ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 121 (2015) 1318 – 1325
Procedia Engineering
www.elsevier.com/locate/procedia

9th International Symposium on Heating, Ventilation and Air Conditioning (ISHVAC) and the 3rd International Conference on Building Energy and Environment (COBEE)

Short-term Prediction of Power Consumption for Large-scale Public Buildings based on Regression Algorithm

Huilong Wang[a], Xing Lu[a], Peng Xu[a]*, Dengkuo Yuan[a]

[a] *School of Mechanical Engineering, Tongji University, Caoan Road 4800, Shanghai, 201804, China*

Abstract

Energy consumption monitoring and regulation on large-scale public buildings are always an indispensable part in building energy conservation. With the mounting establishment of the building sub-metering platform, massive amounts of historical power consumption data are provided. In this paper, different building types of six large-scale public buildings in Shanghai are selected, with their sub-metering data deeply analyzed. The concept of CDHs/HDHs (cooling/heating degree hours) is introduced and weekly prediction models of total building power consumption are proposed by the way of multiple linear regression algorithm which is relatively simple and easy to understand. The prediction models are validated to have great accuracy and general applicability in the paper, offering reliable instructions to the building facility manager and relevant competent authorities in terms of decision making and policy implementation.

Peer-review under responsibility of the organizing committee of ISHVAC-COBEE 2015

Keywords: Large-scale public buildings; Sub-metering; Power consumption prediction; Multi-linear re-gression algorithm;

1. Introduction

Large-scale public buildings generally refer to the building covering an area of over 20000m^2, using the central air-conditioning systems, diversified in types including office building, commercial building, education center and transportation hub, etc. The area of the large public buildings in China only account for 4% of the overall area in cities and towns, while the annual power consumption take up as much as 22% of the whole energy consumption [1].

* Corresponding author. Tel.: +86 136 0197 1494; fax: +86 136 0197 1494.
E-mail address: xupengwb@gmail.com

Peer-review under responsibility of the organizing committee of ISHVAC-COBEE 2015
doi:10.1016/j.proeng.2015.09.011

In 2011, the area of large public buildings is 7.97 billion m^2, among which the large-scale public building occupies 0.57 billion m^2. The energy consumption excluding heating consumption reaches 90-200 kWh/(m^2·a), roughly 2~4 times than normal public buildings [1]. Unless energy-saving measures being taken, by 2020, merely power consumption of newly-built large public buildings will increase to 200 billion kWh [2]. Thus it can be concluded that the energy consumption of large public buildings possesses a massive proportion in building energy consumption and it is high time that we adopted effective approaches to achieve conservation and efficiency.

Recent years, sub-metering has been widely promoted by the encouragement of relevant government departments. Since 2005, thousands of large public buildings in big cities have installed sub-metering to implement real-time monitoring. Through wireless data acquisition and transmission, sub-item energy consumption in large public buildings can be precisely measured, tackling the problem of lacking detailed sub-item energy consumption data on building energy conservation work [3]. Undoubtedly, energy consumption monitoring platform offer us overwhelming historical data. However, what majority of the monitoring systems are facing at present is the dilemma of data accumulation. Hence, the digging and analysis of the sub-metering data become increasingly important.

We can take full advantage of the data from the sub-metering platform to figure out the characteristics of a certain building and make a study and forecast of energy consumption on the whole-building level and subsystem level. Further diagnosis of the abnormal consumption can be realized by utilizing the prediction model and the conservation strategy can be proposed by analyzing the reason for the abnormal consumption. Moreover, the result of the prediction model can offer reliable instructions to the building facility manager and relevant competent authorities in terms of decision making and policy implementation.

Numerous papers concerned with methods used in building energy prediction can be found in academic databases. The methods can be categorized into the following groups, for example, engineering methods, regression analyses, artificial neural networks (ANNs), support vector machines (SVMs) and grey models [4,5]. Engineering methods are on the basis of physical principles to calculate energy performance on the whole building level or for sub-level components. Simulation programs calibrated with real measured data, such as DOE-2, EnergyPlus, BLAST, ESP-r, are often used to predict the energy consumption. Regression methods use a simple or multivariable regression analysis to correlate the outputs to inputs parameters like climate information, operation data and occupancy behavior. ANNs and SVMs are both intelligent computer systems, which developed from machine learning algorithms that are capable of 'make decisions' with an interpretation of data. ANNs are good at solving nonlinear problems to predict building energy consumption. SVMs are effective in solving nonlinear problems even with small quantities of training data.

