buildings of the university architectural complex as the nodes in a graph $G = (V, E, A)$, where V is a set of N nodes; E is a set of edges that depict the relationship between the buildings; $A \in R^{N \times N}$ indicates the adjacency matrix of graph G . Each timestep contains load and key factor data for each node, which may be seen as input feature vectors.

Considering the people flows between the buildings of the university architectural complex, load data of the buildings can be seen as nature graph structure data (shown in Fig. 4). The spectral-based GCN is adopted to extract the topological properties of university architectural complex loads in the spatial dimension.

In a spectral-based GCN, a graph is represented by its corresponding Laplacian matrix L , which is defined as $L = D - A$, and the normalized Laplacian matrix L_N is

$$L_N = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \quad (1)$$

where I_N is a unit matrix, and the degree matrix $D \in R^{N \times N}$ is the diagonal matrix, $D_{ii} = \sum_{j \in N} A_{ij}$.

The eigenvalue decomposition of the Laplacian matrix is $L = U \Lambda U^T$, where U is the orthogonal matrix composed of eigenvectors, $\Lambda = \text{diag}([\lambda_0, \dots, \lambda_{N-1}]) \in R^{N \times N}$, is a diagonal matrix composed of eigenvalues.

The spectral-based GCN performs convolutional operations through Fourier transformation on the graph. The graph convolutional operation of feature vector x with a filter g can be defined as

$$g^* x = \mathcal{F}^{-1}(\mathcal{F}(x) \odot \mathcal{F}(g)) =$$

$$U(U^T x \odot U^T g) = U g_w U^T x \quad (2)$$

where $*$ indicates the graph convolutional operation; \odot is the Hadamard product; and $g_w = \text{diag}(U^T g)$ is the filter parameterized by U .

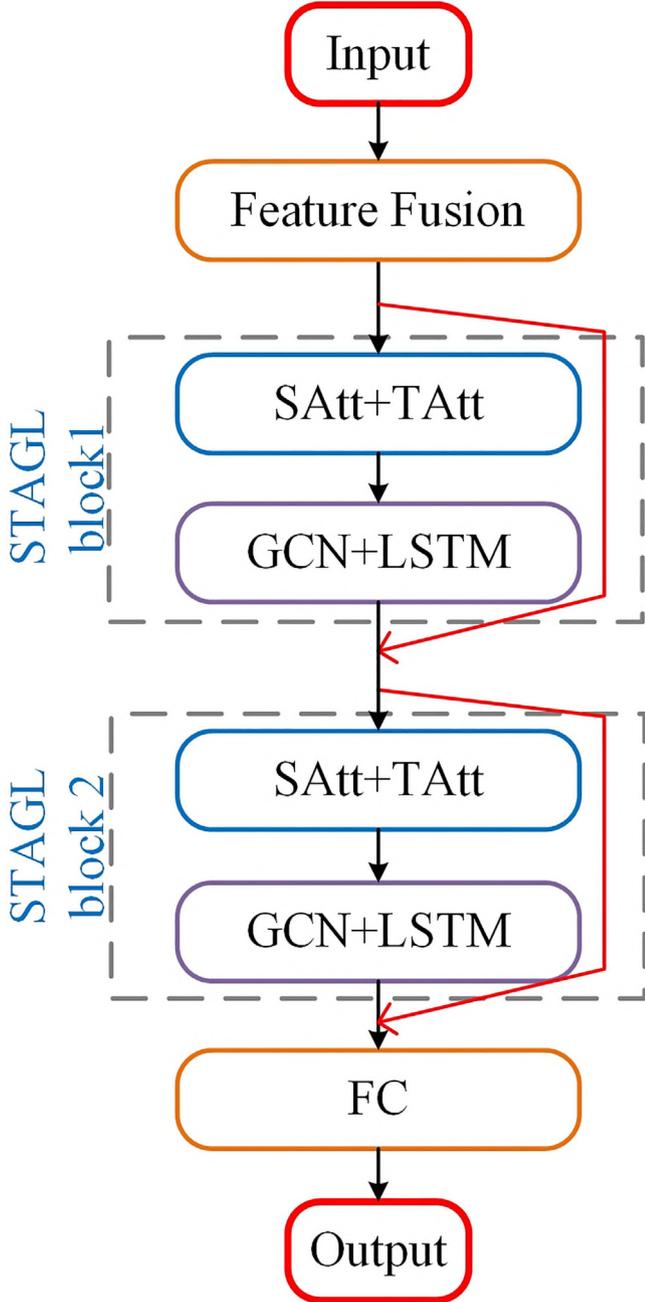


Fig. 3. The framework of MF-STAGL, SAtt: Spatial Attention; TAtt: Temporal Attention; FC: Fully-connected.

