Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 118 (2015) 1258 – 1265

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Sustainable Design, Engineering and Construction

Determining the Feasibility of Statistical Techniques to Identify the Most Important Input Parameters of Building Energy Models

***R. Arababadi*[a], *H. Naganathan*[a]*, *K. Parrish*[a], *W.K. Chong*[a]**

[a]*Arizona State University, University Drive & Mill Avenue, Tempe, AZ 85281, United States*

Abstract

Previous studies in building energy assessment clearly state that to meet sustainable energy goals, existing buildings, as well as new buildings, will need to improve their energy efficiency. Thus, meeting energy goals relies on retrofitting existing buildings. Most building energy models are bottom-up engineering models, meaning these models calculate energy demand of individual buildings through their physical properties and energy use for specific end uses (e.g., lighting, appliances, and water heating). Researchers then scale up these model results to represent the building stock of the region studied.

Studies reveal that there is a lack of information about the building stock and associated modeling tools and this lack of knowledge affects the assessment of building energy efficiency strategies. Literature suggests that the level of complexity of energy models needs to be limited. Accuracy of these energy models can be elevated by reducing the input parameters, alleviating the need for users to make many assumptions about building construction and occupancy, among other factors. To mitigate the need for assumptions and the resulting model inaccuracies, the authors argue buildings should be described in a regional stock model with a restricted number of input parameters. One commonly-accepted method of identifying critical input parameters is sensitivity analysis, which requires a large number of runs that are both time consuming and may require high processing capacity.

This paper utilizes the Energy, Carbon and Cost Assessment for Buildings Stocks (ECCABS) model, which calculates the net energy demand of buildings and presents aggregated and individual- building-level, demand for specific end uses, e.g., heating, cooling, lighting, hot water and appliances. The model has already been validated

* Corresponding author.
E-mail address: reza.arababadi@asu.edu

Peer-review under responsibility of organizing committee of the International Conference on Sustainable Design, Engineering and Construction 2015
doi:10.1016/j.proeng.2015.08.478

using the Swedish, Spanish, and UK building stock data. This paper discusses potential improvements to this model by assessing the feasibility of using stepwise regression to identify the most important input parameters using the data from UK residential sector. The paper presents results of stepwise regression and compares these to sensitivity analysis; finally, the paper documents the advantages and challenges associated with each method.

Peer-review under responsibility of organizing committee of the International Conference on Sustainable Design, Engineering and Construction 2015

Keywords: Statistical Tenchniques, Energy Modeling, Stepwise Regrission, Sensitivity analysis

1. Introduction

Existing models aim to approximate the baseline energy use of existing building stock and provide an estimation of the future of residential energy demand. Most of the previously built regional residential stock models are bottom up engineering models and need assumptions both in the absence of direct data and in the application of input values where some supporting data are available.

Energy, Carbon and Cost Assessment for Building Stocks (ECCABS) model is designed to assess the effects of Energy Saving Measures (ESM) for building stocks[1]. The main outputs from the model are net energy demand by end-uses, delivered energy (to the building), CO2 emissions, and costs associated with the implementation of ESM.

This model is a bottom-up engineering model, which means that calculation of the energy demand of individual buildings is based on the physical properties of the buildings and their energy use (e.g., for lighting, appliances, and water heating). The results are scaled-up to represent the building stock and thus, the model assumes that a number of buildings can be assigned as representative of the region to be evaluated. The ECCABS model was used to simulate the UK building stock [2] where 192 sample buildings in the residential sector were simulated. The current project uses the data presented in the previous work and presents a method to reduce the number of input parameters required for the model. Reducing the amount of input data will support efforts to gather data in regions for which information is lacking. Therefore, the buildings in this model are described with a restricted number of parameters. The levels of input data required to describe the energy system and their possible energy saving scenarios are also limited. The purpose of this paper is to identify the best technique among sensitivity and stepwise regression analysis to determine the most important input parameters for the ECCABS model in the UK residential sector.

