Short-term Electricity Load Forecasting for Building Energy Management System

Phanumat Saatwong Surapong Suwankawin
Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University

Abstract

Short-term electricity load forecasting is essential for Building Energy Management System (BEMS) in various aspects, e.g. peak-shaving application, planning for self-consumption with renewable energy, net-zero energy building. This paper presents the forecast models for load demand in buildings by using the time-series approach. The load-forecast models are created from the step-by-step procedure of Box-Jenkins Methodology and the Seasonal Autoregressive Integrated Moving Average (SARIMA) models are obtained. The resultant models are evaluated with the actual load of Electrical Engineering Building at Chulalongkorn University. The proposed models can fairly forecast the load pattern for the workdays with roughly 20% Mean Absolute Percentage Error (MAPE). In addition, the models are moderately successful to predict the peak-load instant.

Keywords: Short-term Electricity Load Forecasting; Box-Jenkins Methodology; Seasonal Autoregressive Integrated Moving Average Model; Building Energy Management System.

1. Introduction

In Thailand, commercial building sector governs high electricity consumption. [1] Therefore, a lot of Building Energy Management Systems (BEMS) are conducted in order to not only reduce the electricity bills, but also to help defer the construction of new power plants. Electricity bills of commercial buildings, including energy charges and demand charges. To reduce the demand charges, Battery Energy Storage System (BESS) is usually employed as shown in Fig. 1. The load profiles of building are needed to achieve the peak-shaving application.

Electricity load forecast is an essential component to generate the command signal for BESS to successfully reduce the peak demand. Short-term electricity load forecast predicts the future value from hours to weeks for energy management. Various models are employed in the literature which can be classified into two main types; the linear models and the non-linear models, L. Hernandez et al. [2]. The linear models are typically based on the time series approach. L. Jong Hun et al. [3] propose an exponential smoothing linear model to forecast the daily load along with temperature in building. J. Massana et al. [4] mention the Multiple Linear Regression (MLR) linear model and Y. Penya et al. [5] draw a comparison between the Autoregressive (AR) linear model and other non-linear models and point out that the AR model can provides a better performance. On the other hand, the non-linear models are usually based on the Artificial Intelligent approach. Y. Tae Chae et al. [6] present the using of an Artificial Neural Network (ANN) model to forecast the 15-minute electricity load in a building by which the moderate prediction results can be obtained in both daily load profile and daily peak load. J. Massana et al. [4] also introduce a model based on the multi-layer perceptron (MLP) for the hourly load forecast.

Regarding the practical point of view, the time-series approach is preferred due to its systematic methodology with rigorous analysis from the basis of statistics and probability theory. In order to employ the suitable linear model for forecasting task, the property of load profile of building should be considered. The explicit properties of the time series of load profile are 1) correlated observation; their values are statistically dependent upon each other, 2) non-stationary; no fixed mean, and 3) daily seasonal.

In this paper, the seasonal autoregressive integrated moving average (SARIMA) model is used to forecast the future load profile. A forecast model is developed by using the Box-Jenkins methodology. The advantages of this approach for load forecast are as follows:

1) Box-Jenkins methodology has procedures to deal with the correlated time series, therefore it is rather more suitable than the regression and exponential smoothing approaches.

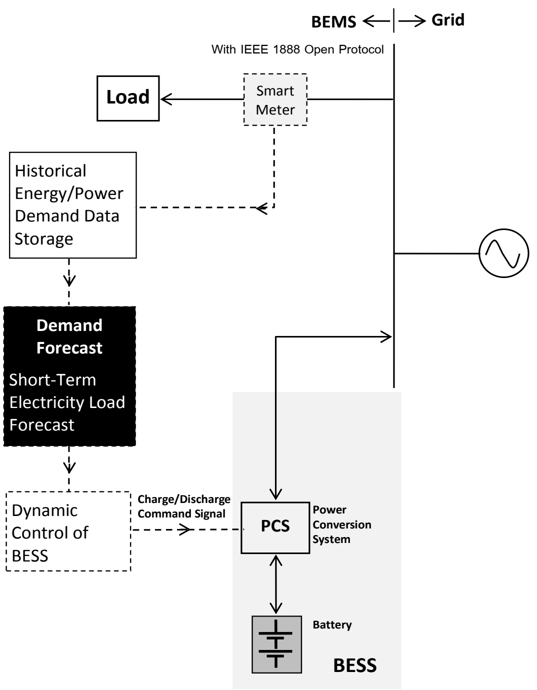

Fig. 1 Demand Forecast for Chulalongkorn University Building Energy Management System (CUBEMS).

