

# A Short-Term Forecast Approach of Public Buildings' Power Demands upon Multi-source Data

Shubing Shan<sup>1,2(✉)</sup> and Buyang Cao<sup>1,2</sup>

<sup>1</sup> School of Software Engineering, Tongji University, 200092 Shanghai, China  
1410786@tongji.edu.cn

<sup>2</sup> China Intelligent Urbanization Co-Creation Center,  
Tongji University, 200092 Shanghai, China

**Abstract.** Due to the significant increase of the global electricity demand and the rising number of urban population, the electric consumption in a city has attracted more attentions. Given the fact that public buildings occupy a large proportion of the electric consumption, the accurate prediction of electric consumptions for them is crucial to the rational electricity allocation and supply. This paper studies the possibility of utilizing urban multi-source data such as POI, pedestrian volume etc. to predict buildings' electric consumptions. Among the multiple datasets, the key influencing factors are extracted to forecast the buildings' electric power demands by the given probabilistic graphical algorithm named EMG. Our methodology is applied to display the relationships between the factors and forecast the daily electric power demands of nine public buildings including hotels, shopping malls, and office buildings in city of Hangzhou, China over the period of a month. The computational experiments are conducted and the result favors our approach.

**Keywords:** Electricity demand · Short-term forecast · Multi-source data · Grey relational analysis · Probabilistic graph

## 1 Introduction

With the population growth and economic development, the global electric consumption is increasing yearly. In the past decades, the proportion of the world's urban population has been rising and the per capita electric consumption has gradually increased. The electric power demand in a city is undergoing drastic changing. According to the latest statistics released by the World Bank in 2016 [1], the global per capita electric consumption rose to 3104.7 kWh (kilowatt hour) in 2013 up from 2027.4 kWh in 2006 and the percentage of the world's urban population increased by 3.9% from 2006 to 2014. As the result, the urban electric consumption will rise sharply worldwide. Meanwhile, the urban public buildings have always been one of major electric consumption groups around the world [2, 3]. The topic of building energy demand-side is the focus of researches today because an accurate prediction of demand is very important for every country to work out the reasonable plan of energy production and reduce carbon emissions [4–6]. It also plays a vital role in rationally

allocating a city's electricity, avoiding peak-hour power shortage, saving public funds, reducing economic risks, and reducing environmental pollution caused by excessive electricity generation [7, 8].

In the past researchers paid close attentions mainly to the influences of four factors including meteorological factors, building attributes, time series, and occupancy rates. Among them meteorological data, building attribute data and time series data can be more easily obtained, but the occupancy is difficult to obtain accurately. There are two possible ways to obtain building occupancy data including manual counting and automatic sensing. The former method is not only time-consuming but also cannot be conducted in real time. Thus, the occupancy data obtained is usually a macroscopic statistical data by such as the annual average occupancy. The automatic sensing way is to install a sensing device and a data collecting system in the building records real-time building occupancy. But this method requires expensive hardware and software investment, which makes it difficult to be applied extensively. It inspires us to invent a way to capture the occupancy upon other data. From the practice, we know the occupancy of a building is influenced by regional functions and regional vitality [9–11]. Therefore, we try to acquire the occupancy rate of a building through the urban multi-source data such as POI and pedestrian volume data in the surrounding area of the building to support the forecast of building electric consumption. Furthermore, few scholars have studied the relationships between factors such as weather and time series and their impacts on the occupancy of a building. We use the probability graph model to represent the relationships between various influence factors and electric consumptions and an approximate inference algorithm to predict public buildings' electric consumptions.

This paper has the following contributions:

- (1) It studies the relationships between urban multi-source data with electric consumptions to predict building electric consumptions.
- (2) It applies the probability graph model to study and express the relationships between various factors, and an approximate inference algorithm known as EMG is proposed to predict building electric consumptions.
- (3) We evaluate the proposed approaches using the real data from nine public buildings in city of Hangzhou, China. In our computational experiments, the MAPE is used for the quality criterion. A comparative analysis is performed by using the regression analysis. The result indicates that our approach has a better accuracy.

This paper is organized as follows. Related work is presented in Sect. 2. Section 3 gives the methodology including grey correlation analysis and probability graph algorithm. Section 4 is the case study and results. Several tests and statistical analysis are provided in Sect. 5. The paper is concluded with some remarks.