2. Related Work

Statistical analysis is useful to manipulate or predict the future of climatic change, natural calamities and other identifiable factors. It is required to develop a framework and to interpret observations from real or digital data [3]. These statistical concepts are key to forecast and identify various real time issues. Finally, Chua states including statistical analysis in reporting increases the credibility of our findings [4]. Sensitivity analysis methods have been applied in various fields including complex engineering systems, economics, physics, social sciences, medical decision-making, and others. Furthermore, sensitivity analysis can play an important role in model verification and validation [5].

2.1 Sensitivity analysis in Energy Modeling

Most quantitative research uses statistical techniques to validate and prove their hypothesis and to analyze the data presented in the research. Sensitivity analysis can help in identifying critical control points, prioritizing additional data collection or research, and verifying and validating a model [5]. Lomas et al. (1992) tested three important sensitivity analysis tools and tested them with building thermal simulation programs. He concluded that Monte Carlo analysis is feasible although it has a disadvantage of determining only total uncertainties. According to Lam et al (1996) , people using sensitivity analyses are not clear that the results are based on how often the simulation model is used. Sensitivity analyses have been used in IMAGE (Integrated Method to Assess Greenhouse

Effect) using Weismann and Morris method and the best input parameters are selected and compared between two methods [8].

Tian (2013) classifies sensitivity analysis into local and global approaches. Local approach also called as one factor at a time approach is focused on effects of uncertain inputs whereas global approaches looks at the influences of uncertain inputs [9]. Hence global sensitivity analysis is more reliable as it the technique deepens the root of reliability.

2.2 Stepwise regression in energy modeling

The regression methods come under global sensitivity analysis approaches. The common technique used in the regression method is stepwise regression. On stepwise regression, the factors are run based on their level of importance and the selection is made with respect to results from statistical techniques [9]. Generally, sensitivity analysis doesn't require all the uncertain variables for the regression model. It is more appropriate to construct regression models in a stepwise manner [10]. According to Helton et al (2006), stepwise regression is first constructed with the most influential variable followed by the second-most influential variable.

Dhar et al (1998) utilized the stepwise regression method to identify the most important parameters for temperature based and generalized Fourier series modeling for energy use. Katipamula (1998) determined the strength of the variables using stepwise regression that is used for dual-duct constant volume (DDCV) system in three different buildings. They also compared two different models and results are based on the variables selected through stepwise regression analysis. Stepwise regression can have impacts on global sensitivities of either design objectives or constraints when utilizing results from optimized models.

2.3 Energy Models and Statistics

A variety of approaches, including annual building energy simulation, statistical analysis based on empirical data, and spreadsheet calculations, can be used to perform whole-building energy analysis. Electricity use for each end use are dependent variables and occupancy levels and utilization factors generally serve as independent factors for energy analysis on buildings. Regression analysis has been used for decades to measure the energy savings associated with building retrofits [13]. It is also used for verification of energy audits following IPMVP protocol 3 [13]. The PRISM® (Princeton Scorekeeping Method) is one of the earliest applications of using regression analysis to measure energy savings in commercial buildings [13]. Most sensitivity analysis methods involve a large number of simulation runs. Hence, it is necessary to automate the process of creating building energy models with the different combinations of the inputs.

Previous studies fail to examine the feasibility of using this method in a building stock model that can simulate the energy performance of large number of buildings. Identifying the most important input parameters in building stock modeling is important because in many regions, most of the data for input parameters is lacking and reducing the number of input parameters thus supports the effort to gather and disseminate the data for future analysis. This paper highlights ECCABS modeling and selection of their input parameters based on sensitivity analysis and stepwise regression analysis.

3. Methodology

In order to accomplish the objectives, a number of steps are taken. It is very common to utilize sensitivity analysis to identify the most important parameters in building energy modeling. While sensitivity analysis provides results that are robust and informative, it requires a large number of runs that is time consuming. Also, the complexity of the analysis requires high-speed processors when the parameters increase exponentially. The paper utilizes the results from both sensitivity analysis and stepwise regression and the advantages of each method over others based on various factors are analyzed. The objective of this paper is to validate the use of stepwise regression in place of sensitivity analysis for determining important input parameters for building stock modeling. The authors use the

ECCABS model to validate the use of stepwise regression by comparing the most important input parameters for ECCABS model resulting from sensitivity analysis and stepwise regression.