Each model has its own pros and cons in certain cases of applications [6]. Engineering method is the only method that can fully develop the energy consumption of the sector without any historical energy consumption information [7]. But a well-accepted demerit of detailed engineering model lies in that it is difficult to perform in practice on account of the lack of input information. Intelligent computer systems provide a better prediction for the energy consumption in spite of high complexity. Nevertheless, they are not based on the physical principles and run as a black-box system which makes the interpretability very difficult [8]. As for the regression model, it is easy to develop but it is commonly acknowledged that they are relatively inaccurate and lack of flexibility. The scholars in the past built the prediction model taking various factors into consideration to discover a universal regression equation which has a fine fitness with the buildings in the same type. But the predictive results in most cases are not satisfied in the practice due to the distinct structure, systems as well as the operation time of different buildings. To solve the problem, they had to introduce complex correction coefficient. Nowadays, we are able to set up prediction model of each building on the ground of historical data on the sub-metering platform, thus enhancing the accuracy of the outcome to a certain extent.

To put it in a nutshell, it is really necessary for us to find a simple but effective energy prediction model. In this paper, an accurate prediction to the weekly power consumption of the large-scale buildings is proposed, using the multi-linear regression algorithm with sub-metering data deeply analyzed. According to the result of the prediction, the model is applicable to different types of public buildings in Shanghai, including commercial buildings, office buildings and building complex, setting solid foundation for the large-scale public buildings in energy efficiency diagnosis, energy saving inspection and supervision work.

Nomenclature

AARD	average absolute relative deviations
RMSD	relative mean square deviation
HDDs	heating degree days
CDDs	cooling degree days
CDHs	cooling degree hours
HDHs	heating degree hours
Tm	daily mean temperature
OB	office building
CB	commercial building
BC	building complex
C	constant

2. Methods

2.1. Regression algorithm

Relevant independent variables that define the energy consumption are selected to develop models obtained by means of regression models, which are validated and discussed for future predictions and reproducibility.

Take the multiple linear regression for example, suppose variable Y as the power consumption of the building and variable X1, X2 as the impact factors, the general regression formula is as shown as follow :

$$\mathrm{Y} = b_0 + b_1 X_1 + b_2 X_2 + \varepsilon \tag{1}$$

Where b_0, b_1, b_2 are the regression coefficients, where ε is the regression residual.

2.2. Predictive validation

The average absolute relative deviations (AARD) and the relative mean square deviation (RMSD) are popular evaluation criterion to analyze the accuracy of the prediction. The effect of the prediction model not only depends on the bias degree of the prediction error but also the variation of error also called variance. The average relation error can reveal the bias degree of the prediction error quite well and the RMSD is also a good indication of the variance.

$$\mathrm{AARD}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{y_i}\right) \tag{2}$$

$$\mathrm{RMSD}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{y_i}\right)^2} \tag{3}$$

Where n is the number of the data, where y_i is the actual value, where $\hat{y}_i$ is the predicted value.

2.3. Influence factors

Power consumption of the large-scale public building is influenced by many aspects, for instance, external climate, urban microclimate, architectural planning, thermal performance of the building envelope, the building type and the human use [9]. The first few are considered as factors related to temperature and the last is with regards to the working duration.

Degree days (DD) represent a versatile climatic indicator frequently used in the analysis of building energy performance, which require less data input and can be used to assess rapidly how energy consumption may be influenced by major design decisions (e.g. insulation level, glazing area rate of the building, assumptions about infiltration, etc.) [10]. Heating degree days (HDDs) are defined as the deviation of the outdoor mean temperature from a heating reference temperature, taking into account only positive values. The reference temperature is also known as the base temperature which, for buildings, is a balance point temperature, that is, the outdoor temperature at which the heating systems do not need to run so as to maintain comfort conditions. Likewise, cooling degree days (CDDs) are calculated from temperatures above the base temperature. In this case, a base temperature is regarded as the outdoor temperature below which a building needs no cooling. However, in this paper, we introduce the cooling degree hours (CDHs) and heating degree hours (HDHs), similar concept of CDDs and HDDs owing to the fact that the precision of degree days is not sufficient enough for the short-term prediction. For a short period of time (daily, weekly, etc.), accumulated cooling/heating degree-hours (ACDHs/AHDHs) are calculated using the following expression:

