Multivariate linear regression model for estimating energy consumption

Elaheh Jalilzadehazhari[1], Krushna Mahaparta[2]

[1]Department of Forestry and Wood Technology, Linnaeus University, Växjö 352 52, Sweden
[2]Department of Built Environment and Energy Technology, Linnaeus University, Växjö 352 52, Sweden

ABSTRACT

Windows as essential elements of buildings have a significant effect on the energy consumption, including heating and cooling demand in Sweden. A reliable statistical model for estimating the energy consumption associated with various window designs and glazing systems helps architects and designers in the early stage of the design process. Most of the introduced models in literature utilized a mathematic sampling algorithm such as Monte Carlo to develop a simple linear regression model for estimating the energy consumption. A simple linear regression model cannot describe the effect of different groups of a categorical variable. Hence this study considers four variables related to the window characteristics, including window size, design model, orientation, glazing system and develops a categorical multiple linear regression model for estimating the energy consumption. 544 simulations were performed by COMFEN Beta5 software. The results were used as a database for developing categorical multivariate linear regression models. The accuracy of the developed model was studied by the coefficient of determination, R-square value (R^2). The obtained R^2 exceeded by 94%. Furthermore, the predicted energy consumptions obtained by the developed regression model were compared with the simulated values by COMFEN software. Results show a strong linear relationship between predicted and simulated values. Developed multivariate linear regression model can be utilized in the early stage of the design process for estimating the energy consumption associated with various window designs and glazing systems.

KEYWORDS: Multiple linear regression, categorical variable, interaction analysis

INTRODUCTION

Designing energy efficient buildings in Sweden is considerably important since the total energy consumption should be reduced by 20% and 50% until 2020 and 2050 respectively (Energy Indicators [In Swedish: Energiindikator], 2016). In this context, windows play an important role. Because a large fraction of transmission heat loss occurs through the windows (Avasoo, 2007). Hence analyzing various window design variables including window size (Kim et.al, 2014), position (Acosta et.al, 2015), orientation (Tavares at.al, 2014), glazing and shading systems (Abdul Fasi and Budaiwi, 2015; Tzempelikos and Shen, 2013) by a simulation tools gained interest among scientific studies. The available simulation tools assist designers in simulating the complex thermal interactions between indoor and outdoor environment and calculating the energy consumption. The need for using simulation tools increases as buildings become more complex. However, performing simulations rely on designers and engineers experienced in the subject. Furthermore, simulating energy consumption for numerous combinations of window design variables requires a significant time and bring the huge amount of results (Hygh et.al, 2012). Hence there is a need for a reliable model which helps to estimate the energy consumption in the early stage of design process.

Many studies have attempted to develop linear regression models by implementing a mathematic algorithm such as Monte Carlo. Asadi et.al (2012) developed a multivariate linear regression model for estimating energy consumption by considering 17 different variables including orientation and glazing system. They performed ten thousand simulations for seven different buildings by implementing Monte Carlo algorithm. Hygh et.al, (2012) developed a multivariate linear regression model based on 27 variables and utilized Monte Carlo algorithm for executing 20000 simulations. They analyzed the effect of various window sizes, glazing systems, orientation and shading projection factor (% of window height) to quantify the energy performance of a building by considering its size, location, and geometry. In another study carried out by Westphal and Lamberts (2007), multiple linear regression models were developed for estimating the electric energy consumption for an air condition system in three cities in Brazil. Scripts routines for executing the simulations automatically were written in BASIC language. The aforementioned studies show that multivariate linear regression model can serve as an effective model for assessing energy consumption in the early stage of design process. However, utilizing mathematic sampling algorithms in these studies requires programming skills. On the other hand, these studies considered variables such as orientation as a continuous variable and introduced only one coefficient value. As the developed linear regression models cannot describe the effect of different groups of a categorical variable. Hence this study attempted to develop a linear regression model by considering categorical variables as factor variables and without implementing a sampling model. This model introduces different coefficient values for various groups of categorical variables and shows their effect on the energy consumption.

METHODOLOGY

A 9 m² office room prototype was modeled using the COMFEN 5BETA software (COMFEN). This software simulates a range of window designs and allows to analyze the performance of a building in term of energy consumption. The simulation conditions, based on common construction practice in Sweden, and climate data of the office room are presented in Table 1 and 2.