It is found that directly achieving the eigenvalue decomposition on the Laplacian matrix is not so efficient, especially when the number of nodes is large. Hence, we utilize the Chebyshev polynomials to approximate g_w . [32] With k^{th} order Chebyshev polynomials $T_k(x)$,

$$g * x \approx \sum_{k=0}^K W_k T_k \left(\frac{2L}{\lambda_{\max}} - I_n \right) x \quad (3)$$

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x),$$

$$T_0(x) = 1, T_1(x) = x \quad (4)$$

where $W_k \in R^K$ is a vector consisting of Chebyshev coefficients; λ_{\max} is the largest eigenvalue. The operation is a K^{th} order polynomial in

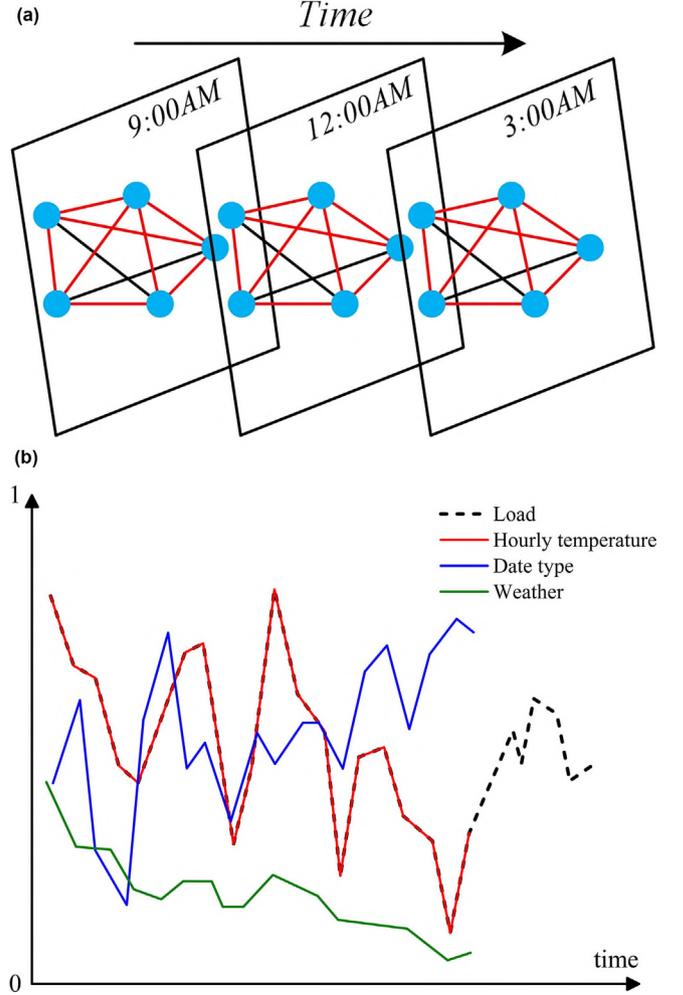


Fig. 4. The spatio-temporal structure of forecasting data, where the data at each time slice forms a graph.

the Laplacian, which means that the information of the surrounding 0 to K^{th} order neighbors centered on each node can be extracted by the convolution kernel g_w . For the connections between loads of different buildings, we set $K = 1$.

The eigenvalue range of normalized Laplacian is $[0, 2]$, then, $\lambda_{\max} = 2$. Further, to decrease the number of parameters, we can set

$$W = W'_0 = -W'_1 \quad (5)$$

The graph convolutional operation [33] can be simplified as

$$g^* x \approx W \left(I_n + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) x \quad (6)$$

W can be seen as the convolution kernel parameters. At last, we use the Rectified Linear Unit (ReLU) as the activation function of the graph convolution module, i.e., the output becomes $\text{ReLU}(g * x)$.

3.2.2. Spatio-temporal attention mechanism

The introduction of an attention mechanism can improve the learning ability of artificial neural networks, allocate the limited resources reasonably, and improve the interpretability of artificial neural networks [34].

In the GCNs, the neighborhood of nodes is commonly aggregated with equal or predefined weights. Loads of different buildings in the university architectural complex are dynamically interrelated. Therefore, their weights should be learned in the training process. In essence, the spatial attention mechanism is used to learn the spatial attention matrix $S \in R^{N \times N}$, which indicates

the impacting weights between nodes. The attention matrix S is dynamically computed according to the input of the current layer. The dynamic extraction of spatial characteristics may be done by updating the adjacency matrix A with the spatial attention matrix S .

In the temporal dimension, there are correlations between the loads in various time steps, and the correlations vary depending on the scenario. The temporal attention mechanism is included to model the importance of the inputs of a different time. The temporal attention matrix E is calculated by the fluctuating inputs. We adapt the inputs by multiplying them by the temporal attention matrix E .

3.3. Spatio-temporal feature extraction module

As a special type of a RNN, LSTM introduces a gated structure and uses memory cells to solve the problem of long-term dependence of RNN. LSTM is very suitable for dealing with problems highly related to time series. Therefore, we introduce LSTM to achieve the temporal feature of each building load. Further, we stack a GCN for spatial feature extraction and a LSTM for sequency feature learning to form the suggested GCN-LSTM module [35].