$$ACDH = \sum_{j=1}^{\mathrm{N}} (CDH_j) \begin{cases} if \;\; T_j > T_b \;\; then \; CDH = T_m - T_j \\ \qquad\qquad\quad else \; CDH = 0 \end{cases} \tag{4}$$

$$A\mathrm{H}DH = \sum_{j=1}^{\mathrm{N}} (\mathrm{H}DH_j) \begin{cases} if \;\; T_j < T_b \;\; then \; HDH = T_b - T_j \\ \qquad\qquad\quad else \; HDH = 0 \end{cases} \tag{5}$$

Where N is defined as the period of time i.e. number of hours in the week. The corresponding number of hours for the accumulated degree-hours for any period of time is determined by summation of the hours with the difference between T_j and T_b.

For office buildings and building complex, there exists an apparent difference in power consumption level between workdays and weekends. CDH, HDH and DAY are chosen as the impact factors in the regression model for weekly prediction, where DAY represents the number of workdays in a week. As for large commercial buildings, the flow of customers increases rapidly in the weekend, inevitably exerting great influence on the power consumption. In view of the huge inner zone and high internal heat, we choose CDH and DAY as the impact factors in the regression model, leaving the significance level of respective factors to be tested.

2.4. Season division

The whole year can be divided into cooling season, heating season and transition season. The base temperature discussed above cannot be determined arbitrarily or customarily. We can infer the base temperature from the historical power consumption data on the sub-metering platform. From Figure 1, it can be deducted that the commercial buildings begin cooling when the outside air temperature is below 15°C. For the office buildings and building complexes, air-conditioning system starts when the outdoor air temperature is over 20°C, and heating starts as long as the outdoor temperature is below 12°C. So for the office buildings and building complexes, the day is defined as the cooling day when Tm (daily mean temperature) is above 20°C. Likewise, the day is defined as the heating day when Tm is lower than 12°C, otherwise it is deemed as a transition day. For the commercial buildings, the day is defined as the cooling day when Tm is above 15°C, otherwise it is defined as a transition day. The detailed season division approach of a day based on Tm is tabulated in Table 1.

Table 1 Season division approach based on T_m

Building type	Cooling day	Heating day	Transition day
OB	>20	<12	12~20
CB	>15	--	≤15

BC	>20	<12	12~20

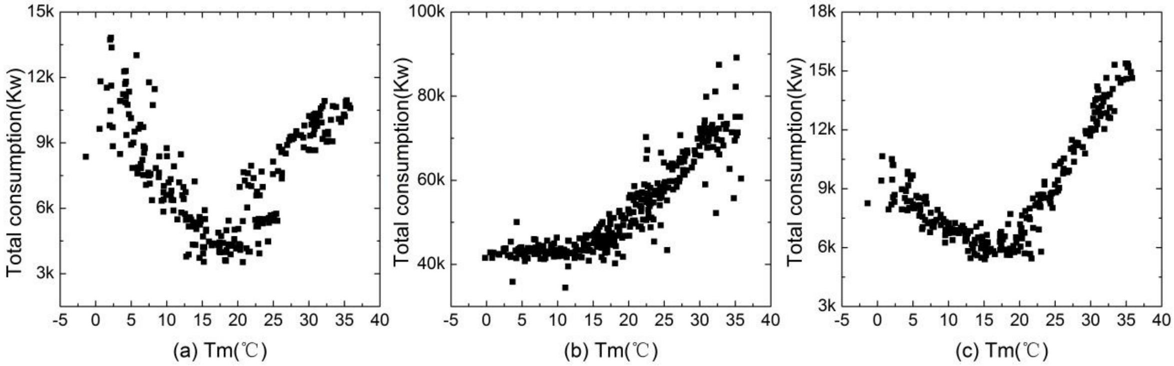

Fig 1. Scatter diagram of total power consumption corresponding to T_m in a year of (a) an office building (b) a commercial building (c) a building complex

The season division of a week depends on the most type of the days in this week. That is to say, if there exists five cooling days and the rest are transition days, then this week is divided into cooling season. For commercial building, the base temperature in this week is correspondingly 15°C. According to the division approach mentioned above, we will omit the discussion of the transition weeks considering that they are few in one year and energy-saving potential is not considerable.