Table 1.
Simulation condition

Building type	Office
Location	Gothenburg
Dimension (W.D.H)	3m x 3m x 2.7m
Exterior wall u-value	0,19 W/K.m²
HVAC	Packaged single zone
Set point	Heating: 18˚C Cooling: 24˚ C
Equipment load	8 W/m²
Lighting load	10 W/m²

Table 2.
Climate data, daylight hours only

	Gothenburg
Latitude	57.7
Longitude	11.97
Annual average direct beam	7125.9
Annual average global horizontal diffuse	13646.6
Annual average clearness index	1.47

2.1. Design variables: Windows size, orientation, and position

To perform the simulations, four groups of design models have been specified (Fig. 1). Which differ from window size (window wall ratio: WWR) and window position. Simulations for calculating the energy performance of each design group were first executed for minimum WWR as 10%. Similar simulations were replicated over again while WWR was increased by 5% up to 90%. Moreover, energy performance of design groups was calculated for four cardinal orientations (north, south, east, and west). However, in performing the simulations, it was assumed that the office room was located in the intermediary level. This means that the façade with window was considered as an external wall, while the other walls, roof, and floor were counted as adiabatic.

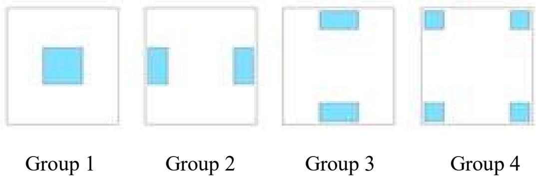

Fig 1. Schematic illustration of design models

2.2. Design variable: Glazing system

Non-operable Elit original window: this window consists of three glasses with 4mm thickness while the interior glass is exfoliated with a low e-coating layer. There are two 16 mm air gaps between each pair of glass. The interior air gap is filled with 10% air and 90% argon. However, the exterior one is filled with only air (Elitfönster, 2016). In order to prevent glare problem, this window in all design groups was modeled with a 1-inch interior Venetian blind. As the blind was completely lowered at the present of glare. Table 2, shows the thermal properties of utilized Venetian blind for each design group.

Table 3.
IR thermal properties of interior Venetian blind

IR properties	IR. Transmittance	0
	IR. Emission, back	0.9
	IR. Emission, front	0.9

Electrochromic (EC) SageGlass window: This window consists of two glasses with 6mm thickness. The exterior glass was laminated with a lithium-based EC film. The EC glass is then fabricated as insulated glass unit (IGU) by combining with a single glass pane. The 12.7 mm air gap was filled with 90% argon and 10% air (Tavares et.al, 2014). The EC SageGlass was equipped with a glare control system, as this window was tinted at the present of unappreciated glare.

Table 4 shows the visible and solar transmittance, u-value and solar heat gain coefficient (SHGC) of Elit Original and EC SageGlass. In clear state, the SHGC and visible transmittance of EC SagaGlass and Elit original are relatively similar.

Table 4.
Window properties

		Transmittance		U-Factor	SHGC
		Visible	Solar		
SagaGlass properties (Sbar et.al, 2012)	Clear	63	38	1,3	0.47
	Intermediate 1	21	9		0.17
	Intermediate 2	6	2		0.11
	Fully tinted	2	0,7		0.09
Non-operable Elit Original	-	66		1	0.48

2.3. Simulations

544 simulations were performed by considering combinations of four design groups, 17 window size, four cardinal orientations and two glazing systems. However, the other properties of the office room, mentioned in Table 1 were kept constant while simulations were performed. As a default, COMFEN software considers an occupancy schedule for performing simulations (Fig 2).

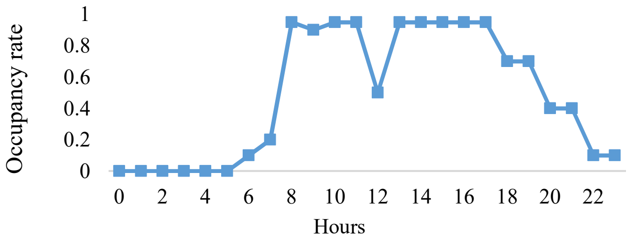

Fig. 2, occupancy schedule of simulation

2.3. Multivariate linear regression model

A multivariate linear regression model fits a linear equation between various independent variables and one dependent variable by considering the relationships between them (Seltman HJ, 2015). According to Berrington de Gonz´alez, and DR. Cox (2007), studying the interactions among the variables and considering them into the model helps to understand the relationships and interpret the model. Interactions occur when the effect of one independent variable on dependent variable depends on another independent variable. This interaction, which in known as two-way interaction can be identified by performing two-way ANOVA test. A multivariate linear regression model, developed by considering the interaction, can be written as;

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{ij} X_i X_j + \cdots + \beta_p X_p + \varepsilon \qquad (1)$$