2) Non-stationary model for no-fixed mean can be conducted by Autoregressive Integrated Moving Average (ARIMA) process.

3) Box-Jenkins methodology provides the models that are useful in forecasting time series having seasonal variation.

2. Electricity Load Profile

The electricity usage of electrical engineering building at Chulalongkorn University is considered as a study case. The patterns of electricity consumption of university buildings are dependent on the working days, holidays, seasons and activities. Because load profile has a daily pattern, we need to use historical load data for forecasting. The historical demand is measured during 9 months (1 April 2015 – 31 December 2015) and stored in the data storage of the CUBEMS (Chulalongkorn University Building Energy Management System) as shown in Fig. 1. The raw data is sampled every one minute and the 15-minute averaged power demand is calculated as shown in Fig. 2 and Fig. 3 respectively. It can be seen in Fig. 2 that the peak demand mostly concentrates in April, which is correspond to the country's peak demand in the summer season, and gradually decreases from May to July. The electricity consumption increases again in August due to the university's classes starting in the first semester.

Fig. 3 shows the 15-minute averaged power demand during one week. The electricity consumption is rather high in workdays (Monday-Friday) while the averaged electricity load is less than 10 kW at weekends. In addition, the electricity load on each workday starts to increase at 7:00 AM and dies down at 8:00 PM. The peak load occurs between 10:00 AM and 4:00 PM.

In this case study, the 5-week of workday's load data in Fig. 4 (between Aug. 3rd, 2015 and Sep. 4th, 2015) is used as the historical data for load forecasting. And the future one-week load data of 5 workdays (from Mon. Sep. 7th, 2015 to Fri. Sep. 11th, 2015) is then estimated.

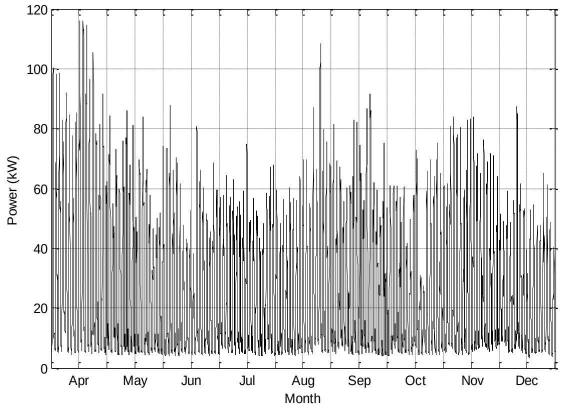

Fig. 2 Power demand on workdays of Electrical Engineering Building at Chulalongkorn University during 9 months (April-December 2015).

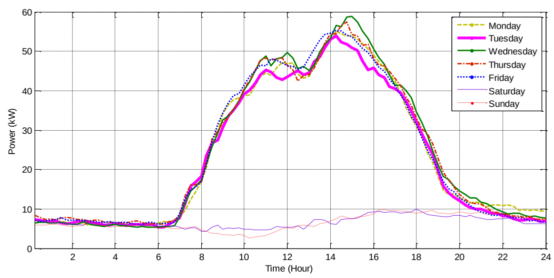

Fig. 3 Averaged daily electricity load in one week of Electrical Engineering Building at Chulalongkorn University.

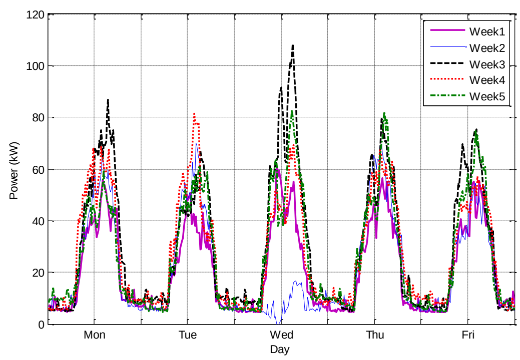

Fig. 4 Averaged daily electricity load for workdays of Electrical Engineering Building at Chulalongkorn University.

