Jurnal Ilmiah Matematika
Vol. 8, No. 1, April 2021, pp. 33-39
ISSN 2774-3241
http://journal.uad.ac.id/index.php/Konvergensi

Modelling and predicting energy consumption in laboratory buildings using multiple linear regression

Nor Hafizah Hussin [a,1,*], Rahaini Mohd Said Aditya [a,2], Nadiah Ishak [a,3]

[a] Fakulti Teknologi Kejuruteraan Elektrik & Elektronik, Universiti Teknikal Malaysia Melaka
[*1]norhafizah.hussin@utem.edu.my
*Correspondent Author

KEYWORDS

Regression Model
Stepwise Selection
Influence factors
Energy consumption

ABSTRACT

This study was carried to improve the energy saving by investigating the influence factors that contribute to high energy consumption in a building particularly related to the building in Technology Campus, UTeM. Correlation analysis was performed to measure the strength of relationship between the influence factors whereby all the factors proven to have a strong linear correlation with the energy consumption. The Stepwise Selection of Multiple Linear Regression (MLR) were used to determine and modelling the most influence factors that affects the energy consumption. The final linear regression models was developed based on the amount of lighting in a building and surrounding temperature in the building which is considered as major influence factors that affect the energy consumption. Comparing the actual and predicted energy consumption in Technology Campus, UTeM showed that the MLR model obtained can be used to predict energy consumption and accounted for around 81% of the variance.

Introduction

Energy consumption in commercial buildings accounts for a significant proportion of worldwide energy consumption. In Malaysia, energy saving campaign has been very crucial due to fast economic development affected by globalization has steadily increasing the energy consumption. Inefficient use of energy or electricity contributes substantially to climate change. In this case, we need to adopt an energy saving equipment, maintenance and operations practices so that the inefficient consumption of energy can be reduces. Hence, it is important to determine the major contributor to the energy consumptions.

Study carried by few researchers have shown that there are many influence factors that contribute to the usage of energy. A research carried by Yuan[1] using a correlation analysis, it can be concluded that there were five factors that contribute to the energy consumption which is age of construction, number of building stories, the occupancy density, and the form of cooling source and heating source. A study conducted by Mora[2] to investigate the behavioral and physical factors influencing energy building performances reveal that floor area and climate are the most significant influence factor for electricity consumption. Another factors affecting the energy consumptions as proven in the research conducted by Chen[3] are equipment used and lighting system in the building. In this study, Technology Campus, UTeM will be selected as the area of research since it has few buildings that has a high demand of electricity and energy consumption.

The objective of this study is to determine and modelling the most influence factors that affect the energy consumption in Technology Campus, UTeM. Few factors have been studied and will be used to analyse the relationship between the factors and the energy consumption. Several researches have been carried out to assess the influencing factors of building energy consumption using multiple regression analysis [4][5][6]. Multiple linear regression models the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data[7] [8] [9].

A study conducted by Amiri[10] on using multiple regression analysis to develop energy consumption indicators for commercial buildings in the U.S. used the MLR Stepwise Selection to correlate the annual energy consumption with the 17 input parameters. Two design parameters, occupancy schedule and exterior wall construction, were found have the highest influence on both cooling and heating load since it has high value of R2 and the annual energy consumption will be more sensitive to the changes in these two design variables.

Using MLR Forward selection, Mottahedi[11] conducted a research on Multi-linear Regression Models to Predict the Annual Energy Consumption of an Office Building with Different Shapes conclude that energy models shows that there is a strong interaction between building shapes, their locations, and level of energy consumptions. According to a research on the development of multi-linear regression analysis to assess energy consumption in the early stages of building design by Asadi[12], the developed model consist of variables such as building materials, their thickness, building shape, and occupant schedule are selected using the minimum Mean Square Error (MSE) selection. This study is motivated to offer a flexible and simple model in order to facilitate the development of regression models to predict the energy consumption for energy saving in Technology Campus, UTeM.