Similar interactions can occur between one independent variable and a group of independent variables. For example, a three- way interaction in multivariate linear regression model with three- way interaction can be shown as;

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{ijk} X_i X_j X_k + \cdots + \beta_p X_p + \varepsilon \qquad (2)$$

Moreover, a multivariate linear regression model can be developed by including factor variables. Considering a categorical independent variable as factor variable means that, various dummies will be created for different groups of a categorical variable. By treating a categorical variable as a factor variable, each group of a categorical variable will obtain a separate coefficient value. For instance, if the X_1 formula (1) is a categorical variable with two groups, then the multivariate linear regression model can be rewritten as;

$$Y_{group\ 1} = \alpha + \beta_{11} X'_1 + \beta_2 X_2 + \cdots + \beta_{ij} X_i X_j + \cdots + \beta_p X_p + \varepsilon \qquad (3)$$
$$Y_{group\ 2} = \alpha + \beta_{12} X'_2 + \beta_2 X_2 + \cdots + \beta_{ij} X_i X_j + \cdots + \beta_p X_p + \varepsilon \qquad (4)$$

In this study, Stata software has been utilized for developing a multivariate linear regression model (Stata). This software allows treating the categorical variables as factor variables and introduces different coefficient values for variables' respective groups. For the purpose of achieving a model with the best fit, various multivariate linear regression models were developed. First, the orientation, design group, and glazing system were considered as categorical variables one by one. Later two- pair combinations were tested. Finally, the mentioned variables were considered as categorical variables all at once. However, while developing various models, the window size was always considered as a continuous variable.

The best fit model was developed first based on 90% of the dataset, selected randomly. The remaining 10% of the dataset was later used for validating the developed model. The validity of the model was further checked with a second multivariate linear regression model, based on 50% of the dataset. However, it should be mentioned that the interaction between variables was analyzed and considered in developing both models, based on 90% and 50% of the dataset. Furthermore, the accuracy of multivariate linear regression models was studied by calculating the coefficient of determination R- square value (R^2), F-value and root MSE.

RESULTS

For the purpose of detecting interactions among variables, the two-way ANOVA test was performed as significance level (P) was less than 0.05 and confidence interval as 95%. Larger F value due to larger variance between groups leads to significant results. As seen in Table 5, the interactions between window size- orientation and glazing- orientation are significant. This means that the effects of window size and glazing on energy consumption varies with orientation. This effect is known as three-way interaction. The three- way interaction can be analyzed by three-way ANOVA test. As seen in Table 6, the three-way interaction between window size- orientation- the glazing system is significant.

Table 5
Two- way ANONA test
Dependent variable is energy consumption

Integration between independent variables	F	Sig
Window size- Glazing systems	0.413	0.979
Window size- Orientation	3.538	0.000
Window size- Design models	0.923	0.621
Glazing systems- Design models	0.005	1.000
Glazing systems- Orientation	4.205	0.006
Design models- Orientation	0.748	0.665

Table 6
Three- way ANONA test
Dependent variable is energy consumption

Integration between independent variables	F	Sig
Window size- Glazing systems- Orientation	3.185	0.000
Window size- Glazing system- Design models	0.031	1
Window size- Orientation- Design models	0.364	1
Glazing system- Design models- Orientation	0.113	0.999

3.1. Multivariate linear regression model based on 90% of dataset

Analyzing the 90% of dataset shows that distribution of energy consumption is not normal. For this purpose, the natural logarithm of energy consumption was defined as the skewness was close to 0;

$$\text{Ln- energy consumption} = \text{Ln (energy consumption- K)}$$

Which happens at K= -5.3. This means that, in developing multivariate linear regression model, "Ln- energy consumption" was defined as;

$$\text{Ln- energy consumption} = \text{Ln (energy consumption+ 5.3)}$$

Various multivariate linear regression models were developed by considering two- way interactions and three- way interaction singly. Later both of interactions were included in the models. However, after analyzing different models, it was recognized that the best fit could be achieved by considering window size, glazing system and design models as continuous variables and orientation as a categorical variable.

Table 7 shows the developed multivariate linear regression model for estimating energy consumption. As seen in this table, it was assumed that windows were oriented to the north in the base case. The R^2 value of the multivariate linear regression model was exceeded to 94%, and the Root MSE was not significant and indicates that the developed model has a good fit to predict the energy consumption.