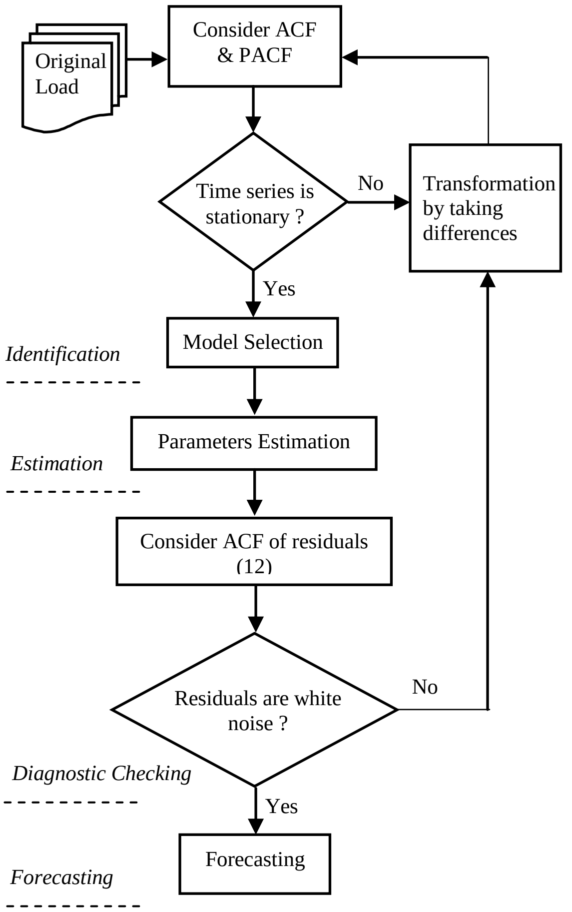

Fig. 5 Flow chart of basic steps of Box-Jenkins Methodology.

3. Forecasting Model by Box-Jenkins Methodology

The load data is denoted by a time series at equally spaced time $t, t-1, t-2, ...$ by $y_t, y_{t-1}, y_{t-2}, ...$. Since the time series of load y_t is a correlated time series, the forecasting model can be assumed by the ARMA (Autoregressive and Moving Average) models as follow:

$$y_t = \underbrace{\mu + \phi_1 y_{t-1} + ... + \phi_p y_{t-p}}_{Autoregressive\ (AR)\ Model} + \underbrace{\varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}}_{Moving\ Average\ (MA)\ Model} \quad (1)$$

where μ is constant mean, ε_t is white noise error term, p is the order of autoregressive model and q is the order of moving-average model.

The general transformation that produces stationary time series values can be written as

$$z_t = \underbrace{\nabla_L^D}_{\substack{seasonal\\difference\\operator}} \underbrace{\nabla^d}_{\substack{non-seasonal\\difference\\operator}} y_t = \underbrace{(1-B^L)^D}_{\substack{seasonal\\difference\\operator}} \underbrace{(1-B)^d}_{\substack{non-seasonal\\difference\\operator}} y_t \quad (2)$$

where B is backward shift operator, ∇ is the non-seasonal difference operator, d is the order of non-seasonal differencing, ∇_L is seasonal difference operator, D is the order of seasonal differencing and L is the seasonal period. And the general Seasonal Autoregressive Integrated Moving Average (SARIMA) model can be given in (3).

$$\underbrace{\phi_p(B)}_{\substack{Generalized\\Non\ Seasonal\\AR\ opeator}} \underbrace{\Phi_P(B^L)}_{\substack{Seasonal\\AR\ opeator}} z_t = \underbrace{\theta_q(B)}_{\substack{Non-seasonal\\MA\ operator}} \underbrace{\Theta_Q(B^L)}_{\substack{Seasonal\\MA\ operator}} \varepsilon_t \quad (3)$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p \quad (4)$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q \quad (5)$$

$$\Phi_P(B^L) = 1 - \Phi_{1L} B^L - \Phi_{2L} B^{2L} - ... - \Phi_{PL} B^{PL} \quad (6)$$

$$\Theta_Q(B^L) = 1 - \Theta_{1L} B^L - \Theta_{2L} B^{2L} - ... - \Theta_{QL} B^{QL} \quad (7)$$

The order of this general multiplicative seasonal model is ARIMA $(p, d, q) \times (P, D, Q)_L$. Box-Jenkins methodology [7-9] provide the systematic procedure to estimate the unknown parameters in the model, it consists of four basic steps including identification, estimation, diagnostic checking and forecasting as shown in Fig. 5.