Metode

Data Description

In this study, data was collected in Technology Campus, UTeM which consist of several building that includes a laboratory buildings and administrative building. It is known that the energy consumption in the laboratory building are higher compared to administrative building during the semester start, hence this project was focusing on auditing the laboratory buildings for improving the energy saving campaign. The data were taken three times a day and the daily energy consumption were used as the response variable, *y*. This study has narrow down few factors that might affect the energy usage in laboratory buildings such as the number of lighting used, the temperature of the surrounding especially the usage of air-conditioning, and also the occupancy in that particular building. All these factors are used in data collection as the explanatory variable, *x,* to determine and modelling the most influence factors that affect the energy consumption in Technology Campus, UTeM. In this research, response variable is the energy consumption (*Y)* while temperature (X_1*)*, number of lighting (X_2), and occupancy (X_3) were explanatory variables.

In this study, the methods that will be used for correlation analysis is Pearson's correlation coefficient method where it is used to investigate the relationship between the energy consumption with each of the influence factors that affect the energy consumption using the Pearson's Correlation Coefficient. Followed by modelling part using Stepwise selection MLR where the final outcome will be a best fitted model that can be used to predict the energy consumption.

Correlation Analysis

The correlation analysis is a method to develop relationship between the response variable and explanatory variable. Before performing a modelling, we might need to check the linearity where there must be a linear relationship between the response variable and the explanatory variable in order to fit a linear model. Using the Pearson's correlation coefficient method, the value r can be calculated using the following formula;

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}} \tag{1}$$

In this research, the value of r will be used to check whether there is a relationship between the energy consumption and each of the influence factors which is number of lighting used, surrounding temperature in the building and number of occupancies. It can also measure the strength of relationship and according to Hussin [13], some of the strength value r obtained by equation (1) are as follows Table 1:

Table 1. Pearson Correlation Coefficient strength

Value of r	***Correlation strength***
0	No correlation
$0.1 \le r < 0.3$	Weak
$0.3 \le r < 0.7$	Moderate
$0.7 \le r < 1$	Strong
1	Perfect

Modelling using Multiple Linear Regression

Multiple Linear Regression (MLR) are widely used in explaining the dependency of response variable with several explanatory variables. It also can be used to forecast effects or impact changes and predict trends and future values. In this study, the multiple linear regression was used to develop a mathematical modelling that can be used to predict future energy consumption. Since there are few factors that might affect the energy consumption, stepwise regression will be used to make sure that only the most influenced factor will be selected in the model. Throughout the process of combination between forward and backward selection, only the variable that were significant at p =0.05 will be maintained in the final model[14]. The MLR model that can be used in predicting the energy consumption are;

$$\hat{y} = \phi_0 - \phi_1\beta_1 + \phi_2\beta_2 + \phi_3\beta_3 + \varepsilon$$

(2)

Where $\hat{y}$ is the response variable which is the energy consumption and ϕ_1, ϕ_2, ϕ_3 are the regression coefficient for explanatory variables while $\beta_1, \beta_2, \beta_3$ are the explanatory variable which is the number of lighting used, the temperature of the surrounding, and the occupancy in that particular building.

Stepwise regression is a technique that are widely used in eliminating the variable that is not related to the response variable. The final model will only contain the explanatory variable that are significantly related to the response variable. Hence, there are two important values that need to be check to make sure the stability and reliability of the final model obtained which is p-value and Variance Inflation Factor (VIF). This is to make sure that the model can be used in predicting the future energy consumption and the value are reliable.

The p-value is known to determine the relationship that is obtained in the sample also exist in the larger population. The p-value for each explanatory variable test the null hypothesis that the variable has no correlation with the response variable. When there is no correlation, there will be no association between changes in the explanatory variable and the shift in the response variable. We can also conclude that there is no sufficient evidence to prove that there is an effect at the population level. If the *p*-value for a variable is less than the significance level, the sample data provide enough evidence to reject the null hypothesis for entire population. Meaning that changes in the explanatory variable will also affect the changes in the response at the population level. This shows that the variable is statistically significant to be added in the regression model.