4. Experimental data processing

4.1. Data collection

All the load data are obtained from the energy-saving monitoring platform system of the considered university, which can automatically collect the load values by hourly communications with the smart meters installed in the buildings. The time range of the load data is from 0:00 on May 1th, 2019 to 23:00 on November 30th, 2019. The load data acquisition interval is one hour. Among the data of key factors, the hourly temperature is obtained from the National Oceanic and Atmospheric Administration (NOAA), and other meteorological data are obtained from the website (<https://tianqi.2345.com/>).

4.2. Data preprocessing

4.2.1. Remapping of the temperature

Given the substantial seasonal temperature variations in Beijing, a fuzzy criterion is employed to remap the hourly temperature and daily average temperature to increase the accuracy and generality of the load forecasting model.

Through the fuzzy membership function, the temperature value is remapped to the low-temperature, medium-temperature, and high-temperature. According to the temperature of Beijing, we define the low-temperature range as $-5\text{ }^{\circ}\text{C} \sim 8\text{ }^{\circ}\text{C}$, the medium-temperature as $2\text{ }^{\circ}\text{C} \sim 17\text{ }^{\circ}\text{C}$, and the high-temperature as $17\text{ }^{\circ}\text{C} \sim 35\text{ }^{\circ}\text{C}$. The temperature membership functions are defined as follows,

$$m_1 = \begin{cases} 1 & T < T_1 \\ \frac{T_2 - T}{T_2 - T_1} & T_1 \leq T \leq T_2 \\ 0 & T > T_2 \end{cases} \quad (7)$$

$$m_2 = \begin{cases} 0 & T < T_3, T > T_4 \\ \frac{T - T_3}{(T_3 + T_4)/2 - T_3} & T_3 \leq T \leq (T_3 + T_4)/2 \\ \frac{T_4 - T}{T_4 - (T_3 + T_4)/2} & (T_3 + T_4)/2 < T \leq T_4 \end{cases} \quad (8)$$

$$m_3 = \begin{cases} 1 & T > T_6 \\ \frac{T - T_5}{T_2 - T_1} & T_5 \leq T \leq T_6 \\ 0 & T < T_5 \end{cases} \quad (9)$$

where, T is the output temperature ($^{\circ}\text{C}$), $T_1 = -5\text{ }^{\circ}\text{C}$, $T_2 = 10\text{ }^{\circ}\text{C}$, $T_3 = 8\text{ }^{\circ}\text{C}$, $T_4 = 22\text{ }^{\circ}\text{C}$, $T_5 = 20\text{ }^{\circ}\text{C}$, $T_6 = 33\text{ }^{\circ}\text{C}$.

4.2.2. Outlier calibration

Outliers in the load data exist for a variety of causes and should be calibrated before being fed into the forecasting model. According to the daily and weekly periodicity and the relative stability of the university architectural complex load, the load difference at the same time of adjacent days of the same date type is not obvious without emergencies. Therefore, we compare the load data at time T with the average load at the same time on the same date type. When the difference is greater than the preset threshold, the data can be regarded as abnormal and calibrated by using the following formula:

$$Y'(t) = \begin{cases} Y_{\text{avg}} + \delta(t) & Y(t) - Y_{\text{avg}} > \delta(t) \\ Y_{\text{avg}} - \delta(t) & Y(t) - Y_{\text{avg}} < \delta(t) \end{cases} \quad (10)$$

where, $Y(t)$ is the original value (kW·h), $Y'(t)$ is the calibrated value (kW·h), Y_{avg} is the average load at the same time with the same date type (kW·h), and $\delta(t)$ is the preset threshold.

4.2.3. Data normalization

The value ranges of different key factors and load are varied widely. Hence, zero-mean normalization is performed to decrease the gaps. During the experiments, the dataset is randomly split into the training set, validation set, and test set with a ratio of 6:2:2.

5. Experimental results and analysis

5.1. Experiment setup

The hardware environment of the experiments is: CPU: Intel (R) Core (TM) i7-8565U, Memory: 16 GB. The software environment is: Windows 10 64 operating system, Visual Studio coding platform, Pytorch 1.8.1 and python 3.6.

The detailed structures and the input/output shapes of the model are listed in Table 5. There are eight features input into the model, which contains hourly load, maximum load of the day before, minimum load of the day before, the fuzzed low temperature, the fuzzed medium temperature, the fuzzed high temperature, weather type, and the date type. The input time length and output time length are both twelve, which means that loads of twelve hours will be predicted by inputting the features of the previous twelve hours with the proposed model. The input dimension is $32 \times 5 \times 8 \times 12$, where 32 is the Batchsize, 5 means the number of buildings, 8 represents the number of the input features, and 12 indicates the forecast period. This input allows multiple buildings' data to be input into the MF-STAGL model at the same time.

Table 5
Model design.