4. Representation of the Load Data by Multiplicative Seasonal Models

In this paper, the historical load data is collected during Aug. 3rd, 2015 to Sep. 4th, 2015. The load data on weekends is excluded. The validity of resultant models is verified by the load prediction for the next 5 workdays (Mon-Fri) between Sep. 7th and Sep 11th, 2015. There are two case studies that are conducted in order to represent the model of the load data. The difference between these two case studies is due to the usage of historical load data. The first case study uses all historical data together to represent a single model that is used to predict the daily load profile for 5 workdays. On the other hand, the second case study separates the historical data into five sets according to each workday (Mon-Fri), and five models are then obtained. These five models are employed to forecast the load of 5 workdays between Sep. 7th and Sep 11th, 2015.

Following the basic steps of Box-Jenkins methodology in Fig. 5, the details of model creation for the first case study are proceeded.

4.1 1st step: Identification

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used to investigate whether the time series data is stationary or non-stationary. The ACF and the PACF are calculated from original historical load data and are plotted in Fig. 6. It unfortunately reveals that the original load data y_t is non-stationary time series, since the ACF dies down slowly with oscillation as shown in Fig. 6.

One route cause of this non-stationary behavior is the daily seasonal property of the load data. In order to transform non-stationary time series (y_t) to stationary time series (z_t), the first seasonal differences ($D=1$) are taken with the daily period $L=96$ (there are 96 averaged-load values per day, since the load demand is averaged for every 15 minutes). It is observed that, with the first seasonal differences, the transformed values of ACF die down quickly without spike at seasonal lags $2L, 3L, 4L, \ldots$. However, the transformed values die down slowly between lag 0 and seasonal lag $1L$. This implies that the transformed values are still non-stationary. The first non-seasonal differences ($d=1$) are additionally taken as written in (8).

$$z_t = \underbrace{\nabla_{96}^{1}}_{\substack{seasonal\\ difference\\ operator}} \underbrace{\nabla^{1}}_{\substack{non-seasonal\\ difference\\ operator}} y_t = \underbrace{(1-B^{96})^{1}}_{\substack{seasonal\\ difference\\ operator}} \underbrace{(1-B)^{1}}_{\substack{non-seasonal\\ difference\\ operator}} y_t \quad (8)$$

Fig. 7 well confirms that the transformation in (8) can produce stationary time series z_t, this is because the transformed values of ACF die down quickly even between lag 0 and seasonal lag $1L$. To identify the orders of SARIMA model, Box-Jenkins methodology provides guidelines by examining the ACF and PACF values in Fig. 7. The sample ACF cuts off after lag 2; a spike appears at lag 2, and the sample PACF dies down quickly after lag 2. This behavior reflects the characteristic of non-seasonal MA model with the order of 2 $(q=2)$. The order of seasonal model can be examined from the significant spikes at lag 96 $(1L)$ of sample ACF and at lags $96\,(1L)$, $192\,(2L)$, $288\,(3L)$, $384\,(4L)$ and $480\,(5L)$ of the sample PACF. The sample ACF cuts off after lag $1L$ and the sample PACF dies down after lags $1L$ in a damped exponential fashion with oscillation at lags $2L$, $3L$, $4L$ and $5L$. This reflects the seasonal MA model with the order of 1 $(Q=1)$. It can be concluded that the identification process is completed with the selection of ARIMA $(0,1,2)\times(0,1,1)_{96}$ as shown in (9).

$$\nabla_{96}^{1}\nabla^{1}y_t = \theta_2(B)\Theta_1(B^{96})\varepsilon_t \quad (9)$$

Table 1 Estimated Parameters for ARIMA $(0,1,2)\times(0,1,1)_{96}$

Parameter	Value
θ_1	-0.230492
θ_2	-0.444753
Θ_1	-0.7912

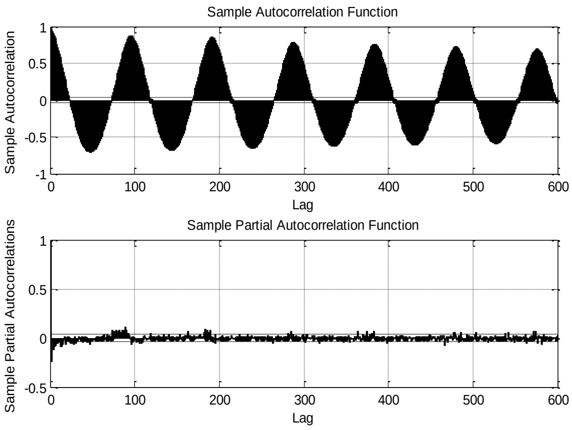

Fig. 6 ACF and PACF for the original load data.