## 2 Related Work

In the past researchers have been continuously exploring four approaches to forecast buildings' energy consumptions including meteorological, architectural attributes, occupancy, and time series predictions. Some studies explore the impact of meteorological

factors on building energy consumption. Zheng et al. study the effects of hour of day and outside air temperature on hot water energy consumption by a data driven method [12]. Yang et al. investigate some variables such as time of day and outdoor air dry-bulb temperature, and apply artificial neural network to make a short-term electric consumption prediction for commercial buildings [13]. Nelson et al. conduct the research on the influence of meteorological factor on residential buildings' energy and use a quadratic regression analysis approach to predict the demand of buildings' energy [14]. Ambera et al. investigate the influence of five important factors (temperature, solar radiation, relative humidity, wind speed, and weekday index) on administration buildings' energy. They use a multiple regression model and a genetic programming model to forecast daily electricity consumption [15]. James et al. look into the impact of climate change on peak and annual building energy consumption [16].

Many scholars pay closer attention to the effect of architectural attributes on the electric consumption. Lu use a physical–statistical approach which includes physical model and the statistical time series model to predict the energy consumption of buildings. The physical model simulates the basic energy consumption of different buildings and the statistical time series model reflects the heterogeneity of various buildings [17]. Akin makes a short-term prediction of electric demand through the detailed data and information of the house [18]. Cara et al. develop the auto-regressive models with building specific inputs for forecasting power demands [19]. Kristopher et al. carry out a study that utilizes statistical learning methods to predict the future monthly energy consumptions for single-family detached homes using building attributes and monthly climate data [20].

Researchers also try to utilize occupancy and time series factors to predict the building energy consumptions. Ferlito et al. use the building properties, occupancy rate and weather to predict buildings' energy consumption by Artificial Neural Network method [21]. Sandels et al. explore the influence of weather, occupancy, and temporal factors on electricity consumptions of a Swedish office building [22]. Kim et al. study the influence of building occupancy and construction area allocation on building electric consumption, and uses the linear equation method to predict the electric consumption of buildings [23]. López-Rodríguez et al. conclude that building electricity demand is highly correlated with occupancy time in buildings, and build an occupancy statistical model for creating active occupancy with the aim to predict electricity consumptions [24]. Kavousian studies the structural and behavioral determinants of residential electricity consumption. This study shows that electric consumption is not significantly related to income level, home ownership, or building age [25].

## 3 Methodology

### 3.1 Grey Correlation Analysis

The grey relational analysis has been widely studied and applied since its birth [26, 27]. It determines the degree of association upon the similarity of the curves represented by the two series. The grey correlation degree is used to represent the degree of association.

The grey correlation degree is computed as follows:

**Step 1:** Set a data column (*reference column*) of the historical electric consumptions as shown in (1). Let  $m$  be the number of records for one of the underlying six buildings' electric consumptions during 92 days.

$$X'_0 = (x_0(1), x_0(2), \dots, x_0(m))^T, \quad m \leq 92 \tag{1}$$

**Step 2:** The reference column together with the comparing columns form a matrix  $A$ . We apply a normalization process to all data in the matrix for the analysis accuracy. The normalized data matrix  $B$  is shown (2). The  $n$  is number of factors. The first column is normalized reference column and others are normalized comparing columns. The normalization process utilizes the Initiative Value method (3) where  $X'_i(1)$  is the value of first row of matrix  $A$ .

$$(X_0, X_1, \dots, X_n) = \begin{pmatrix} x_0(1) & \dots & x_n(1) \\ \vdots & \ddots & \vdots \\ x_0(m) & \dots & x_n(m) \end{pmatrix} \tag{2}$$

$$x_i(k) = \frac{x'_i(k)}{x'_i(1)} \quad i = 0, 1, \dots, n; k = 1, 2, \dots, m \tag{3}$$

**Step 3:** We compute one by one the absolute difference between the elements in normalized comparing and reference ones. Correlation coefficients between normalized comparing and normalized reference columns are calculated. In formula (4),  $\rho$  is the distinguishing coefficient that takes values in the range (0, 1). The smaller the  $\rho$  value is, the greater the difference between correlation coefficients is. It may be adjusted based on the practical needs of the system.

$$\xi_i(k) = \frac{(\min_k |x_0(k) - x_i(k)| + \rho \max_k |x_0(k) - x_i(k)|)}{(|x_0(k) - x_i(k)| + \max_k |x_0(k) - x_i(k)|)} \quad k = 1, \dots, m \tag{4}$$

**Step 4:** Using the outcome obtained in step 3, the mean value of correlation coefficients for each *comparing column* can be calculated upon (5) respectively. The purpose of this is to acquire the correlation between every pair of comparing column and reference column and yield the grey correlation degree. The larger the value, the greater the influence of the comparing column.

$$r_{0i} = \frac{1}{m \sum_{k=1}^m \xi_i(k)} \tag{5}$$

### 3.2 Probability Graph Model

There are three steps to construct a probabilistic graphical model: structure learning, parameter learning, and inference.