In modelling using a multiple regression as shown in equation (1), there might also chances of having two or more explanatory variables that are related to each other. This is called multicollinearity. It effects the coefficient estimates which makes it difficult to interpret the result and unstable. It can cause the coefficient to switch signs, making it more difficult to identify and determine the perfect model. In order to estimate multicollinearity, the variance inflation factor (VIF) can be used to assess the variance of estimated regression coefficient were increased when the explanatory variable is correlated. If the value of VIF is equal to 1, there are no multicollinearity among the explanatory variable while if the value of VIF is greater than 1 and less than 5, it shows that the variable might be moderately correlated but is acceptable. For a value of VIF between 5 to 10, it shows a high correlation that might cause a problem on the model and if it is more than 10, it can be considered as poor estimation of regression coefficient due to multicollinearity[15].

Once the MLR model were obtained, it is also important to determine how well the model fits the data. This can be done using the one of the goodness-of-fit tests which is coefficient of determination, R^2. It shows how close the data are fitted in the regression line. The range of R^2 will be always between 0 to 100% where 0% means that the model explains no variability of the response data and 100% indicates that the model explains all the variability of the response variable. This concludes that the higher the value of R^2, the better the model fits the data. Once all the model checking and validation have been done, the model will be appropriate to be used in predicting the energy consumption for energy saving in Technology campus, UTeM.

Results and Discussion

The result of this study was based on the data collected every day in a laboratory building in Technology Campus, UTeM for few months including weekend. The daily energy consumption was measured using the kilowatt meter reading (kWh). Based on the data of three influenced factors that affect the energy consumption which is surrounding building temperature, number of lighting used in the building, and number of occupancies in the building, the results for the research was divided into two parts.

Correlation Analysis

Table 2 below shows the result for correlation analysis where the correlation between the response variable and the explanatory variable were investigated and the strength were measured using Pearson's correlation coefficient.

Table 2. Pearson Correlation Coefficient strength

Influence Factor	*Pearson's Correlation, r*	*P-value*
Surrounding Temperature	-0.821	0.000
Number of Lighting	0.877	0.000
Occupancy	0.870	0.000

Based on the result shown in Table 1, each of the influence factors has a linear correlation with the energy consumption. The P-value obtained for each factor proven that there is sufficient evidence to conclude that there is a significant linear relationship between surrounding temperature, number of lighting, and occupancy with the energy consumption

 ISSN 2774-3241

because the *p*-value is less than the significance level (α=0.05). From this table, it shows that there is a strong correlation between the energy consumption and surrounding building temperature (-0.821), number of lighting used in the building (0.877), and number of occupancies in the building (0.87).

Modelling using Multiple Linear Regression

Using the stepwise selection for modelling using MLR, the results shows that the energy consumption in Technology Campus, UTeM is dependent with the number of lighting used in the building and the temperature in the building. These two factors were determined using the Stepwise regression as the most influenced factors that affect the energy consumption. The result is shown in Table 3 below.

Table 3. Coefficient, p – value, and VIF for each term in the Model

Term	*Coefficient*	*P-Value*	*VIF*
Constant	149.3	0.000	
Surrounding Temperature	-3.95	0.001	2.71
Number of Lighting	0.0877	0.000	2.71

The equation for the most influenced factor affecting the energy consumption in Technology Campus, UTeM is:

$$\hat{y} = 149.3 - 3.95\beta_1 + 0.0877\beta_2 + \varepsilon \tag{3}$$

In other word, the energy consumption can be predicted using the model = 149.3 – (3.95*surrounding temperature) + (0.0877*number of lighting) as shown in equation (2). Based on the p-value obtained, both explanatory variable which is surrounding temperature and number of lighting used in the building are statistically significant because the p-value is less than the significance level (α=0.05). This shows that there is enough evidence to conclude that these two factors are a worthwhile addition to the regression model. While the value of VIF for the two variables are both 2.71 which is greater than 1 and less than 5, it shows that the variable might be moderately correlated but is acceptable to be included in the model. There is no multicollinearity among the explanatory variable, hence, the variable included in the model are reliable to be used in predicting the energy consumption in Technology Campus, UTeM. The result is shown in Figure 1 and Figure 2 below.

