A RANDOMISED APPROACH TO MULTIPLE REGRESSION ANALYSIS OF BUILDING ENERGY SIMULATION

Sam C. M. Hui
Department of Architecture
University of Hong Kong
Pokfulam Road, Hong Kong

ABSTRACT

Multiple regression analysis (MRA) is useful for developing energy prediction equations from the results of building energy simulation. However, if many design parameters are involved, a very large number of simulations is needed to generate data for the MRA. To tackle this problem, a randomised approach to MRA is proposed so that less simulations will be needed to generate the data. This paper presents a research study in Hong Kong that investigates building energy performance using a randomised MRA method. A parametric study has been conducted using the DOE-2.1E building energy simulation program to derive a set of energy equations relating 12 design parameters. The parameters are selected from a sensitivity analysis and are taken from three groups: load, system and plant. It is found that the energy equations generated from a set of 100 randomised simulations can do energy predictions quite well. It is believed that this approach has good potential for reducing the efforts required for MRA and making MRA easier for average building design professionals.

INTRODUCTION

Regression analysis is a statistical technique used to relate variables (Bowerman and O'Connell, 1990). Its primary objective is to build a mathematical model relating a dependent variable to independent variable(s). If there are more than one independent variable involved, the association is known as a multiple regression model (Draper and Smith, 1981).

Multiple regression analysis (MRA) is often used to study the effects of different design parameters on building energy performance (Misuriello and Fireovid, 1989; Sullivan, et al., 1985; Sullivan and Nozaki, 1984; Chou and Chang, 1993). It has also been employed for developing simplified energy equations and design tools for building energy standards (Wilcox, 1991; O'Neill, Crawley and Schliesing, 1991; Sander, et al., 1993; Sullivan, et al., 1992). To provide the data for deriving algebraic expression using MRA, a set of simulation results is usually generated by varying the input parameters of building energy simulation.

However, if the number of design parameters involved is large, a very large number of simulations will be needed to generate data for the MRA. For example, if there are 12 parameters with 3 perturbations in each of them, the total number of simulations required for all the combinations is equal to 3 to the 12^{th} power, which is 531,441 simulations. This may present great difficulty to carrying out of the study and analysis. To tackle this problem, a randomised approach to MRA is proposed so that less simulations will be needed to generate the data. By creating randomised inputs to the simulation, a set of 'normally distributed' simulation results can be obtained for deriving the regression model.

This paper presents a research study in Hong Kong that investigates building energy performance using the randomised MRA approach. A parametric study has been conducted using the DOE-2.1E building energy simulation program to derive a set of energy prediction equations relating 12 design parameters. It is believed that this approach has good potential for reducing the efforts required for MRA. By making MRA procedure easier and economical for average building design professionals, the study of building energy performance can be facilitated.

REGRESSION ANALYSIS TECHNIQUES AND ENERGY EQUATIONS

Regression analysis is a method for analysing data of ordered pairs or groups. The data is usually plotted, one versus the other and then modelled. The model in a two-dimensional case is a best fit curve. The modelling assumption is a best fit line computed by the least squares method (Milton and Arnold, 1990).

The purpose of regression analysis is to improve our ability to predict the next 'real world' occurrence of the dependent variable. Such a regression equation provides the ability to predict one variable on the basis of the knowledge of the other variable. This is useful for the study of building energy performance which is intricately and indirectly related to the design parameters.

MRA, or more precisely multiple 'linear' regression analysis, is a way to relate the building energy performance to many design variables in the simulation input using a linear form of equation. A database of energy use is created by performing a number of simulations for several important parameters, and then the database is used to develop simplified energy prediction equations (Chou, Chang and Wong, 1993; Turiel, et al., 1984).

Energy equations are established to provide an effective means for analysing building energy performance and energy targets (Briggs and Brambley, 1991). They also provide a guidance on the relative importance of design parameters. Significant parameters usually have large regression coefficients in the energy equation.