Group	Name	Input shape	Output shape
Pre-processing	Feature Fusion	$32 \times 5 \times 8 \times 12$	$32 \times 5 \times 4 \times 12$
STAGL block1	TAtt	$32 \times 5 \times 4 \times 12$	$32 \times 5 \times 4 \times 12$
	SAtt	$32 \times 5 \times 4 \times 12$	$32 \times 5 \times 4 \times 12$
	GCN	$32 \times 5 \times 4 \times 12$	$32 \times 5 \times 32 \times 12$
	LSTM	$32 \times 5 \times 32 \times 12$	$32 \times 5 \times 16 \times 12$
STAGL block2	TAtt	$32 \times 5 \times 16 \times 12$	$32 \times 5 \times 16 \times 12$
	SAtt	$32 \times 5 \times 16 \times 12$	$32 \times 5 \times 16 \times 12$
	GCN	$32 \times 5 \times 16 \times 12$	$32 \times 5 \times 32 \times 12$
	LSTM	$32 \times 5 \times 32 \times 12$	$32 \times 5 \times 16 \times 12$
Output	FC	$32 \times 5 \times 16 \times 12$	$32 \times 5 \times 12$

Table 6
Training hyperparameters.

Parameters	Value
Learning rate	0.001
Optimizer	Adam
Epoch	100
Batchsize	32

Mean Squared Error Loss (MSE) is selected as the loss function, which allows faster convergence compared to other loss functions. The training hyperparameters are shown in Table 6.

5.2. Evaluation metrics

Four evaluation metrics are calculated to estimate the predicting performance: Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are used to measure the prediction error, the coefficient of determination R^2 and Accuracy are introduced to evaluate the goodness of the fit. The functions of the four metrics are defined as:

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (\bar{Y} - y_i)^2} \quad (13)$$

$$Accuracy = 1 - \frac{\|Y - \hat{Y}\|_F}{\|Y\|_F} \quad (14)$$

where N is the total number of forecasting results; Y are the actual value of the load (kW·h), \hat{Y} is the forecasting value of the load (kW·h); y_i is the actual load of point i (kW·h); \hat{y}_i is the forecast load of y_i (kW·h), and \bar{Y} is the mean value of the actual value of the load (kW·h).

5.3. Performances of the proposed model

5.3.1. Comparison of different buildings

With the trained MF-STAGL model, loads of all five buildings can be predicted at the same time. To evaluate the MF-STAGL model more comprehensively, the MAPE, RMSE, R^2 , and Accuracy were calculated for all five buildings, respectively. The results are shown in Table 7.

It can be seen from Table 7 that the MAPE of most buildings are generally less than 5 % except for the building Second Canteen. Accordingly, the RMSE of Second Canteen is the largest. Fig. 5

Table 7
MAPE, RMSE, R^2 and Accuracy of prediction results.

	MAPE (%)	RMSE (kW·h)	R^2	Accuracy
Second Canteen	6.86	14.96	0.96	0.89
Second Teaching Building	4.98	8.59	0.96	0.90
Experimental Building	5.03	2.49	0.88	0.90
Dormitory Building 7	4.22	8.97	0.87	0.85
Main Building	3.54	8.82	0.96	0.95

shows the 48-hour forecast results of five buildings from October 25, 2019 to October 27, 2019.

Further, we calculated the load fluctuations of five buildings with Eq. (15), and Table 8 shows the results.

$$flu = \frac{y_{max} - y_{min}}{y_{max}} \quad (15)$$

where flu is the fluctuation, y_{max} and y_{min} are the maximum and minimum of loads in certain period (kW·h), respectively.

Among loads of 48 h, the Second Canteen has the largest fluctuation, which is 2.15 %. And the maximum load is 303.3 kW·h, which is the highest of all the buildings. As a result, although R^2 and the Accuracy of the Second Canteen are relatively high, the MAPE and RMSE of the Second Canteen are still the largest. Conversely, due to the load fluctuation of the Dormitory Building 7 is not so high, the MAPE and RMSE are not the highest, despite the smallest R^2 and Accuracy. Furthermore, the Main Building has the smallest fluctuation, which is 0.60 %, then the smallest MAPE is achieved with the highest R^2 and Accuracy. These findings demonstrate that the change in MAPE is related to both the goodness of fit and the amount of load fluctuation. Based on this conclusion, as well as the fact that the load of the Experimental Building is much smaller than the load of the Dormitory Building 7 (the former is nearly one-tenth the latter), the load fluctuation of the former is also much smaller, as a result, even though their R^2 are nearly the same, leading to the lower accuracy of the Dormitory Building 7. To sum up, the highest MAPE is only 6.86 %, which indicates that the model still has good forecasting ability for the load with large variations.

The Main Building is an office building. The routine of getting to and from work is identical to that of getting to and from class to some extent. Compared with Fig. 5(b) and (c), it is found that the change law of the load of the Second Teaching Building is a little more complex than that of the Main Building. Hence, all the evaluation metrics of them are almost equal, except for the MAPE of the Second Teaching Building which is considerably higher.

Unlike the other four buildings, it is obvious that the load of the Experimental Building has no significant day periodicity (shown in Fig. 5(e)), which can be attributed to the uncertainty of the operations of the large experimental equipment. The forecasting load values are not fitted with high accuracy. Moreover, although the fitting performance is not so well ($R^2 = 0.88$, Accuracy = 0.90), MAPE and RMSE are not very poor, which is mainly due to the low load.