3. Results and discussions

The result is shown inTable.2 after training the regression model according to the sub-metering weekly data in 2013 and testing it according to the data in 2014.

Table 2 Regression results of total power consumption in types of large public buildings

Building type	Consumption type	Impact factor	Regression coefficient	F-test	t-test	R^2
OB1	Cooling weeks	C	18481.32	2.17E-09	0.003493	0.86
		DAY	2033.565		0.098986	
		CDH	15.24238		1.51E-09	
	Heating weeks	C	-6667.5	2.58E-05	0.411892	0.83
		DAY	7466.171		1.42E-04	
		HDH	22.74376		1.26E-04	
CB1	Cooling weeks	C	176002.7	1.49E-20	2.68E-16	0.95
		DAY	-1135.6		0.622632	
		CDH	47.64427		2.42E-21	
BC1	Cooling weeks	C	13811.4	2.11E-20	1.75E-06	0.99
		DAY	4892.462		4.52E-10	
		CDH	20.43565		4.68E-20	
	Heating weeks	C	24106.84	8.98E-06	1.09E-05	0.86
		DAY	2802.804		0.000423	
		HDH	12.54942		1.13E-05	
OB2	Cooling weeks	C	85931.82	1.62E-12	2.59E-07	0.92
		DAY	17492.49		5.60E-07	
		CDH	37.07204		6.37E-11	
	Heating weeks	C	68027.67	4.01E-07	5.44E-07	0.90

		DAY	14019.38		1.95E-07	
		HDH	15.61237		0.000244	
		C	298278.2		1.50E-19	
CB2	Cooling weeks	DAY	-355.182	4.73E-22	0.905672	0.96
		CDH	69.89282		8.07E-23	
BC2	Cooling weeks	C	78865.14	5.32E-10	4.24E-05	0.88
		DAY	12777.49		0.000962	
		CDH	40.09		4.79E-09	
	Heating weeks	C	21055.78	4.03E-06	0.390374	0.85
		DAY	20815.48		0.000447	
		HDH	78.46933		3.97E-06	

The table illustrates that the model result fits the historical power consumption data with an average coefficient of a high multiple determination (R2-value). Testing the model by F-test, a validation for the significance of all regression coefficients, we can draw the conclusion that all factors have remarkable effects on the building power consumption. Using t-test, which represents the significance of a single variable, we can see that the variable DAY of commercial buildings is not significantly correlated with the building power consumption. An acceptable reason is that the influence of weekend is insignificant on the power consumption though customers and sales surge at weekends. So we redo the regression process after eliminating the factor-DAY. The result is presented in Table 3.

Table 3 Modified regression results of total power consumption in commercial buildings

Type	Season	Impact factor	Regression coefficient	F-test	t-test	R^2
CB1	Cooling weeks	C	170720.9	5.58E-22	1.99E-31	0.95
		CDH	47.53724		5.58E-22	
CB2	Cooling weeks	C	296627.7	1.4E-23	2.28E-35	0.96
		CDH	69.8585		1.40E-23	

As seen from the table, the fitness of the model is still good. For there is only one factor left, F-test and t-test are the same, indicating that the CDH has a remarkable influence on the power consumption of the commercial building.

We use the trained prediction model to predict the power consumption in 2014 and validate the prediction effect by real-time monitoring data. Table 4 demonstrates the predictive effect of total power consumption of different buildings types in cooling season.