To derive energy equations from building energy simulation, the input design parameters can be categorised into three main types (Lam and Hui, 1993):

(a) Building load — LOAD

(b) HVAC system — SYSTEM

(c) HVAC refrigeration plant — PLANT

If the parameters of these three types are considered as three separate group functions, the general form of energy equation will look like this:

$$E = Function\ [(Load), (System), (Plant)] \quad (1)$$

where E = energy or load index, such as annual MWh or peak kW demand
(Load) = function of LOAD parameters
(System) = function of SYSTEM parameters
(Plant) = function of PLANT parameters

The simplest formula for expressing the above function will involve adding of the three group functions, like this:

$$E = K + (Load) + (System) + (Plant) \quad (2)$$

where K = regression constant

Another way of expression is to multiply all group functions, like this:

$$E = (Load) \times (System) \times (Plant) \quad (3)$$

There are of course many other forms of equation possible, but to simplify the analysis and facilitate future applications, the above two general forms are commonly employed. These two forms are used to develop the energy equations in this study.

THE RANDOMISED APPROACH

The randomised approach makes use of a random number generator to produce perturbation values for the input design parameters under study. By submitting the randomised input files to the simulation program, a set of simulation results representative of the defined ranges can be generated for the MRA. The procedure begins by selecting the design parameters, defining the range of perturbation of each of them and generating random values in this range. Input files for simulation are then prepared based on the randomised inputs and are eventually submitted to the simulation program. Figure 1 shows a flow chart of the process. The process is best carried out by automated simulation batch file or program.

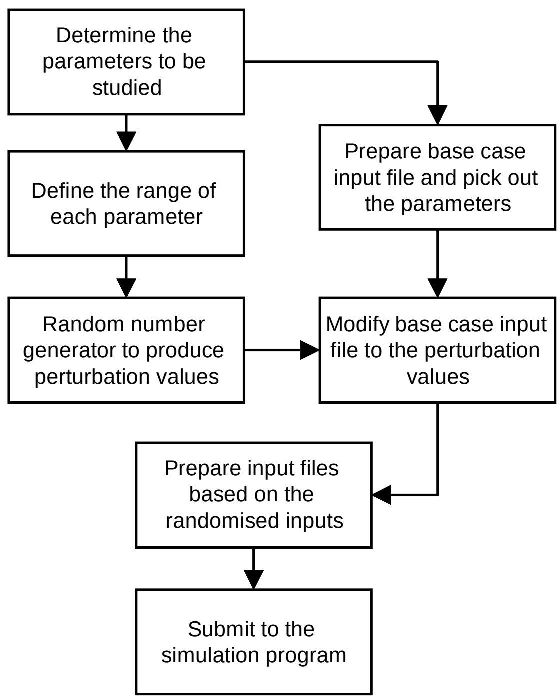

Figure 1 Flow chart of the randomised approach

By selecting the input values randomly for each parameter under study, the input domain, and hopefully the output response, will be randomly or normally distributed. The technique can be applied to most numerical parameters (those defined by numbers), but may be limited for non-numerical ones, such as system types and schedules.

Depending on the objective of the study, different simulation results may be chosen as the 'dependent variable' for the regression analysis. The annual building energy consumption and peak energy demand are often considered in the study of building energy performance.

Although the concept of the randomised approach to MRA is straightforward, implementation of the approach will require particular care since the existing simulation programs do not allow for that. A systematic way to perform the simulation and to manage the data is essential. Table 1 gives an example of standard form for recording the major inputs and outputs from the simulation. This can facilitate the analysis and avoid errors in the computations.

Table 1 Example of standard form for recording major simulation inputs and outputs

Run No.	Values of design parameters					Results	
	X_1	X_2	X_3	X_4	…	Y_1	…
1							
2							
3							
…							

To examine the application of this approach in practice, a parametric study has been conducted in Hong Kong using a detailed building energy simulation program.

PARAMETRIC STUDY USING DOE-2

The simulation tool used in this study is the personal computer (PC) version of the DOE-2.1E program developed by the Lawrence Berkeley National Laboratory (Acrosoft International, Inc., 1994). A series of simulations has been performed to develop 12-parameter regression models for studying the energy performance of office buildings in Hong Kong.

To facilitate the analysis, a supporting utility program developed by the author has been used to perform the simulations (Hui and Lam, 1995). This program automates the simulation process, extracts key results and handles simulation input and output. Another batch program has been used to prepare the randomised inputs and modify the base case input file for the parametric study.