5.3.2. Correlation analysis of loads of buildings

We further calculated the Pearson correlation coefficients using the forecasting loads of the buildings to study the relations among the buildings. The results are shown in the form of a heat map in Fig. 6. Leaving aside the diagonal, the second row is the darkest of all rows, or the second column is the darkest of all columns. It indicates that the correlations between the Second Teaching Building and the other buildings are stronger than those between the other two buildings. The Second Teaching Building can be seen as the load flow center, which is closely linked with the office, laboratory, and living areas. There is an obvious flow of students and

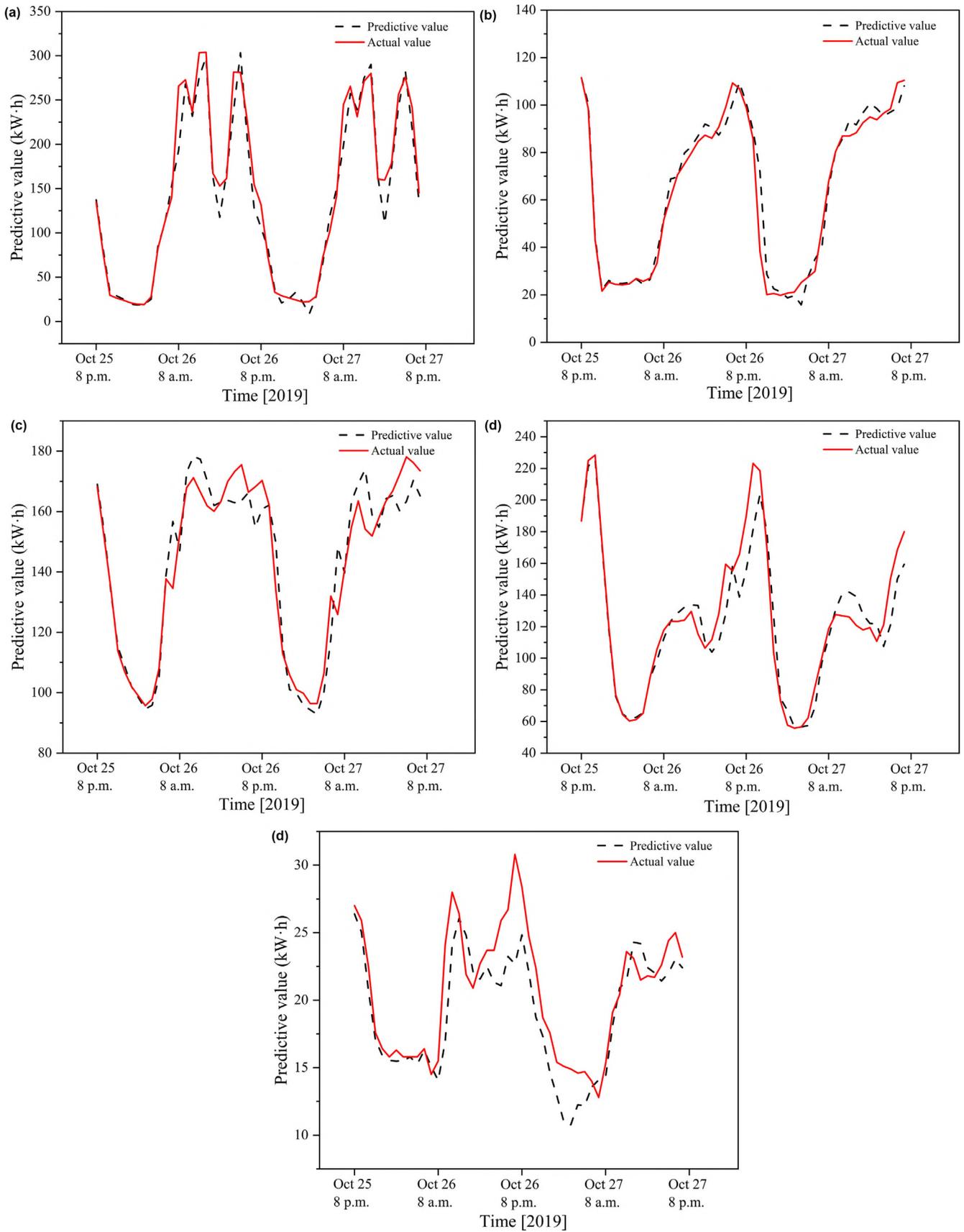


Fig. 5. 48-hour forecast results of five buildings.

Table 8
The fluctuations of five buildings.