- Structure learning

The obtained data may be incomplete, so we use the SEM (Structural Expectation-Maximization) algorithm [28] to learn structure. It is adopted in the structure learning based on an incomplete data set. Figure 1 shows the Pseudo code of the algorithm.

---

```

SEM (
  G0, // Initial network structure
  Θ0,0, // Initial parameter value
  D, // Missing dataset
  R // Optimization of the number of parameters between the two structural optimization
)
1 For t=0,..., until convergence
2 For r=0 to R-1
3   Θt,r+1 ← Expectation-Maximization(Gt, Θt,r, D)
4   Θt = Θt,R
5 L ← All of the Gt to do a plus side, edge reduction or edge of the candidate model structure
6 Gt+1 ← arg max Score(G|D, Θt)
7 Θt+1,0 ← Estimate Gt+1 parameter values
8 If (Score(Gt+1, Θt+1, 0|D) ≤ Score(Gt, Θt|D))
9   Return (Gt, Θt)
10 Return(Gt+1, Θt+1)

```

---

**Fig. 1.** The pseudo code of SEM algorithm

- Parameter learning and inference

In the Probabilistic Graphical model, EM (Expectation-Maximization) algorithm is an approximate learning and inference algorithm [29], which can resolve the incomplete graph problem. The Gibbs algorithm [30] is data sampling algorithm upon the Monte Carlo method to reduce the amount of data and speed up the calculation. Here we propose a hybrid algorithm, EMG algorithm, which utilizes the advantages of both algorithms.

The pseudo code of the EMG algorithm is shown in Fig. 2. The algorithm includes three parts, the first is to generate a sample data set for each entity variable by the Gibbs algorithm. The distribution of the sample data set is similar to the real data set. Second, it obtains the current parameters of each entity from the sample dataset, and uses the current parameters and graph structure to compute the expected value of each entity variable. Finally, it recalculates the parameters of each entity by the expected value of it. It iterates until the estimated parameter reaches the local optimum or reaches the specified number of iterations.

- The sample dataset

The approximate inference algorithm based on Gibbs sampling is one of the simplest and the most popular methods of data sampling. It uses each node as a variable to conduct a random sampling, and then assigns an initial value to each variable to get an initial state. It computes each node's conditional probability to achieve a new value and

---

```

// Dataset generation algorithm based on Gibbs-sample
Procedure Gibbs-Sample (
  X // The set of variables that need to be sampled
   $\Phi$  // Define Factor Sets of  $P_{\Phi}$ 
   $P^{(0)}(X)$  // Initial State Distribution
  T // Sampling Times
  D, // Sample dataset
  G, // Bayesian network structure
   $\theta$ , // Bayesian network parameters
   $\theta^0$ , // Bayesian network initial parameters
)
//Producing a sample  $x^{(0)}$  from the distribution  $P^{(0)}(X)$ 
1 For  $t=1, \dots, T$ 
2    $X^{(t)} \leftarrow x^{(t-1)}$ 
3   For each  $X_i \in X$ 
4     Sampled  $x_i^{(t)}$  From  $P_{\Phi}(X_i|x_i)$ 
5     Replace  $X_i$  in  $x^{(t)}$ 
6 Return  $x^{(0)}, \dots, x^{(T)}$ 
7 Return  $D = \{x^{(0)}, \dots, x^{(T)}\}$ 
//EM parameter learning algorithm
// E-step
8 For each  $i=1, \dots, n$  // Initialize the data structure
9 For each  $x_i, \mu_i \in \text{Val}(X_i, Pa_{X_i}^G)$ 
10  $\bar{M}xi|\mu_i \leftarrow 0$ 
11 For each  $m=1, \dots, M$ 
12 // The data model D [m] is used to carry out the reasoning operation on the network
13 // structure and //the current parameters
14 Run inference on (G,  $\theta$ ) using evidence D[m]
15 For each  $i = 1, \dots, n$ 
16 For each  $x_i, \mu_i \in \text{Val}(X_i, Pa_{X_i}^G)$ 
17  $\bar{M}xi|\mu_i \leftarrow \bar{M}xi|\mu_i + p(x_i, \mu_i|D[m])$ 
18 Return  $\{\bar{M}xi|\mu_i: \forall i = 1, \dots, n, \forall x_i, \mu_i \in \text{Val}(X_i, Pa_{X_i}^G)\}$ 
// M-step
19 For each  $t=0, 1, \dots$ , until convergence
20  $\{\bar{M}^t_{xi|\mu_i}\} \leftarrow \text{Compute-ESS}(G, \theta, D)$ 
21 For each  $i=1, \dots, n$ 
22 For each  $x_i, \mu_i \in \text{Val}(X_i, Pa_{X_i}^G)$ 
23  $\theta^{t+1}_{xi|\mu_i} \leftarrow \bar{M}^t_{xi|\mu_i} / \bar{M}^t_{\mu_i}$ 
24 Return  $\theta^t$ 

```

---

**Fig. 2.** The pseudo code of EMG algorithm

state based on the Markov Cover. The above steps repeat until the number of samples reaches a given threshold and the sample data set is obtained.