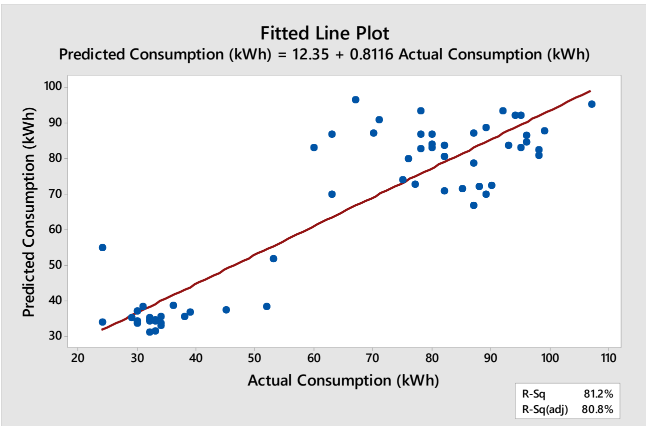

Figure 1. Relationship between actual and predicted energy consumption based on the MLR model

ISSN 2774-3241

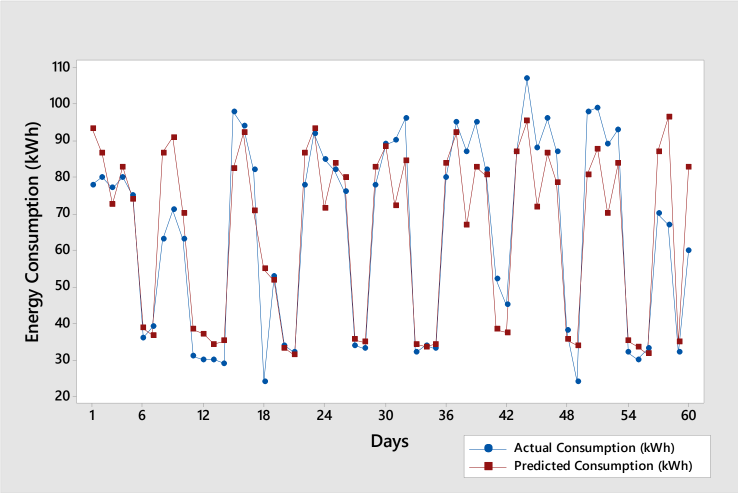

Figure 2. Actual and predicted energy consumption for the building in Technology Campus, UTeM

The value of R^2 obtained as shown in Figure. 1 is 0.8116 means that around 81.2% of the total variation in kilowatt meter reading or the energy consumption can be explained by the relationship with the surrounding temperature and the number of lighting used in the building. This means that these two factors are highly affecting the energy consumption while the other factor which is the number of occupancies in the building might be affecting the energy consumption but in a smaller scale. The results also show that the final MLR model could be fitted to energy consumption data and estimate around 81.2% of the variance. The result in Figure. 2 shows the actual and predicted data matched together for the energy consumption in Technology Campus, UTeM. Based on the result, it appear that some difference might be in the peak energy consumption. Most of the predicted value are closely in the range of actual energy consumption. It can be concluded that the predicted and actual data were significantly correlated and the linear model obtained can be used to predict energy consumption with an acceptable error for energy saving in Technology Campus, UTeM.