Building Name	Fluctuation
Second Canteen	2.15 %
Second Teaching Building	1.46 %
Experimental Building	0.82 %
Dormitory Building 7	1.38 %
Main Building	0.60 %

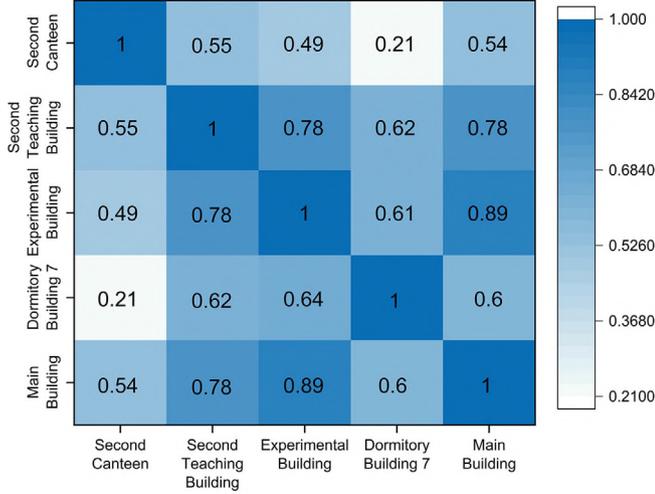


Fig. 6. The correlation coefficients among the buildings.

teachers between the Second Teaching Building and other buildings, which brings the load flow.

In addition, the light-colored areas in the figure are mostly found between the Second Canteen and other buildings. The primary reason for this is that the load of Second Canteen Building has little to do with the people flow, resulting in distinct disparities between the Second Canteen Building and the other buildings. On the one hand, the fluctuation in the load of the Second Canteen is larger than the other buildings; on the other hand, the peak and valley times of the Second Canteen are distinctly different from those of the other buildings. As a result, the relationship between the load of Second Canteen and that of the other buildings is weak. In summary, this analysis reveals that the addition of the graph convolutional neural network increased the interpretability of the forecasting model for forecasting the architectural complex loads from the viewpoints of load flow and load change law.

5.4. Performance comparison with other models

Five commonly used forecasting models, namely History Average (HA), SVR, ARIMA, LSTM, and a graph neural network model without spatio-temporal attention mechanism (MF-STGL) are compared with the MF-STAGL model. Fig. 7 shows the forecasting results of the Main Building from October 24, 2019, 3:00 pm to October 27, 2019, 3:00 am using the above six models.

As shown in Fig. 7, except for the HA model, all the models can track the fluctuation of the load rather effectively. Table 9 shows the evaluation results of six models for the Main Building. The MAPE of HA is as high as 43.38 %, the R^2 is only 0.58, and the Accuracy is just 0.46, which is insufficient for forecasting rapidly varying loads. The performance of ARIMA and SVR is similar, and the R^2 of SVR is higher, which makes the MAPE of SVR smaller. Compared with a classical statistical model, HA and the traditional machine learning models ARIMA and SVR, the LSTM, the MF-STGL, and the MF-STAGL models achieve more accurate forecasting results by

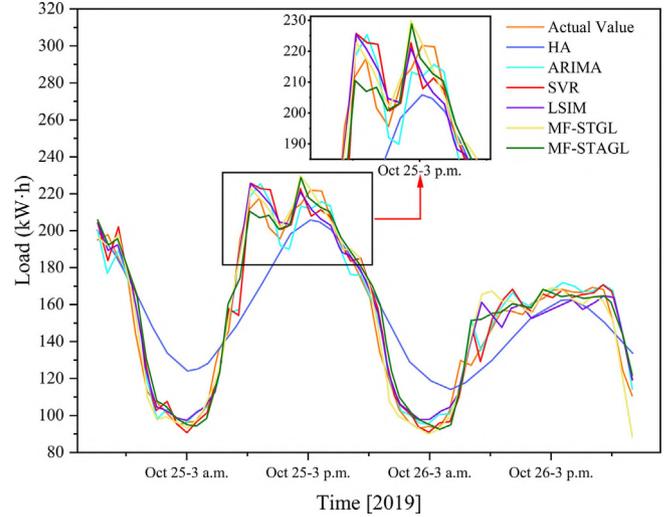


Fig. 7. Forecasting results of six different models.

Table 9

The evaluation results of the Main Building with six models. The best performances are marked in Bold.

Model	MAPE (%)	RMSE (kW·h)	R^2	Accuracy
HA	43.38	52.34	0.58	0.46
ARIMA	7.12	10.02	0.87	0.85
SVR	6.87	9.05	0.89	0.88
LSTM	3.84	8.96	0.95	0.93
MF-STGL	3.77	8.89	0.96	0.94
MF-STAGL	3.54	8.82	0.96	0.95

introducing neural networks, and the MAPE declines obviously. Compared with SVR, the MAPE and RMSE of LSTM decrease by 3.03 % and 1.00 %, respectively. MF-STGL combines the temporal and spatial features, all the evaluation metrics are better than using LSTM, which only considers the temporal feature. The performance of MF-STAGL is the best. With the attention mechanism, the MAPE and RMSE reduce by 0.23 % and 0.79 % compared with MF-STGL respectively, which means the attention mechanism is a benefit for tracking detail changes of the loads. The enlarged portion of Fig. 7 clearly demonstrates this.