- Parameter learning

The weights of each entity in the graph are obtained by the parameter learning of the EMG algorithm. Its parameter learning is similar to the EM parameter learning algorithm.

**E-Step.** It uses the graph structure and the current parameters to calculate the expected value of missing variables.

**M-Step.** In the M-step by scanning the inferred results from the E-step, the algorithm recalculates the new maximum parameter distribution and replaces the old parameters with new ones. It repeats until the parameters converge, and we have learned the unknown parameters.

- Inference

In the E-step (Line 12–16) of EMG algorithm, we call exact inference method, i.e., use the simple Bayesian rule, to compute the values of the hidden entity nodes, for each instance of the observed data. This is actually an inference process.

## 4 Case Study and Result

This paper uses the nine public buildings in Hangzhou, China for the case study. Among them there are shopping malls, hotels and office buildings to illustrate the (predicting) methodology. The paper explores the correlations between buildings' electric consumptions and influencing factors including: architectural property, weather, air quality, population and POI data. All the different sorts of data obtained will be further processed.

### 4.1 Data Preparation

- The architectural property data

In spite of different functions and structures, the buildings possess some common attributes or properties. The property data collected includes building age, number of stories (including ground and underground), and total area ( $m^2$ ) as well as window/wall ratio.

- Historical Electric Consumption Data

The historical data of daily electric consumption is acquired for these public buildings from January 1, 2015 to January 31, 2016. The daily electric consumptions spanned from 0:00 a.m. to 23:59:59 p.m. The electricity unit is KWH (kilowatt hour). In order to verify the prediction, this paper divides the data set into two subsets: data for May, June and July used as the training sets while data for August used as the testing one to validate the model.

- Weather and Air Quality Data

The weather data collected contains the daily average temperature and humidity from January 1, 2015 to November 31, in City of Hangzhou, China. The temperature unit is degree centigrade, and the humidity unit is percentage.

- POI Data

POI data contains a number of specific functional facilities such as restaurants, bus stops, etc. The dataset has the name, address, coordinate and other attributes of the functional properties. In our paper six functional facilities (POI) within 200 meters around concerned buildings are included. The six types of POI are office buildings, shopping malls, restaurants, hotels, metro stations and bus stations. The number of bus stations is calculated according to distinguish bus routes. For example, bus line 12 and bus line 39 stop at the same station A then the number of stops at A will be counted as two.

- Pedestrian Volume Data

We collect pedestrian volume data within 50 meters around the building from January 1, 2015 to November 31, 2015.

Due to the widespread use of mobile phones, the number of mobile phone users can accurately reflect the changes in pedestrian volume.

- Occupancy Data

We collect the average statistical data for each month in 2015. Occupancy data is the ratio of average number of people in a building to the total building accommodation capacity

- Time series Data

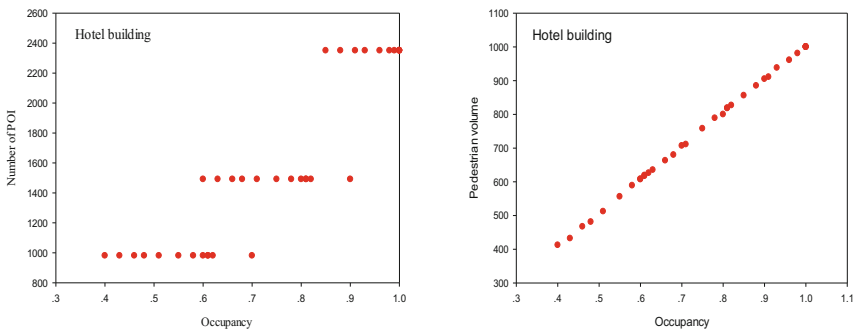
We divide the year into four quarters and use the vector to represent it. For example, the first quarter can be expressed as 1, 0, 0, 0.

## 4.2 Influence Factors

### 4.2.1 Scatter Diagram

The following scating diagrams disclose that the correlation of the occupancy between with pedestrian volume and number of POI.

Figure 3 is the scatter diagram of the three hotel buildings. Figure 4 is the scatter diagram of the three office buildings. Figure 5 is the scatter diagram of the three shopping buildings. The occupancy rate of various public buildings is highly correlated with number of POI and pedestrian volume. Therefore, number of POI and occupancy rate are the impact factors of building electric consumption.