A base case model building has been established to serve as a reference point for evaluation. The model building is a 40-storey square office building (35 m by 35 m) with curtain-wall construction and a central air-conditioning system. Figure 2 shows the base case model building. The input design parameters and detailed configuration of the base case model building can be found in Lam and Hui (1996).

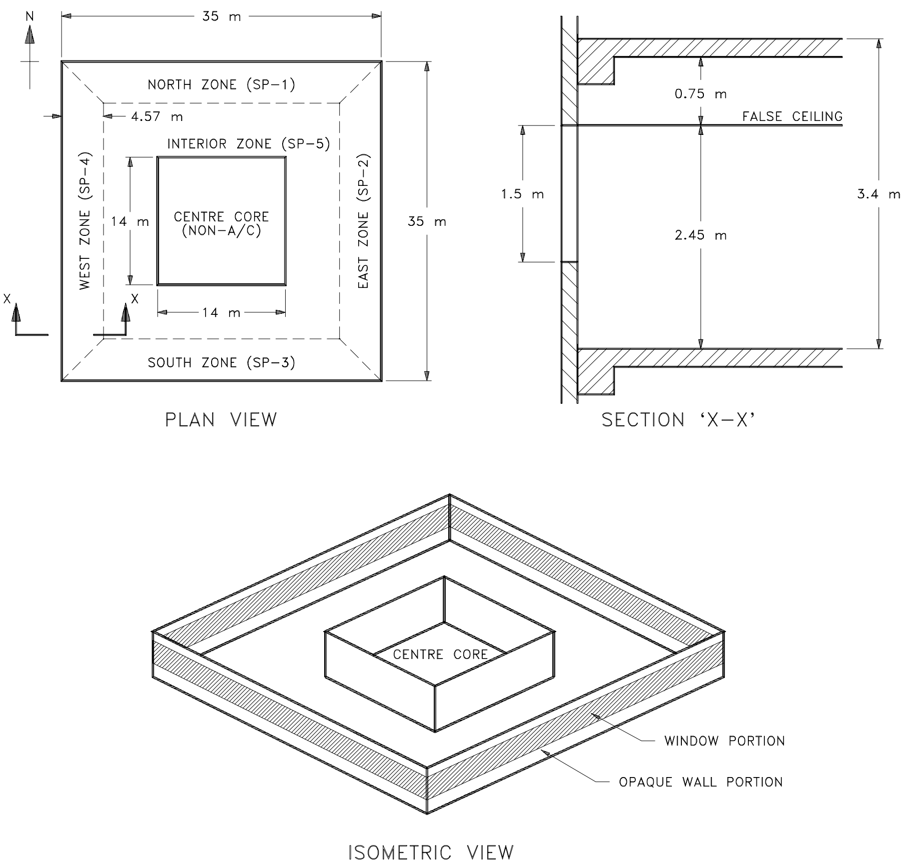

Figure 2 Typical floor of base case model building

The weather file used in this study contains full hourly data of Hong Kong for the year 1989. The year 1989 has been selected as the Test Reference Year (TRY) for Hong Kong (Hui and Lam, 1992) and is suitable for comparative energy studies. The weather file used in this study is stored in Typical Meteorological Year (TMY) format (NCC, 1981).

To help assess and select the variables, techniques like sensitivity analysis and graphical diagnostic are useful before the actual regression process. The selected parameters should make physical sense as well as being useful predictors (Sullivan, et al., 1985). Engineering judgement is often required to make a reasonable choice.

In this study, a sensitivity analysis has been performed before to identify the parameters that are most important to building energy consumption in Hong Kong (Lam and Hui, 1996). Twelve significant parameters have been determined in the sensitivity analysis. Six parameters are from building load, four from HVAC system and two from HVAC refrigeration plant. Table 2 gives a list of the parameters, their base case values and the simple statistics of the randomised values for each parameter.

In order to develop the regression models, some 100 simulations has been performed to generate the data which are then taken to the regression analysis tool. In this study, the MRA was performed on a statistical package, SPSS (Norusis, 1993a). Energy equations were determined for both annual building energy consumption (in MWh) and peak electricity demand (in kW). They are considered the most important outcomes of the simulation.