3.1 Sensitivity analysis method

Sensitivity analysis examines the changes in the model's output variables based on minor changes in the model's inputs (Firth, et al. (2009). The sensitivity analysis procedure helps to determine which variables have larger effects on outputs. It is not obvious that which parameter needs to be included which parameters could be ignored because of their level of importance. Based on Firth, et al. (2009) Sensitivity analysis should be undertaken in the following steps:

- Each input parameter should be assigned a set value (k_i)
- Each input parameter faces a small change Δk_i while the other input parameters are kept constant, i.e. $\pm$1% change in the input parameter.
- For each change in the input parameters the model is run
- New output variables are used to calculate the sensitivity coefficients and normalized sensitivity coefficients.

Sensitivity coefficients are given by:

$$\frac{\partial y_i}{\partial k_i} \approx \frac{y_i(k_i + \Delta k_i) - y_i(k_i - \Delta k_i)}{2\Delta k_i} \qquad i=1, n \quad \text{and} \quad j=1, m \tag{1}$$

Where y_i is the *i*th output variable, k_i is the *j*th input parameter, $\frac{\partial y_i}{\partial k_i}$ is the sensitivity coefficient for output variable y_i and input parameter k_i , and y_i ($k_{i}+\Delta k_i$) is the value of y_i when the input parameter k_i is increased by Δk_i.

In order to be able to make a comparison of sensitivity coefficients of input parameters with different units the normalized sensitivity coefficients must be calculated. Firth, et al. (2009) suggests a formula to calculate this coefficient (Equation 2).

$$S_{i,j} = \frac{k_i}{y_i}\frac{\partial y_i}{\partial k_i} \qquad i=1\dots n \quad \text{and} \quad j=1\dots m \tag{2}$$

3.1. Stepwise Regression Method

Stepwise regression is actually the step-by-step iterative construction of a regression model that involves automatic selection of independent variables [1]. Stepwise regression is done either by selecting one independent variable at a time and including it in the regression model if it is statistically significant (forward selection), or by including all independent variables in the model and removing those that are not statistically significant (backward). In this project the forward method is used to select the most significant variables.

Prior to the stepwise regression we have done a normality test on the response variable (net energy use) to check if it its distribution is normal. Box-cox transformation could be used to normalize data in case it was not normally distributed. After checking the normality of the response variable the predictors must be prepared for stepwise regression. Some predictors might be constant for all sample buildings and some might be strongly correlated which will not enter the regression.

4. Analysis

ECCABS model calculates the energy demand of building stocks based on the thermal and physical properties of the sample buildings. Table 1 shows results for the sensitivity analysis for ECCABS based on initial set values of the UK building stock. The indoor air temperature results in the most sensitivity (1.63 in residential sector). This can be explained as a 1% increase in the indoor air temperature leads to a 1.63 percent increase in the energy consumption of the buildings. This is considerably higher than the other Si, j values and suggests that the indoor air temperature is the key determinant energy use in buildings. The external surface and average U-value of the building have the second largest sensitivity. The negative sensitivity is because an increase in a number of parameters will make a decrease in space heating energy consumption.

Table 1. Results for sensitivity analysis in residential building stock obtained in this work

Input Parameters	Sensitivity Coefficient	Normalized Sensitivity Coefficient
Total windows surface of the building (Sw)	2.401	0.070
Coefficient of solar transmission of the window (Ts)	-56.428	-0.070
Mean U value of the building (Umean)	426.956	0.882
Shading coefficient of the window (Wc)	-56.428	-0.070
Frame coefficient of the window (Wf)	-60.769	-0.073
Boiler eff.	-440	-0.596
Area of heat floor space (A)	1.314	0.212
Average constant consumption of the appliances (Ac)	5.058	0.038
Total external surfaces of the building (S)	2.142	0.875
Consumption of the hydro pumps (HyP)	20.000	0.001
Average constant gain due to people in the building (Oc)	-13.913	-0.028
Demand of hot water (Hw)	23.000	0.206
Minimum indoor temperature (Tmin)	53.236	1.63
Sanitary ventilation rate (Vc)	120.000	0.007