(a) Scatter diagram of POI and average occupancy

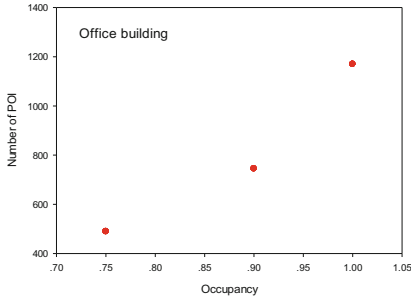
(b) Scatter diagram of pedestrian volume and occupancy

**Fig. 3.** Hotel building scatter diagram

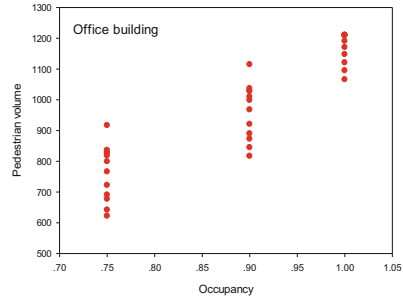
### 4.2.2 Remove Noisy Factors

Although the relationships between factors and electric consumptions can be represented by scatterplots, to accurately analyze the correlations between factors and public buildings' electric consumptions we use the gray relational analysis method introduced in Sect. 3.1 to remove the noise factors.



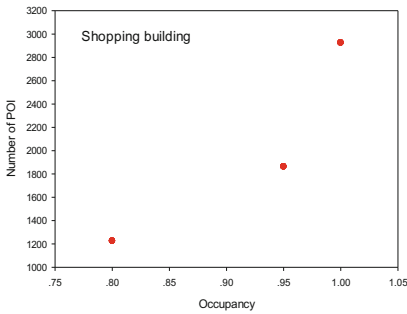


(a) Scatter diagram of POI and average Occupancy

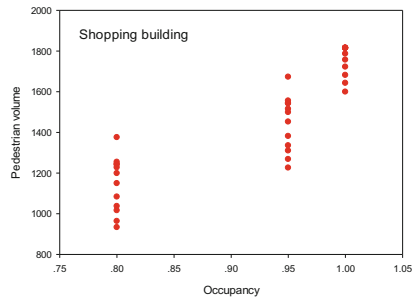


(b) Scatter diagram of pedestrian volume and occupancy

**Fig. 4.** Office building Scatter diagram



(a) Scatter diagram of POI and average Occupancy



(b) Scatter diagram of pedestrian volume and Occupancy

**Fig. 5.** Shopping building scatter diagram

As shown in Table 1, we extracted potential fifteen factors from the prepared multi-source data in Sect. 4.1. The Grey Correlation analysis is used to determine whether all fifteen factors ( $X_1, \dots, X_{15}$ ) listed in Table 1 have significant impacts on the underlying public buildings' electric consumptions.

The grey theory [33] has the advantage of using less data while producing higher accuracy. It has been widely studied and applied since its birth.

The grey correlation degrees between potential 15 factors and buildings' electric consumptions are shown in Table 2. According to the grey theory, the correlation degree above 0.5 (the threshold value) will be treated as key influence factors.

As shown in Table 2, it is interesting to see the correlations of influence factors vary significantly with building types. For instance, Pedestrian volume has the greatest impact on the electricity consumption of a shopping building, while its impact on an office building's one is minimal. The more pedestrians in a shopping building the more

**Table 1.** Fifteen factors

Daily electricity consumption	Age of building	No. of stories	Total area	Window-wall Ratio	Daily average temperature	Daily average humidity	Time series
X0'	X1	X2	X3	X4	X5	X6	X7
Pedestrian Volume	No. of office buildings	No. of malls	No. of restaurants	No. of hotels	No. of metro stations	No. of bus stations	Total No. of POI
X8	X9	X10	X11	X12	X13	X14	X15

**Table 2.** Analysis Result by Grey Theory

Factor	Correlation degree (Hotel building)	Correlation degree (Office building)	Correlation degree (Shopping building)
Age of building, x1	0.49	0.37	0.41
No. of stories, x2	0.43	0.62	0.33
Total area, x3	0.77	0.81	0.79
Window-wall Ratio, x4	0.57	0.55	0.43
Daily average temperature, x5	0.71	0.66	0.59
Daily average humidity, x6	0.49	0.38	0.31
Time series, x7	0.68	0.72	0.61
Pedestrian Volume, x8	0.58	0.48	0.71
No. of office buildings, x9	0.44	0.48	0.37
No. of shopping malls, x10	0.48	0.39	0.31
No. of restaurants, x11	0.59	0.66	0.69
No. of hotels, x12	0.48	0.41	0.37
No. of metro stations, x13	0.36	0.39	0.41
No. of bus stations, x14	0.67	0.70	0.69
Total No. of POI, x15	0.65	0.69	0.70

consumptions in that building. Nevertheless, during the normal working hours the number of staff members in an office building is relatively stable.