Table 4 Predictive effect of models of different buildings types in cooling season

Type	AARD	RMSD	Type	AARD	RMSD	Type	AARD	RMSE
OB 1	-2.70%	12.84%	CB 1	0.26%	2.67%	BC 1	-2.51%	4.93%
OB 2	10.98%	11.44%	CB 2	-3.93%	5.40%	BC 2	5.55%	6.96%

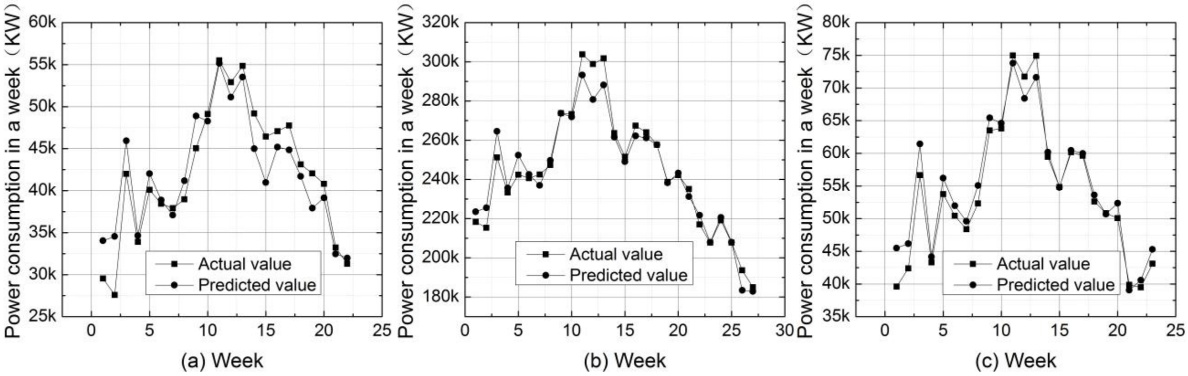

Fig 2. Curve-fitting of actual and predicted total power consumption in cooling season in (a) an office building (b) a commercial building (c) a building complex

The accuracy of the prediction model is satisfactory according to the data and curve above (Figure 2), and it is applicable to the public buildings including office buildings, building complexes and commercial buildings. The result of the prediction has a high reliability which can forecast the consumption of the next week. This method can be popularized to other region along with the mounting establishment of the building sub-metering platform.

4. Conclusions

By establishing sub-metering platform, real-time monitoring of building power consumption can be achieved. Massive power consumption data will be provided, laying a good foundation for the future research. By means of regression algorithm, the short-time prediction of power consumption in large-scale public buildings can also be realized, offering a gist for the building facility managers and relevant decision-makers. But we should make classification of the public buildings before the regression because of their variance in construction features and operating characteristics, therefore improving the accuracy of the prediction.

Short-term prediction models of building power consumption in cooling and heating season are established according to the sub-metering consumption of six large-scale public buildings in Shanghai. The weekly power consumption of office buildings and building complexes mainly depends on the factors of temperature and workdays in a week, while, for the commercial buildings, it only depends on the temperature for the impact of the workdays in whole week is not remarkable. The impact of temperature can be described by CDH and HDH. The prediction models are validated to have great accuracy and general applicability which provide a good foundation for the large-scale buildings in diagnosing, energy monitoring, energy-saving reform and regulatory work.

Acknowledgements

The authors gratefully acknowledge the support provided by Shanghai Science and Technology Committee through the Project 'Innovative Research and Practice of Energy Saving in Large-scale Public Building and Energy Monitoring Platform in Shanghai Central Area'.

References

[1] Building energy conservation research center, Tsinghua University (BECRC), 2012 Annual report on China building energy efficiency, 2012, 262 pages.
[2] Y. Jiang, Z. Jiang and Q. Wei 2010, Energy management in large public buildings & energy-saving diagnosis technique, Construction science and technology 22(2010) 20-23.
[3] X. Chen, Application of sub-metering data based on the energy efficiency monitoring system. Building energy efficiency, 08(2014), 86-92.
[4] H. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew Sustain Energy Reviews 16(2012) 86–92.
[5] ASHRAE, ASHRAE handbook chapter 19– fundamentals, Energy estimating and modeling methods, 2009.
[6] F. Nelson, A review on the basics of building energy estimation. Renewable and Sustainable Energy Reviews 31(2014) 53–60.

[7] G. Lukas and U. Ismet Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews 13(2009) 1819-1835.
[8] H. Alberto, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy and buildings 40(2008) 2169–2176.
[9] Y. Ma, Research on energy-consumption forecasting and energy-saving management system for the large-scale public buildings, Master Thesis, Xian University of Architecture and Technology, China, 2011, 74 pages.
[10] M. Monjur, Relationship between annual mean temperature and degree-days. Energy and buildings 54(2009) 418-425.