5.5. Forecasting stability over longer forecast period

As we all know, the lengthening of the forecast period makes load forecasting more challenging. Hence, we choose forecast periods of 1, 6, 12, and 24 h respectively, then predict the load of the Main Building using four load forecasting models, SVR, LSTM, MF-STGL, and MF-STAGL during the period from October 25, 2019, 3:00 pm to October 26, 2019, 3:00 pm. The calculated RMSEs are shown in Fig. 8. The variation of the RMSEs of the four models is almost linear with the increase of the forecast period, the rate of change is the largest for SVR and the smallest for MF-STAGL. With LSTM, the variation rates of LSTM, MF-STGL, and MF-STAGL are relatively small, which benefit from the memory ability of LSTM. Comparing LSTM and MF-STGL, it is shown that the spatial information of the surrounding buildings can be used to reduce the impact of the forecast period. For MF-STAGL, we believe that the temporal attention mechanism is the main reason that makes the variation rate of MF-STAGL minimal. Overall, the combined use of spatio-temporal features and temporal attention mechanisms

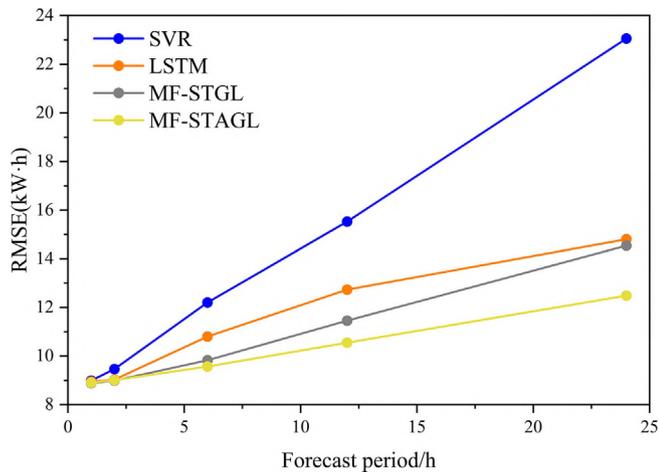


Fig. 8. RMSEs vary with the forecast period.

can maintain the stability of the forecasting to a certain extent with a longer forecast period.

6. Conclusion

To address the issue of traditional load forecasting methods focusing on a single building, we proposed the MF-STAGL model, a short-term load forecasting model that can extract the inherent spatio-temporal characteristics of nonlinear and complicated load data and achieve accurate load forecasting of all buildings in an architectural complex at the same time. In forecasting the actual load data from a public university in Beijing, the MF-STAGL model beat conventional statistical approaches and typical machine learning methods. For the forecasting performances of all buildings, the largest MAPE is only 6.86%, and the smallest R^2 and Accuracy can reach 0.87 and 0.85, respectively. Furthermore, as the forecast period expanded, the use of the attention mechanism helped to maintain considerable forecasting stability. With the proposed method, the rise in RMSE with increasing forecast period is only one-fifth of that of the SVR method and half of that of the LSTM model. Overall, in comparison to traditional load forecasting approaches, which need many predictions using different models, the MF-STAGL model can obtain robust and accurate load forecasting results for all buildings in an architectural complex at the same time, effectively saving model training time. In future efforts, we will try to introduce the transformer like architectures to improve the performance and extend them to practical applications.

Data availability

Data will be made available on request.

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

We would like to express our gratitude to the Beijing Municipal Natural Science Foundation (No. 6214040), the Fundamental Research Funds for the Central Universities (No. 2021ZY70) and China Scholarship Council (202106515010).