Table 7.
Developed multivariate linear regression

Ln- energy consumption	coefficient	Std. Err	t	P> | t |	95% conf. Interval	
Window size	0.015	0.000	83,98	0.000	0.014	0.015
Glazing system	0.038	0.009	4.38	0.000	0.209	0.055
Design model	-0.004	0.004	-3.47	0.001	-0.021	-0.006
Orientation						
North	0					
South	-0.129	0.012	-10.49	0.000	-0.153	-0.104
East	-0.068	0.012	-5.48	0.000	-0.091	-0.043
West	-0.059	0.012	-4.88	0.000	-0.084	-0.035
Cons	2.802	0.204	137.03	0.000	2.762	2.843

R^2	0.94
F (6, 489)	1210.59
Root MSE	0.09

In developed multivariate linear regression model, predicted "Ln- energy consumption" for the base case is;

Ln- energy consumption= 0.015 Window size + 0.038 Glazing system – 0.004 Design model + 2.802

Comparing the absolute coefficient of variables shows that orientation has a significantly higher effect on energy consumption than the window size, glazing system, and design models. In the meantime, analyzing the coefficients of four groups of orientation shows that, a window in south façade decreases the energy consumption more than other facades. Moreover, the negative coefficient of design model indicates that by a one-unit increase in the design model, Ln- energy consumption is decreased around – 0.004. Since the group 1,2,3 and 4 of design model were coded as 1,2,3 and 4 respectively, the developed multivariate linear regression model shows that group 1 has the lowest performance considering energy consumption.

3.1.2 Validation of regression model based on 90% of data set

According to Asadi et.al (2014), calculating the R^2 value does not meet requirements considering the validity of a multivariate linear regression model. For this purpose, the predicted value of 10% of the dataset by the developed model was compared with calculated energy consumption by COMFEN software. As seen in fig. 3, there is a strong correlation between predicted and calculated energy consumption, as R^2 was exceeded to 95%. Hence the developed multivariate linear regression model can effectively be utilized in early stages of design process.

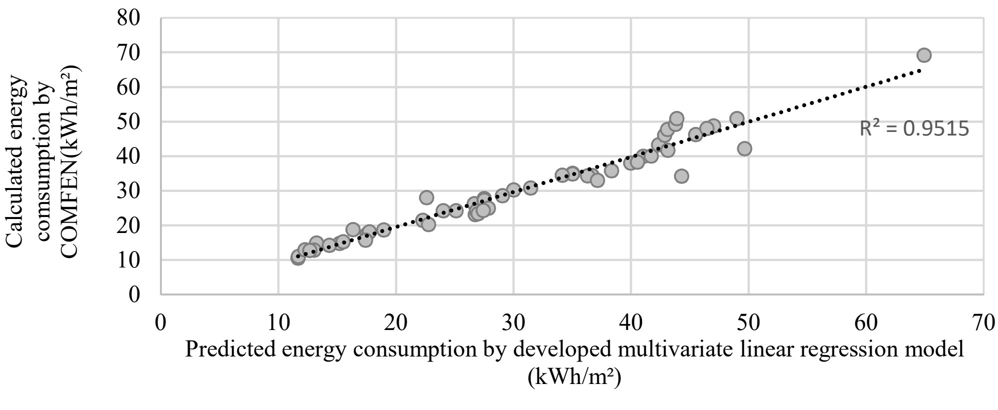

Fig 3. Validation of the multivariate linear regression model

3.1.3 Validation of regression model based on 50% of data set

For the purpose of revalidating developed model above, a similar multivariate linear regression model was developed based on 50% of the dataset, while the remaining 50% of the dataset was used for validating the model. The skewness of the energy consumption was close to 0 while K= -3.56.

$$Ln\_\ energy\ consumption = Ln\ (energy\ consumption + 3.56)$$

The coefficient of variables and their significance in the second multivariate linear regression model were similar to first one. The R^2 value remained constant as 95% however the Root MSE was increased to 0.10.

Fig. 4 shows the correlation between predicted values of remaining 50% of the dataset by second developed model and simulated energy consumption by COMFEN. As seen in this figure, the second model can produce reasonable results in about 91% of cases.