### 4.3 Prediction

Section 4.2 explores the key factors that influence public buildings' electric consumptions. The prediction of electric consumption is realized via probabilistic graphical model for the sake of sorting out the interrelations between the factors and their influences on the buildings' electric consumption. According to the modeling method given in Sect. 3.2, we construct the corresponding probability graph models for various public buildings. Figure 6 shows the probabilistic graphical model for three kinds of public buildings.

Based on the obtained probabilistic graph models, the parameter learning, and inference algorithms described in Sect. 3.2, we are able to predict the electricity consumption of nine public buildings for a given month. Our results are compared and analyzed based on the outcomes yielded by another classic forecasting algorithm: multivariable linear regression model. The prediction results are depicted in Fig. 7.

In Fig. 7, the consumption patterns predicted by our approach look similar to the actual electric consumptions. It is hard to tell the best forecasting model by visualization. In order to quantify the qualities further, we apply MAPE (Mean Absolute Percentage Error) to evaluate the algorithm.

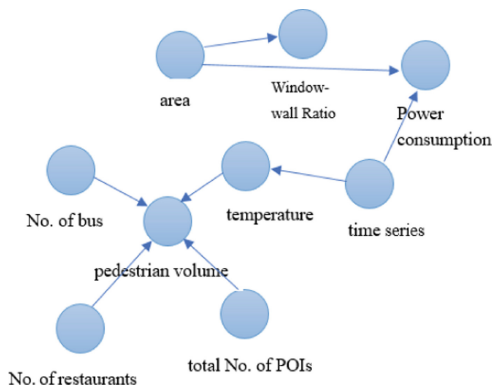
## 5 Error Analysis and Discussion

The mean absolute percentage error (MAPE) indicates the prediction accuracy of a forecasting method. Generally, a lower MAPE interprets better prediction accuracy.

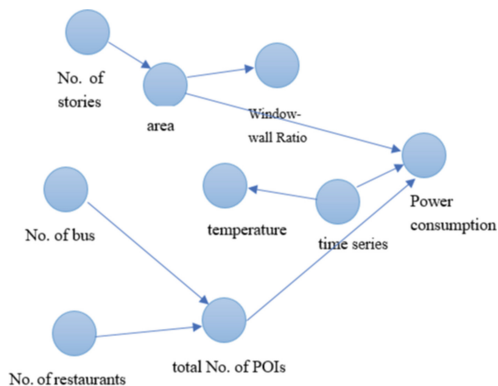
Table 3 summarizes the results yielded by our approach and other benchmarked approaches.

The average prediction error of our approach is 6.98%. The predictions' errors of other four methods are 7.86%, 7.69%, 8.29% and 14.99%, respectively. However, the error of five methods is within the recommended ASHRAE limits—30% for predictions [31]. TC method: total consumption forecast using the proposed ANN and only using total consumption data. EUs method: total consumption forecast using the proposed method and obtaining the prediction as the aggregation of the different EUs [32]. The urban multi-source based methods proposed in this paper produces the predictions with better accuracy while the datasets needed for the model are relatively easy to access. It is expected that in real applications, our approach would be easier to be deployed.

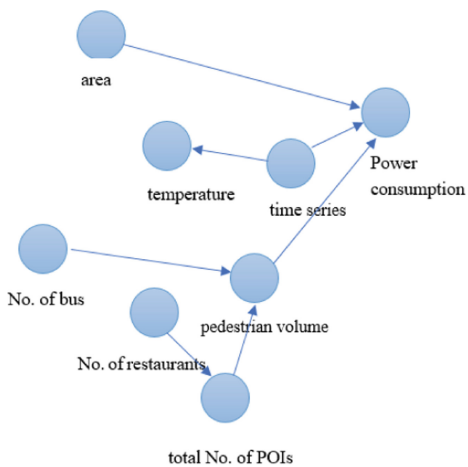
In addition, our approach extracts crucial factors from total fifteen potential ones and uses grey correlation analysis to conduct predictions. This indicates that reducing unimportant factors mitigates some noisy influence due to loosely-related factors and produces a better prediction accuracy. Using different critical influence factors, we have constructed the special probability graph models for various public buildings to help improving the prediction accuracy. Our approach uses the probabilistic graphical algorithm (EMG) is the combination of the Expectation-Maximization and Gibbs