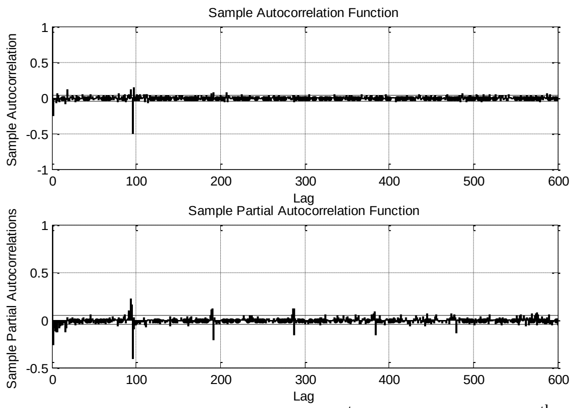

Fig. 7 ACF and PACF for of the 1st-differenced and 96th-differenced transformed values.

4.2 2nd step: Estimation

The SARIMA model in (9) can be rewritten in (10) and (11).

$$(1-B^{96})(1-B)y_t = (1-\theta_1 B-\theta_2 B^2)(1-\Theta_1 B^{96})\varepsilon_t \quad (10)$$

$$y_t = \underbrace{y_{t-1}+y_{t-96}-y_{t-97}}_{AR\ Model} + \underbrace{\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\Theta_1\varepsilon_{t-96}+\theta_1\Theta_1\varepsilon_{t-97}+\theta_2\Theta_1\varepsilon_{t-98}}_{MA\ Model} \quad (11)$$

There are three parameters that should be estimated θ_1, θ_2 and Θ_1. These estimates are obtained by using a least square criterion. The Matlab program is applied for the iterative search procedure in order to find the least square estimates. Finally, the model parameters are carried out in Table 1.

4.3 3rd step: Diagnostic Checking

The adequacy of tentative model in (11) is tested prior to using it to forecast. The diagnostic checking is performed by the calculation of the ACF of residual as shown in (12).

$$r_k = \frac{\sum_{t=a}^{n-k}(e_t-\bar{e})(e_{t+k}-\bar{e})}{\sum_{t=a}^{n}(e_t-\bar{e})^2} \quad (12)$$

where e is the differences between the original load data and the predictions from tentative model in (11). In the sake of simplicity, if the values of ACF of residuals, at all lags, are nearly zero, it implies that the average of residuals is zero and white noise and the model can be considered to be adequate. The ACF of residuals is depicted in Fig. 8, it can be observed that the values of ACF at some lower lags 2 and 3 are larger than the threshold, and the residuals of this model are non-white noise and therefore, the ARIMA $(0,1,2)\times(0,1,1)_{96}$ is deemed inadequate.

Nevertheless, the Box-Jenkins methodology provides the procedure to make the improvement of the model. The model is modified to be ARIMA $(1,1,2)\times(0,1,1)_{96}$. The order of non-seasonal AR model is increased to be equal to one, this is to arrange more influence from the non-seasonal AR model. The modified ARIMA $(1,1,2)\times(0,1,1)_{96}$ can be written in (13)-(15).

$$\phi_1(B)\,\nabla_{96}^1\nabla^1 y_t \;=\; \theta_2(B)\Theta_1(B^{96})\,\varepsilon_t \tag{13}$$

$$(1-\phi_1 B)(1-B^{96})(1-B)\,y_t \;=\; (1-\theta_1 B-\theta_2 B^2)(1-\Theta_1 B^{96})\,\varepsilon_t \tag{14}$$

$$\begin{aligned} y_t = &\underbrace{y_{t-1}+\phi_1 y_{t-1}-\phi_1 y_{t-2}+y_{t-96}-y_{t-97}-\phi_1 y_{t-97}+\phi_1 y_{t-98}}_{Modified\ AR\ Model} + \\ &\underbrace{\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\Theta_1\varepsilon_{t-96}+\theta_1\Theta_1\varepsilon_{t-97}+\theta_2\Theta_1\varepsilon_{t-98}}_{MA\ Model} \end{aligned} \tag{15}$$

The procedures of parameter estimation and diagnostic checking are repeated over again. The modified SARIMA model contains four parameters ϕ_1, θ_1, θ_2 and Θ_1 and the Table 2 indicates the estimated parameters and Fig. 9 illustrates the ACF of residuals of the modified model, it can be seen that all samples of ACF residuals are less than the threshold and nearly equal to zero. This means that the modified ARIMA $(1,1,2)\times(0,1,1)_{96}$ in (15) is adequate and appropriate model and can be adopted for the forecasting process.