4.1. Stepwise Regression

The statistical analysis is starts with performing a normality test on the response variable. The normality probability plot is used to interpret the issues and also to identify the outliers. Figure 1 shows the test results. The deviations from the straight line suggest the deviations from normality and it is observed that the p-value for this data is notably smaller than the standard. This indicates that the null hypothesis is rejected and the response variable, which is the net energy demand of sample buildings, is not normally distributed.

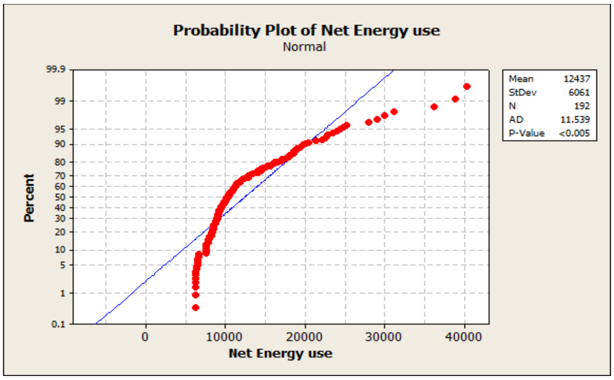

Fig. 1. Normality test on the response variable

We need normally distributed data to use the of statistical analysis tools, stepwise regression in this work. We had to take the appropriate remedial actions to make the data normally distributed. Data transformation, and particularly the Box-Cox power transformation, is a powerful tool that may help to make data normal [14]. The next step was trying to improve the data by use of Box-Cox transformation which is a useful data transformation technique used to make the data normally distributed. We know that as the Box-Cox power transformation actually does not really check for normality it is not an assurance for normality [14]. Therefore, it is absolutely necessary to always check the transformed data for normality using a probability plot.

Figure 2 shows the normality test done on the transformed data. Comparing figure 1 and figure 2 shows a significant improvement in the normality test. It is not still a perfect normal data set but with α=0.015 the null hypothesis is not rejected. We assume the transformed data as normally distributed data and build the regression on this data.

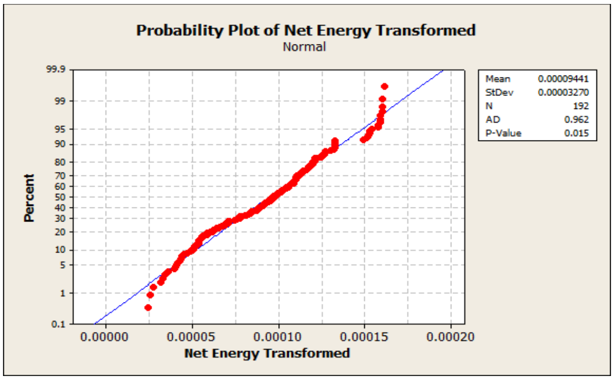

Fig. 2. Normality test on the response variable after the Box-Cox transformation

Data is now ready for stepwise regression. In this project we have decided to use the forward selection method which involves starting with no variables in the model, testing the addition of each variable using a chosen model comparison criterion, adding the variable (if any) that improves the model the most, and repeating this process until none improves the model. Below are results of stepwise regression on the 192 sample buildings. Notice that some input parameters could not enter to the regression process because they were either correlated or they were constant for all 192 samples:

Table 2. Results for stepwise regression in residential building stock obtained in this work

Steps	Input Parameters	RMSE	R-sq	R-Sq adjusted
1	S	0.000022	52.93	52.69
2	U	0.000010	90.99	90.90
3	A	0.000009	92.39	92.27
4	Oc	0.000008	94.34	94.22
5	Pfh	0.000007	95.29	95.17
6	Vcn	0.000007	95.45	95.30
7	Sw	0.000007	95.52	95.35
8	Ac	0.000007	96.00	95.82
9	tcc	0.000006	96.24	96.05

4.2. Stepwise Regression Model validation

Regression model validation is the method of determining whether the results obtained from regression analysis, are accurate and acceptable as descriptions of the data. The validation process can include checking the appropriateness of fit of the regression, and analysis of the regression residuals [15].