(a) PGM structure of Hotel building

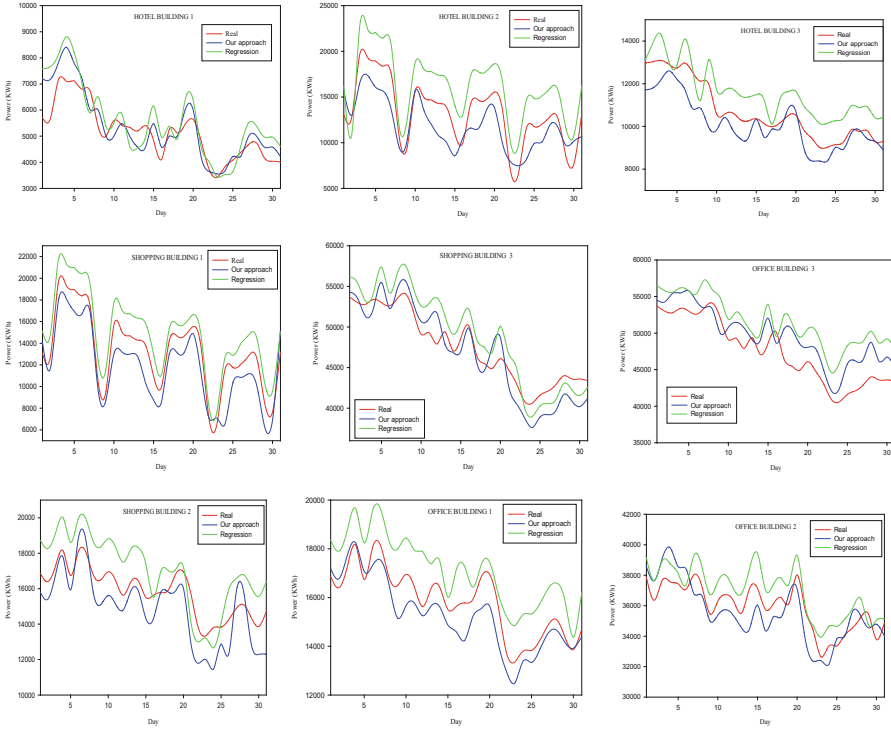


(b) PGM structure of Office building



(c) PGM structure of Shopping building

**Fig. 6.** Probabilistic Graphical models for various public buildings



**Fig. 7.** Forecast results of six buildings by different algorithm

**Table 3.** Experiment Results

Average Error difference	Methods				
	Our approach	(multivariable linear regression model)	(EUs method) Ref 32	TC method Ref 32	(Heuristic Method) Ref 32
MAPE (%)	6.98%	7.86%	7.69%	8.29%	14.99%

methods. On the average, the predictions of our algorithm produce less error other methods that demonstrates our algorithm’s capability of dealing with different types of public buildings.

## 6 Conclusion

In this paper, we investigate the influences of various factors on public buildings’ electric consumptions. The critical influencing factors ranging from the architectural properties to some spatiotemporal attributes such as POI, pedestrian volume, etc. are collected and studied. This research reveals the profound influence of spatiotemporal data on electric consumptions from a new perspective. Furthermore, integrating various building influencing

factors in our approach is unique and more efficient comparing to other methods. However, there are some issues to be addressed in the future research. For example, the dataset is still not big enough in terms of time span due to data acquisition restrictions and costs. The number of investigated buildings is relatively small. We will explore more how different forecasting algorithms fare as more data (longer time span, more public buildings, etc.) being collected and provide better insights in selecting suitable prediction methods under different circumstances for urban electric power demands.

**Acknowledgements.** This work was supported by CIUC and TJAD [grant number CIUC20150011] and National Natural Science Foundation of China [grant number 61271351].

## References

1. The World Bank: Indicators. <http://data.worldbank.org/indicator>. Accessed 06 Apr 2017
2. Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H., Saidur, R.: A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renew. Sustain. Energy Rev.* **33**, 102–109 (2014)
3. Santamouris, M., Cartalis, C., Synnefab, A., Kolokotsa, D.: On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—a review. *Energy Buil.* **98**, 119–124 (2015)
4. Jorn, K.G., Milan, P., Raúl, A.: Estimation and analysis of building energy demand and supply costs. *Energy Procedia* **83**, 216–225 (2015)
5. Moulay Larbi, C., Medjdoub, B., Michael, W., Raid, S.: Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: a review. *Renew. Sustain. Energy Rev.* **64**, 761–776 (2016)
6. Afees, A.S., Taofeek, O.A.: Modeling energy demand: Some emerging issues. *Renew. Sustain. Energy Rev.* **54**, 1470–1480 (2016)
7. Radu, P., Vahid, R.D., Jacques, M.: Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. *Energy Buil.* **92**, 10–18 (2015)
8. Fabiano Castro, T., Reinaldo Castro, S., Fernando Luiz Cyrino, O., Jose Francisco Moreira, P.: Long term electricity consumption forecast in Brazil: a fuzzy logic approach. *Socio-Econ. Plann. Sci.* **54**, 18–27 (2016)
9. Yamauchi, T., Michinori, K., Tomoko, K.: Development of quantitative evaluation method regarding value and environmental impact of cities. *Fujitsu Sci. Tech. J.* **50**, 112–120 (2014)
10. Liang, H., Yu, Z., Duncan, Y., Jingbo, S., Lei, Z.: Detecting urban black holes based on human mobility data. In: *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 1–10. ACM, Bellevue (2015)
11. Jing, Y., Yu, Z., Xing, X.: Discovering regions of different functions in a city using human mobility and POIs. In: *KDD 2012 Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 186–194. ACM, New York (2012)
12. Zheng, O.N., Charles, O.N.: Development of a probabilistic graphical model for predicting building energy performance. *Appl. Energy* **164**, 650–658 (2016)
13. Young Tae, C., Raya, H., Youngdeok, H., Young, M.L.: Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy Buil.* **111**, 184–194 (2016)