During the statistical analysis, the 'adding' expression (Equation (2)) can be established using simple multiple regression procedure or even a spreadsheet program, but the 'multiplying' expression (Equation (3)) requires more complicated procedure. To develop the multiplying expression, nonlinear regression is used to derive prediction equation. The nonlinear regression procedure solves the regression problem by iteration (Norusis, 1993b). The 'goodness of fit' of the model is measured by the coefficient of determination (R^2) (Norusis, 1993a). R^2 is equal to unity if a perfect fit is found.

EVALUATIONS AND DISCUSSIONS

As mentioned earlier, the annual building energy consumption (MWh) and the peak electricity kW are the objective functions for the regression analysis. To get a better regression fit, some of the parameters has been transformed (such as FE, the inverse of fan efficiency). New terms created by multiplying two parameters (such as SC x WR) have also been added during the model evaluation (Lam and Hui, 1996). Several variations of the regression model have been tested and the acceptable models are eventually selected based on interpretability, clarity and ease of use. Table 3 gives the regression models for MWh and peak kW developed based on the 'adding' and 'multiplying' forms of equations.

Table 2 Design parameters selected for the multiple regression analysis

Group	Abbr.	Design parameter	Unit	Base case value	For the 100 randomised inputs		
					Min.	Max.	Mean
Load	SC	Shading coefficient of windows	---	0.4	0.1	0.98	0.54
	WR	Window-to-wall ratio	---	0.44	0.1	0.89	0.47
	AT	Space air temperature	^oC	25.5	21.0	29.9	25.2
	EQ	Equipment load	W/m^2	15	0.9	29.6	14.3
	LL	Lighting load	W/m^2	20	0.1	29.9	14.5
	OC	Occupant density	$person/m^2$	0.2	0.1	1.0	0.28
System	OA	Outdoor air flow	l/s/person	7	1.9	18.6	10.8
	TS	Cooling thermostat setpoint	^oC	25.5	21.0	29.9	25.2
	FE	Inverse of fan efficiency	---	1.82	1	10	2.89
	FS	Supply fan static	Pa	1369	0	2962	1516
Plant	CH	Chilled water supply temperature	^oC	6.7	4.0	9.9	6.7
	CP	Chiller coefficient of performance	kWe/TR	1.2	0.51	1.99	1.20

Note: The minimum, maximum and mean are calculated from the 100 randomised inputs.

Table 3 Regression models for the twelve parameters

Multiplying Model for 12 Parameters (SC, WR, AT, EQ, LL, OC, OA, TS, FE, FS, CH, CP)

MWh = *(Load) x (System) x (Plant)*
= (-1.38 - 7.3 SC x WR - 0.0151 AT - 0.167 EQ - 0.206 LL - 9.6 OC)
x (-340 - 73.7 OA - 412 FE - 0.406 FS - 0.0000384 FS x FS - 0.644 TS x TS
+ 2.49 OA x TS + 16.7 TS x FE + 0.0215 TS x FS - 0.0872 FS x FE)
x (0.508 - 0.0109 CH + 0.311 CP)
[R^2 = 0.9880]

Peak kW = (0.00073 + 0.00134 SC x WR - 0.0000139 AT + 0.0000152 EQ
+ 0.0000159 LL + 0.0014 OC) x (57381 + 39439 OA + 106289 FE
+ 138 FS + 0.0103 FS x FS + 321 TS x TS - 1352 OA x TS - 4019 TS x FE
- 6.55 TS x FS + 23.8 FS x FE) x (3.45 + 0.113 CH + 4.76 CP)
[R^2 = 0.9389]

Adding Model for 12 Parameters (SC, WR, AT, EQ, LL, OC, OA, TS, FE, FS, CH, CP)

MWh = *K + (Load) + (System) + (Plant)*
= -4107 + (4757 SC x WR - 20.5 AT + 166 EQ + 223 LL + 9120 OC)
+ (-415 OA + 4315 FE + 6.79 FS + 0.000672 FS x FS + 2.26 TS x TS
+ 18.9 OA x TS - 193 TS x FE - 0.367 TS x FS + 1.09 FS x FE)
+ (- 304 CH + 3891 CP)
[R^2 = 0.9202]