Conclusions

This study presented a mathematical modelling using Multiple Linear Regression (MLR) to predict the energy consumption for energy savings in Technology Campus, UTeM. The correlation analysis proved that there is a significant linear relationship between surrounding temperature, number of lighting, and occupancy with the energy consumption because the *p*-value is less than the significance level (α=0.05). Using the stepwise selection, the linear model obtained consist of the most influenced factors which is the number of lighting used and surrounding temperature in the building where the equation is $\hat{y} = 149.3 - 3.95\beta_1 + 0.0877\beta_2 + \varepsilon$. This model can be used to predict the energy consumption and accounted for 81% of the variance which is within the acceptable error. In future work, the finding of this study can be compared with other modelling methods by measuring the level of accuracy for each models.

Acknowledgement

The authors would like to acknowledge Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik (FTKEE), Universiti Teknikal Malaysia Melaka (UTeM) and those who give energetic and full support in carrying this research under the grant vote number PJP/2018/FTK (15A)/S01641.

References

[1] L. Yuan, Y. Ruan, G. Yang, F. Feng, and Z. Li, "Analysis of factors influencing the energy consumption of government office buildings in Qingdao," *Energy Procedia*, vol. 104, pp. 263–268, 2016.

[2] D. Mora, C. Carpino, and M. De Simone, "Behavioral and physical factors influencing energy building performances in Mediterranean climate," *Energy Procedia*, vol. 78, pp. 603–608, 2015.

[3] Y.-T. Chen, "The factors affecting electricity consumption and the consumption characteristics in the residential sector—a case example of Taiwan," *Sustainability*, vol. 9, no. 8, p. 1484, 2017.

[4] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices," *J. Financ. Data Sci.*, vol. 4, no. 3, pp. 183–201, 2018, doi: 10.1016/j.jfds.2018.04.003.

[5] M. C. Tsai, C. H. Cheng, M. I. Tsai, and H. Y. Shiu, "Forecasting leading industry stock prices based on a hybrid time-series forecast model," *PLoS One*, vol. 13, no. 12, pp. 1–24, 2018, doi: 10.1371/journal.pone.0209922.

[6] P. E. Ong, A. K. C. Huong, X. T. I. Ngu, F. Mahmud, and S. P. Philimon, "Modified lambert beer for bilirubin concentration and blood oxygen saturation prediction," *Int. J. Adv. Intell. Informatics*, vol. 5, no. 2, pp. 113–122, 2019, doi: 10.26555/ijain.v5i2.363.

[7] M. R. Braun, H. Altan, and S. B. M. Beck, "Using regression analysis to predict the future energy consumption of a supermarket in the UK," *Appl. Energy*, vol. 130, pp. 305–313, 2014.

[8] R. Gamberini, F. Lolli, B. Rimini, and F. Sgarbossa, "Forecasting of sporadic demand patterns with seasonality and trend components: An empirical comparison between holt-winters and (s)ARIMA methods," *Math. Probl. Eng.*, vol. 2010, 2010, doi: 10.1155/2010/579010.

[9] S. M. Chen and K. Tanuwijaya, "Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10594–10605, 2011, doi: 10.1016/j.eswa.2011.02.098.

[10] S. S. Amiri, M. Mottahedi, and S. Asadi, "Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the US," *Energy Build.*, vol. 109, pp. 209–216, 2015.

[11] M. Mottahedi, A. Mohammadpour, S. S. Amiri, D. Riley, and S. Asadi, "Multi-linear regression models to predict the annual energy consumption of an office building with different shapes," *Procedia Eng.*, vol. 118, pp. 622–629, 2015.

[12] S. Asadi, S. S. Amiri, and M. Mottahedi, "On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design," *Energy Build.*, vol. 85, pp. 246–255, 2014.

[13] G. W. Snedecor and W. G. Cochran, "Statistical methods.," *Stat. methods.*, no. 6th ed.(repr.), 1968.

[14] M. Safa, J. Allen, and M. Safa, "Predicting energy usage using historical data and linear models," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, 2014, vol. 31, p. 1.

[15] M. O. Akinwande, H. G. Dikko, and A. Samson, "Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis," *Open J. Stat.*, vol. 5, no. 07, p. 754, 2015.