14. Nelson, F., Rafe Biswas, M.A.: Regression analysis for prediction of residential energy consumption. *Renew. Sustain. Energy Rev.* **47**, 332–343 (2015)
15. Ambera, K.P., Aslamb, M.W., Hussainc, S.K.: Electricity consumption forecasting models for administration buildings of the UK higher education sector. *Energy Build.* **90**, 127–136 (2015)
16. James, A.D., Willy, J.G., John, H.H., Daniel, C.S., Michael, J.S., Trenton, C.P., Maoyi, H., Ying, L., Jennie, S.R.: Impacts of climate change on energy consumption and peak demand in buildings: a detailed regional approach. *Energy* **79**, 20–32 (2015)
17. Xiaoshu, L., Tao, L., Charles, J.K., Martti, V.: Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach. *Appl. Energy* **144**, 261–275 (2015)
18. Akin, T., Borhan, M.S.: Short-term residential electric load forecast-ing: a compressive spatio-temporal approach. *Energy Build.* **111**, 380–392 (2016)
19. Cara, R.T., Rakesh, P.: Building-level power demand forecasting framework using building specific inputs: development and applications. *Appl. Energy* **147**, 466–477 (2015)
20. Kristopher, T.W., Juan, D.G.: Predicting future monthly residential energy consumption using building characteristics and climate data: a statistical learning approach. *Energy Build.* **128**, 1–11 (2016)
21. Ferlito, S., Mauro, A.G., Graditi, S., De Vito, M., Salvato, A., Buonanno, G., Di, F.: Predictive models for building’s energy consumption: an Artificial Neural Network (ANN) approach. In: XVIII AISEM Annual Conference, pp. 1–4, IEEE, Trento (2015)
22. Sandels, C., Widén, J., Nordström, L., Andersson, E.: Day-ahead predictions of electricity consumption in a Swedish office building from weather, occupancy, and temporal data. *Energy Build.* **108**, 279–290 (2015)
23. Yang-Seon, K., Jelena, S.: Impact of occupancy rates on the building electricity consumption in commercial buildings. *Energy Build.* **138**, 591–600 (2017)
24. López-Rodríguez, M.A., Santiago, I., Trillo-Montero, D., Torriti, J., Moreno-Munoz, A.: Analysis and modeling of active occupancy of the residential sector in Spain: an indicator of residential electricity consumption. *Energy Policy* **62**, 742–751 (2013)
25. Amir, K., Ram, R., Martin, F.: Determinants of residential electricity consumption: using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants’ behavior. *Energy* **55**, 184–194 (2013)
26. Zhifeng, Z., Chenxi, Y., Wenyang, C., Chenyang, Y.: Short-term photovoltaic power generation forecasting based on multivariable grey theory model with parameter optimization. In: *Mathematical Problems in Engineering*, pp. 1–9 (2017)
27. Ju-Long, D.: Control problems of grey systems. *Syst. Control Lett.* **5**, 288–294 (1982)
28. Koller, D.: *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, Cambridge (2009)
29. McLachlan, G., Krishnan, T.: *The EM Algorithm and Extensions*, 2nd edn. Wiley-Interscience press, New York (2008)
30. William, M.D.: A theoretical and practical implementation tutorial on topic modeling and gibbs sampling. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 642–647. Springer, Oregon (2011)
31. ASHRAE: *ASHRAE Guideline 14: Measurement of Energy and Demand Savings*, ASHRAE, Atlanta (2002)
32. Guillermo, E., Carlos, Á., Carlos, R., Manuel, A.: New artificial neural network prediction method for electrical consumption forecasting based on building end-uses. *Energy Build.* **43**, 3112–3119 (2011)