Peak kW = -3814 + (4707 SC x WR - 47.1 AT + 62.1 EQ + 80.9 LL + 6215 OC)
+ (88.6 OA + 1654 FE + 3.24 FS + 0.000439 FS x FS + 2.74 TS x TS
- 1.22 OA x TS - 68.1 TS x FE - 0.18 TS x FS + 0.445 FS x FE)
+ (-39.2 CH + 2922 CP)
[R^2 = 0.9361]

Note: Please see Table 2 for units of the parameters

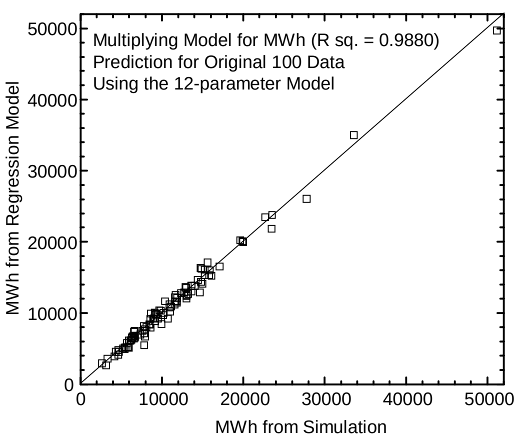

Figure 3 Goodness of fit for MWh prediction by the multiplying model

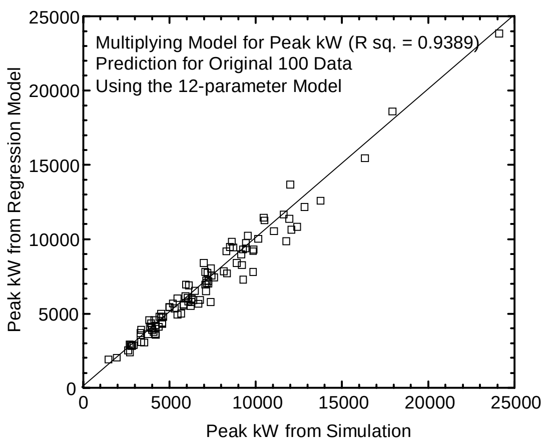

Figure 4 Goodness of fit for peak kW prediction by the multiplying model

It can be seen that the multiplying form gives a slightly better fit (R^2 = 0.988 for MWh and R^2 = 0.9389 for peak kW) than the adding form (R^2 = 0.9202 for MWh and R^2 = 0.9361 for peak kW). These equations can be used to predict the effect of the design parameters on building energy performance, provided that the value of the parameters lies within the ranges concerned. The relative importance of the design parameters can also be observed from the regression coefficients. For example, the terms OC (occupancy density) and SC x WR (product of shading coefficient and window-to-wall ratio) are found very important in the energy equations.

Figure 3 shows the goodness of fit for MWh prediction by the 12-parameter multiplying model. The data points indicate the distributions of MWh in the original 100 randomised sets of input which are used to build the model. The diagonal line is the perfect fit line. Figure 4 shows similar information for peak kW.

To check how well the models can predict the next occurrence, some test cases has been performed on the multiplying models. Another 30 sets of randomised input have been generated and simulations were carried out to calculate the values of annual MWh and peak kW. The results from the simulation were then compared with those predicted using the regression models. Figures 5 and 6 show the test case performance of the models for predicting MWh and peak KW, respectively. Table 4 gives the key statistics of the results for the test cases. It can be seen that the model can perform quite well for the test cases. The averages of the predicted results obtained from the regression models are close to those of the simulated results. The root mean square difference (RMSD) deviates more and this implies that the prediction for an individual case has larger uncertainty --- a very basic statistical interference (Bowerman and O'Connell, 1990).