4.4 4th step: Forecasting

The model in (15) with the parameters in Table 2 is used to forecast future load values. Fig. 10 depicts the comparison between the actual load data and the forecasted load for 5 workdays. The resultant model in (15) can fairly predict the loads. To evaluate the accuracy of the forecast model, the Mean Absolute Percentage Error (MAPE) is used to calculate the forecast error as follow:

$$MAPE \;=\; \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t-\hat{y}_t|}{y_t}\times 100 \tag{15}$$

where y_t is actual load at time t, $\hat{y}_t$ is forecast load at time t and n is the number of forecast load.

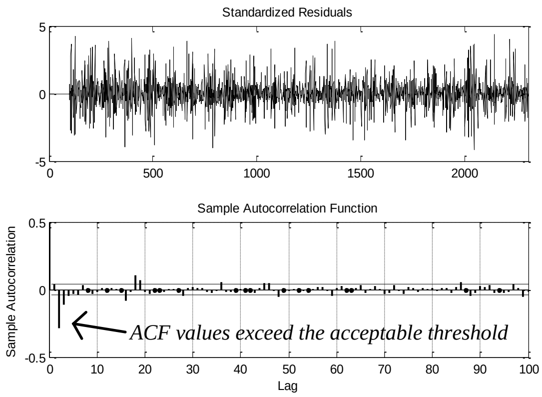

Fig. 8 Standardized residuals and ACF for the residuals of the tentative SARIMA model $(0,1,2)\times(0,1,1)_{96}$.

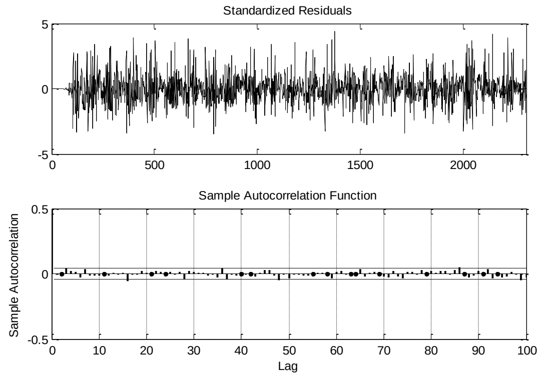

Fig. 9 Standardized residuals and ACF for the residuals of the tentative SARIMA model $(1,1,2)\times(0,1,1)_{96}$.

Table 2 Estimated Parameters for ARIMA $(1,1,2)\times(0,1,1)_{96}$

Parameter	Value
ϕ_1	0.405931
θ_1	-0.58512
θ_2	-0.30291
Θ_1	-0.797306

5. Evaluation of Forecast Models

In this first case study of load forecast, the MAPE is 19.27%. Although the MAPE is rather large, the prediction for the peak-load instant is fairly accurate as shown in Table 4. The ARIMA $(1,1,2)\times(0,1,1)_{96}$ (first case study) can successfully predicts the peak-load instants for all 5 workdays. And this can be useful for the peak-shaving application.

For the second case study, the prediction models can be obtained by using the same procedure of Box-Jenkins methodology. Five SARIMA models for each workday are given in Table 3. Fig. 11 shows the comparison between the actual load data and forecasted load data. The MAPE values are calculated and also listed in Table 3, the average of MAPE in this second case study is 18.82%. The MAPE of second case study is better than the first case study, this reveals that the each workday has the salient behavior of energy usage. However, the prediction of peak-load instant of this second case study is less accurate in comparison to the first case, this is probably due to the amount of history data is five-time less than that of the first case study.

By comparison with other linear models introduced in the literature, the research works in [3-5] present the hourly load forecasting and the values of MAPE are less than 10%. These hourly load models are seemingly suitable for some applications, nevertheless, these models should be reexamined for the shorter-term forecasting, e.g. peak-shaving application which requires the 15-minute forecast model.

6. Conclusion

This paper presents a short-term electricity load forecasting in building by using the SARIMA model. The SARIMA models are obtained by the Box-Jenkins methodology. With regarding to the systematic approach of Box-Jenkins methodology, the model can be constructed by considering the characteristics of historical load data. The validity of SARIMA models are evaluated with the actual load data and the accuracy of the models are acceptable for load forecasting in the buildings, especially the prediction of the peak-load instant. It can be expected that the proposed SARIMA models can be used to further development of various applications in BEMS, e.g. peak shaving, self-consumption and net-zero building.