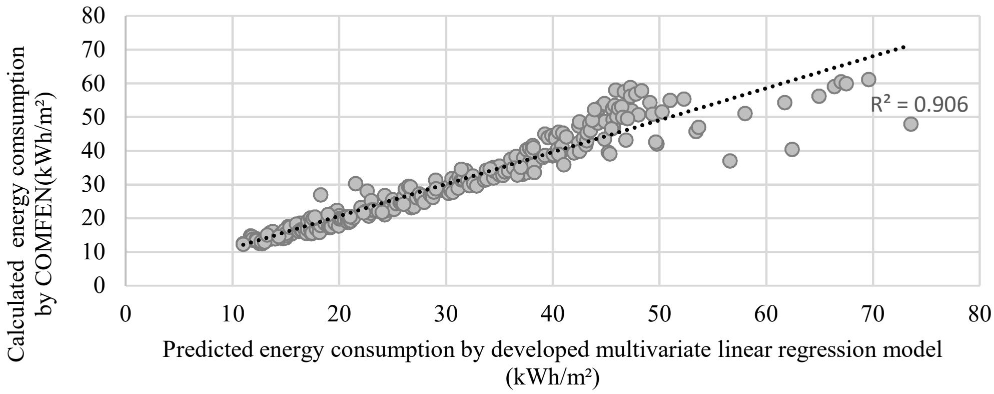

Fig 4. Validation of the multivariate linear regression model

CONCLUSION

544 simulations were performed by COMFEN software for creating a dataset used in developing a multivariate linear regression model. The developed model, which was based on 90% of the dataset, aimed to predict the energy consumption of an office room. The energy consumption in the developed model was dependent on various window sizes, orientations, design models and glazing systems. The R^2 value of the developed model exceeded to 94% which shows an excellent fit. Analyzing the absolute coefficient of the variables shows that orientation has a significantly larger effect on energy consumption than other variables.

The validity of the model was retested by remaining 10% of the dataset. The strong correlation between predicted values by model and simulated energy consumption by COMFEN software show the utility of the model in predicting energy consumption in the early stage of design process. Moreover, the developed model was revalidated by the second multivariate linear regression model, which was developed based on 50% of the dataset. However, the correlation between predicted values by the second model and simulated energy consumption by COMFEN software (R^2 value) were decreased to 91%, which shows the strength of the first model over the second one.

Despite the utility of COMFEN in performing simulations, this software can calculate the energy consumption of the four main cardinal orientations. This limitation restricts the utilization the developed model in predicting energy consumption. However, in future, the multivariate linear regression model can be developed further by considering other common orientations.

ACKNOWLEDGMENT

The authors would like to thank KK- stiftelse Sweden for supporting this research and Inwido Sweden for providing the required information.

REFERENCES

Abdul Fasi M, Budaiwi IM. 2015. Energy performance of windows in office buildings considering daylight integration and visual comfort in hot climates. Energy and Buildings. Vol 108, p 307-316

Acosta I, Munoz C, Campano MA, Navarro J. 2015. Analysis of daylight factor and energy saving allowed by windows under overcast sky conditions. Renewable Energy. Vol 77, p 194-207

Asadi S, Shams Amiri S, Mottahedi M. 2014. On the development of multi- linear regression analysis to assess energy consumption in the early stages of building design. Energy and Buildings. Vol 85, p 246-25

Avasoo D. 2007. The European window energy labeling challenge. WSP Environmental group

Berrington de Gonz´alez A, Cox DR. 2007. Interpretation of interaction: A review. The Annual of applied statics. Vol 1(2), p 371- 385

COMFEN. www. windows.lbl.gov/software/comfen/5. Last accessed on 10 June 2016

Elitfönster, www.elitfonster.se, last accessed on 15 June 2016

Energy Indicators, Follow up by Swedish Energy Policy Goals [In Swedish: Energiidikatorer, Uppföljning av Sveriges energipolitiska mål]. 2016. Energy Association of Sweden [In Swedish: Energimyndigheten]

Hygh JS, DeCarolis JF, Hill DB, Ranjithan SR. 2012. Multivariate regression as an energy assessment tool in early building design. Building and Environment. Vol 57, p 165-175

Kim SH, Kim SS, Kim KW, Cho YH. 2014. A study on the proposes of energy analysis indicator by the window elements of office buildings in Korea. Energy and Buildings. Vol 73, p 153-165

Seltman HJ. 2012. Experimental design and analysis. Chapter 9: Simple linear regression. Carnegie Mellon University. P 213-240

Tavares PF, Gasper AR, Martin AG, Frontini F. 2014. Evaluation of electrochromic windows impact in the energy performance of buildings in Mediterranean climates. Energy Policy. Vol 67, p 68-81

Tzempelikos A, Shen H. 2013. Comparative control strategies for roller shades with respect to daylighting and energy performance. Building and Environment. Vol 67, p 179-192

Sbar NL, Podbelski L, Yang HM, Pease B. 2012. Electrochromic dynamic windows for office buildings. International journal of sustainable built environment. Vol 1, p 125-139

Stata. www.stata.com. Last accessed on 10 June 2016

Westphal FS, Lamberts R. 2007. Regression analysis of electric energy consumption of commercial buildings in Brazil. Proceeding Building Simulation. p 1543-1550