As the stepwise regression results show the R-Sq (adj) is 96% and the RMSE is very small. We know that, a high R^2 cannot assure that the model is valid. R^2 can always be improved by addition of variables into the model. To avoid such fake increases of the R^2, we have instead used the adjusted $R^{2.}$ Adjusted R^2 improves just if the added variable can notably improve the model [16].

To observe how the Box-Cox transformation helped improving the fitted regression model the residual plots for both before and after Box-Cox transformation is illustrated in figure 4. When model fits to the data the residuals are estimated random errors that make the relationship between the explanatory variables and the response variable.

Thus, if the residuals have a random appearance we conclude that the model fits the data well. If non-random behaviour observed in the residuals, it means model fits the data poorly [1]. Figure 3 makes it clear that the fitted model after the Box-cox transformation is more accurate. In regression analysis, the variance of the error terms must have a mean of zero. A residuals versus fitted values plot can check this. Below is the plot from the regression analysis above.

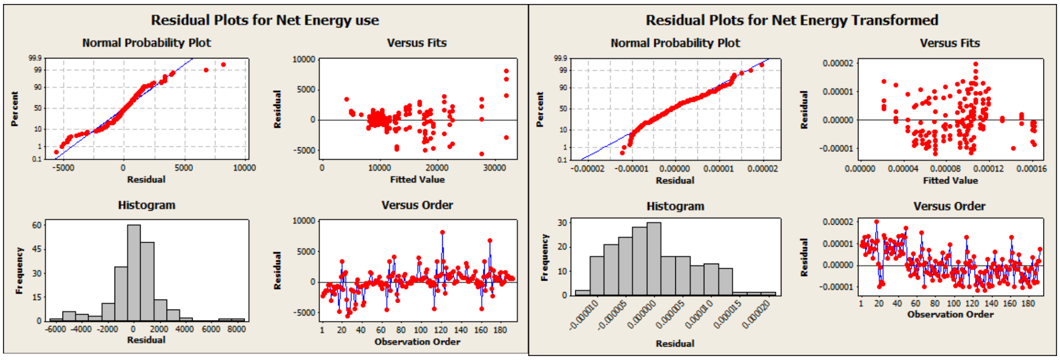

Fig.3. Residual plots before and after Box-Cox transformation

As figure 3 shows the points on fit plot are randomly distributed around zero, so we assume that the error terms have a mean of zero. At this stage buildings are divided into two groups and the stepwise regression is done on each group to see which parameters are more important in newer and older buildings.

5. Discussion

Table 3 compares the most important input parameters found by both the stepwise regression and sensitivity analysis. It is clear that they have both identified almost the same parameters with relatively the same order of importance. In the following parts sensitivity analysis is compared to stepwise regression and strengths and weaknesses of each are discussed.

Table 3. Variables of critical importance based on results of sensitivity analysis and stepwise regression

Rank	Sensitivity analysis	Stepwise regression
1	S	S
2	U	U
3	A	A
4	Oc	Oc
5	Vc	Pfh
6	Pfh	Vcn
7	Vcn	Sw
8	Tc	Ac

In this section, pros and cons of stepwise regression are analysed. Above all it should be mentioned that stepwise regression in this project could produce relatively accurate results. We know that sensitivity analysis requires frequent running of the simulation model which is both time consuming and requires an advanced processor if a large number of buildings are to be modelled. On the other hand the simulating model might not be available (as it was not available for this project). But, the stepwise regression can use a set of database to produce almost the same results. Moreover by use of the statistical measures it would very easy to assess the accuracy of the generated model.