Table 3 Forecast Models for Second Case Study.

Case	Model	MAPE (%)
Monday	ARIMA $(2,1,1)\times(0,1,1)_{96}$	17.71
Tuesday	ARIMA $(2,1,1)\times(0,1,1)_{96}$	16.81
Wednesday	ARIMA $(2,1,2)\times(0,1,1)_{96}$	19.17
Thursday	ARIMA $(2,1,1)\times(0,1,1)_{96}$	17.33
Friday	ARIMA $(2,1,2)\times(0,1,1)_{96}$	23.09
Average		18.82

Table 4 Prediction of Peak-load Instant.

Day	Actual	First Case	Second Case
Monday	14.15	14.30	14.30
Tuesday	14.45	14.30	14.45
Wednesday	14.30	14.45	15.00
Thursday	14.45	14.30	14.30
Friday	15.00	14.30	14.00

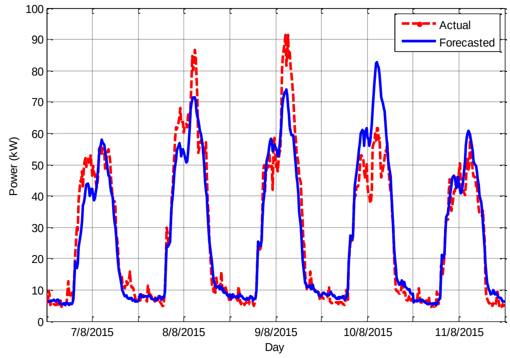

Fig. 10 Load forecasts given by ARIMA $(1,1,2)\times(0,1,1)_{96}$ for 5 workdays.

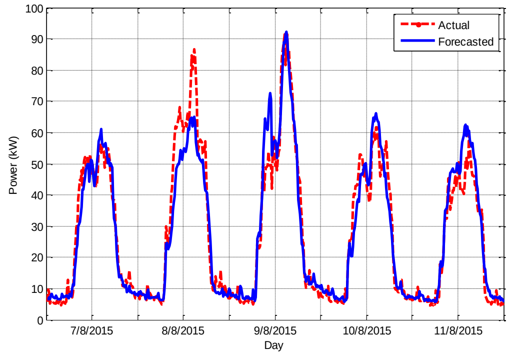

Fig. 11 Load forecasts given by SARIMA models in Table 3 for 5 workdays.

7. Acknowledgement

The financial support by Center of Excellence in Electrical Power Technology is gratefully acknowledged. The authors would like to also thank CU-BEMS project at the Department of Electrical Engineering for providing research data and facilities.

8. References

[1] Energy Forecast and Information Technology Center, “Thailand’s Energy Situation in 2014,” April 2015.

[2] L. Hernandez, C. Baladron, J.M. Aguiar, B. Carro, A.J. Sanchez-Esguevillas, J. Lloret, and J. Massana, “A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings,” IEEE Communications Surveys & Tutorials, vol. 16, no. 3, pp. 1460-1495, 2014.

[3] J. Lim, O. Kwon, K. Song, and J. Park, “Short-term load forecasting for educational buildings with temperature correlation,” in proc., 2013 International Conference on Power Engineering, Energy and Electrical Drives (POWERENG), May 2013, pp. 405-408, 2013.

[4] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, “Short-term load forecasting in a non-residential building contrasting models and attributes,” Energy and Buildings, vol. 92, pp. 322-330, 2015.

[5] Y.K. Penya, C.E. Borges, and I. Fernandez, “Short-term load forecasting in non-residential Buildings,” 2011 AFRICON, pp. 1-6, 2011.

[6] Y. Chae, R. Horesh, Y. Hwang, and Y. Lee, “Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings,” Energy and Buildings, vol. 111, pp. 184-194, 2016.

[7] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel, “Time Series Analysis: Forecasting and Control,” 4th ed: Wiley, 2008.

[8] B.L. Bowerman, and R.T. O’Connell, “Time Series and Forcasting: an Applied Approach,”: Duxbury Press, 1979.

[9] S.G. Makridakis, S.C. Wheelwright and R.J. Hyudman, “Forcasting: Methods and Applications,” 2nd ed: Wiley, 1983.