References

- [1] Professional Committee of Building Energy and Emissions. Building Energy and Emissions Database, CBEED, www.cbeed.cn. 2022.09.22.
- [2] H. Dagdougui, F. Bagheri, H. Le, et al., Neural network model for short-term and very-short-term load forecasting in district buildings, *Energy Build.* 203 (2019).
- [3] L. Zhang, J. Wen, Y. Li, et al., A review of machine learning in building load prediction, *Appl. Energy* 285 (2021).
- [4] E. Chodakowska, J. Nazarko, Ł. Nazarko, Arima models in electrical load forecasting and their robustness to noise, *Energies* 14 (23) (2021) 7952.
- [5] J. Wu, Y.G. Wang, Y.C. Tian, et al., Support vector regression with asymmetric loss for optimal electric load forecasting, *Energy* 223 (2021).
- [6] H. Dong, Y. Gao, Y. Fang, M. Liu, Y. Kong, The short-term load forecasting for special days based on bagged regression trees in Qingdao, China, *Comput Intell Neurosci.* 2021 (2021) 3693294.
- [7] A. Bellahsen, H. Dagdougui, Aggregated short-term load forecasting for heterogeneous buildings using machine learning with peak estimation, *Energy Build.* 237 (2021).
- [8] Y.T. Chae, R. Horehsh, Y. Hwang, Y.M. Lee, Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings, *Energy Build.* 111 (2016) 184–194.
- [9] S.-V. Oprea, A. Bara, Machine learning algorithms for short-term load forecast in residential buildings using smart meters, sensors and big data solutions, *IEEE Access* 7 (2019) 177874–177889.
- [10] L. Xu, S. Wang, R. Tang, Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load, *Appl. Energy* 237 (2019) 180–195.
- [11] Fujimoto Y, Fujita M, Hayashi Y. Deep reservoir architecture for short-term residential load forecasting: an online learning scheme for edge computing. *Appl. Energy.* 2021;298.
- [12] Y. Liu, W. Wang, N. Ghadimi, Electricity load forecasting by an improved forecast engine for building level consumers, *Energy* 139 (2017) 18–30.
- [13] J. Ma, W. Gong, Z. Zhang, A multivariate short-term load forecasting model for regional integrated energy systems based on Copula theory combined with KPCC-GRNN, *Adv. Technol. Electr. Eng. Energy* 39 (2020) 24–31.
- [14] Y. Liang, D. Niu, W.C. Hong, Short term load forecasting based on feature extraction and improved general regression neural network model, *Energy* 166 (2019) 653–663.
- [15] Xu J, Chi F, Ge L, Li J, Zhang L, Xian Y. SARIMA-GRNN-SVM based short-term commercial electricity load combination forecasting method. *Proceedings of the CSU-EPSA.* 2020;32:85-91.
- [16] G. Sun, C. Jiang, X. Wang, X. Yang, Short-term building load forecast based on a data-mining feature selection and LSTM-RNN method, *IEEJ Trans. Electr. Electron. Eng.* 15 (2020) 1002–1010.
- [17] T. Yang, L. Santamouris, Building energy consumption raw data forecasting using data cleaning and deep recurrent neural networks, *Buildings* (2019) 9.
- [18] L. Wen, K. Zhou, S. Yang, Load demand forecasting of residential buildings using a deep learning model, *Electr. Power Syst. Res.* 179 (2020).
- [19] W. Guo, L. Che, M. Shahidehpour, X. Wan, Machine-Learning based methods in short-term load forecasting, *Electricity J.* 34 (2021).
- [20] C. Tian, J. Ma, C. Zhang, P. Zhan, A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network, *Energies.* (2018) 11.
- [21] Y. Jin, M.A. Acquah, M. Seo, et al., Short-term electric load prediction using transfer learning with interval estimate adjustment, *Energy Build.* 258 (2022).
- [22] R. Li, F. Sun, X. Ding, Ultra short-term load forecasting method for user-level integrated energy system considering spatio-temporal coupling of multi-energy, *Power System Technol.* 44 (2020) 4121–4134.
- [23] S. He, C. Li, X. Liu, X. Chen, M. Shahidehpour, T. Chen, et al., A per-unit curve rotated decoupling method for CNN-TCN based day-ahead load forecasting, *IET Gener. Transm. Distrib.* 15 (2021) 2773–2786.
- [24] K. Li, J. Zhang, X. Chen, et al., Building's hourly electrical load prediction based on data clustering and ensemble learning strategy, *Energy Build.* 261 (2022).
- [25] M.M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, P. Vandergheynst, Geometric deep learning: going beyond euclidean data, *IEEE Signal Process Mag.* 34 (2017) 18–42.
- [26] B. Kan, G. Liu, K. Mahdi, J. Zhou, L. Wei, J. Tan, et al., Distributed photovoltaic generation prediction based on graph machine learning, *Distrib. Util.* 36 (2019) 20–27.
- [27] M. Yu, Z. Zhang, X. Li, J. Yu, J. Gao, Z. Liu, et al., Superposition Graph Neural Network for offshore wind power prediction, *Fut. Gen. Comput. Syst.* 113 (2020) 145–157.
- [28] M. Khodayar, J. Wang, Spatio-temporal graph deep neural network for short-term wind speed forecasting, *IEEE Trans. Sustainable Energy* 10 (2019) 670–681.
- [29] X. Geng, L. Xu, X. He, J. Yu, Graph optimization neural network with spatio-temporal correlation learning for multi-node offshore wind speed forecasting, *Renewable Energy* 180 (2021) 1014–1025.
- [30] Battaglia PW, Hamrick JB, Bapst V, Sanchez-Gonzalez A, Zambaldi V, Malininowski M, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:180601261.* 2018.
- [31] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016. p. 770-778.

- [32] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, *Adv. Neural Inf. Process. Syst.* 29 (2016) 3844–3852.
- [33] Berg RVD, Kipf TN, Welling M. Graph convolutional matrix completion. arXiv preprint arXiv:170602263. 2017.
- [34] S. Chaudhari, V. Mithal, G. Polatkan, R. Ramanath, An attentive survey of attention models, *ACM Trans. Intell. Syst. Technol. (TIST)* 12 (2021) 1–32.
- [35] Y. Seo, M. Defferrard, P. Vandergheynst, X. Bresson, Structured sequence modeling with graph convolutional recurrent networks, in: *International Conference on Neural Information Processing*, Springer, 2018, pp. 362–373.