Although stepwise regression has some significant benefits over the sensitivity analysis but there are some drawbacks which requires more attentions. For example:

- Input parameter which are strongly correlated would not be able to enter the regression process. Notice that in building modelling some input parameters are calculated by multiplying a constant into another input parameter. For example that lighting energy use and appliances energy use are calculated by multiplying the lighting and appliances energy density onto total floor area which makes them correlated.

- Input parameters which are identical for all buildings would not be able to enter the regression process. We know that in building stock modelling it is common to assume the same values for some input parameters in all sample buildings.

6. Conclusion

This project makes it clear that the stepwise regression is feasible to be used as a substitution for sensitivity analysis provided that the input data parameters are not strongly correlated nor being identical for all samples. It is still possible to use the stepwise regression, as it is usually easy to identify the correlated and non-varying parameters in building stock modelling. Stepwise regression can help the energy modellers to save plenty of time, which is required to run the simulating model in sensitivity analysis.

Results of this project can help the model developers to identify the most affluent inputs. Once they know the most important parameters they will be able to reduce the number of inputs to the model which makes it easier to use the simulating model. On the other hand the missing data which is very common in building stock modelling cannot prevent the use of simulating model. It would be easier less time consuming to gather the input data when the number of input parameters is reduced.

As part of future work the same method which is used in this project can be tested on the data from other countries to both compare to UK and to see if the affluent parameters are different in different countries.

7. References

[1] É. Mata, A. S. Kalagasidis, and F. Johnsson, "A modelling strategy for energy, carbon, and cost assessments of building stocks," *Energy Build.*, vol. 56, pp. 100–108, 2013.

[2] R. Arababadi, "Energy Use in the EU Building Stock-Case Study: UK," 2012.

[3] F. W. Zwiers and H. Von Storch, "On the role of statistics in climate research," *Int. J. Climatol.*, vol. 24, pp. 665–680, 2004.

[4] P.-P. Chua, "Statistics," 2005.

[5] H. C. Frey and S. R. Patil, "Identification and review of sensitivity analysis methods," in *Risk Analysis*, 2002, vol. 22, pp. 553–578.

[6] K. J. Lomas and H. Eppel, "Sensitivity analysis techniques for building thermal simulation programs," *Energy and Buildings*, vol. 19. pp. 21–44, 1992.

[7] J. C. Lam and S. C. M. Hui, "Sensitivity analysis of energy performance of office buildings," *Build. Environ.*, vol. 31, pp. 27–39, 1996.

[8] F. Campolongo and R. Braddock, "Sensitivity analysis of the IMAGE greenhouse model," *Environ. Model. Softw.*, vol. 14, pp. 275–282, 1999.

[9] W. Tian, "A review of sensitivity analysis methods in building energy analysis," *Renew. Sustain. Energy Rev.*, vol. 20, pp. 411–419, 2013.

[10] J. C. Helton, J. D. Johnson, C. J. Sallaberry, and C. B. Storlie, "Survey of sampling-based methods for uncertainty and sensitivity analysis," *Reliab. Eng. Syst. Saf.*, vol. 91, no. 10–11, pp. 1175–1209, Oct. 2006.

[11] A. Dhar, T. Reddy, and D. Claridge, "Modeling hourly energy use in commercial buildings with Fourier series functional forms," *J. …*, vol. 1, no. August 1998, 1998.

[12] S. Katipamula, "Multivariate regression modeling," *J. …*, vol. 1, no. August 1998, 1998.

[13] M. Donnelly, "LEAN Energy Analysis Using Regression Analysis to Assess Building Energy Performance Issue Brief," no. March, 2013.

[14] G. E. P. Box and D. R. Cox, "An analysis of transformations," *J. R. Stat. Soc. Ser. B*, pp. 211–252, 1964.

[15] R. Snee, "Validation of regression models: methods and examples," *Technometrics*, vol. 19, no. 4, pp. 415–428, 1977.

[16] D. B. Montgomery and D. G. Morrison, "A Note On Adjusting R[Sup2]," J. Finance, vol. 28, pp. 1009–1013, 1973.