It is believed that the number of randomised simulations required for developing the regression model depends on the number of the design parameters and their characteristics. Generally speaking, the ability of the regression model can be improved by performing more simulations and supplying more data to the MRA. A feedback mechanism may be installed in the simulation cycle to check for the necessity of including more simulations, if a certain criteria is met, say, R^2 = 0.98. It should be noted that the randomised approach is

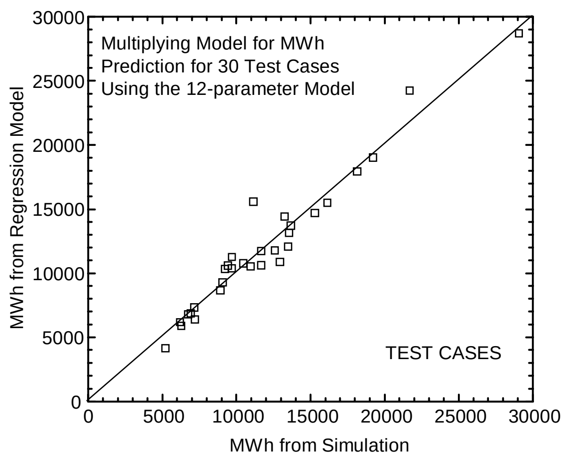

Figure 5 Test cases for MWh prediction

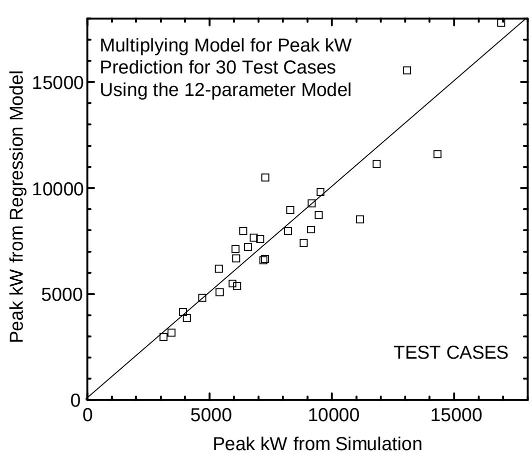

Figure 6 Test cases for peak kW prediction

Table 4 Key statistics of the results for the 30 test cases

	Mean of simulated results	Mean of predicted results	Diff. between the two means		RMSD of the data	
			in unit	in % of mean	in unit	in % of mean
MWh	11874 MWh	11997 MWh	123 MWh	1.0%	1221 MWh	10.3%
Peak kW	7752 kW	7802 kW	50 kW	0.6%	1219 kW	15.7%

Note: 1. RMSD = root mean square difference $= \sqrt{\dfrac{\sum(\text{Predicted} - \text{Simulated})^2}{\text{Total Number of data}}}$

2. The % of mean refers to the mean of the simulated results.

more effective when the number of design parameters is large. It may become less attractive if only a few parameters are being studied.

The methodology presented here can be applied to other simulation programs and building models. The simulation tool is not necessarily a full hourly one; simplified energy programs can be used to perform the MRA. Regression equations for some other simulation results, such as building cooling and heating loads (Sullivan, et al., 1992), can also be developed in the same way.

When applying MRA to develop energy equations, care should be taken to the collinearity between independent variables as this may affect the stability of the model and the relative importance of the parameters (Reddy and Claridge, 1994). Another limitation is the range of data. Results of MRA should not be extrapolated beyond the data range because the underlying relationships between dependent and independent variables may change considerably (Leslie, Areta and Sliwinski, 1986).

CONCLUSIONS

Building energy simulation is playing an increasingly important role in building design and building energy standards. By using the MRA technique, energy prediction equations can be established for estimating building energy performance and for building energy code compliance. Many regression studies nowadays require generation of a large database of simulation results (Crawley, 1995; Sander, et al., 1993; O'Neill, Crawley, and Schliesing, 1991). If the randomised approach can be adopted, the efforts required for performing and interpreting the simulations can be reduced. This will facilitate the investigation of building performance and encourage design professionals to use the MRA method.

ACKNOWLEDGEMENT

The author is grateful to Dr. Joseph C. Lam of City University of Hong Kong for his guidance in the early simulation research for this study.

REFERENCES

Acrosoft International, Inc., 1994. *MICRO-DOE2 E Version POWERPATH/NETPATH Edition User's Guide*, Denver, Colorado, March.

Bowerman, B. L. and O'Connell, R. T., 1990. *Linear Statistical Models: An Applied Approach*, Second Edition, Duxbury Press, Belmont, California.

Briggs, R. S. and Brambley, M. R., 1991. Whole-building energy targets: a methodology for future performance-based standards, In *Proc. of the Building Simulation '91 Conference*, August 20-22, 1991, Sophia-Antipolis, Nice, France, pp. 631-637.

Chou, S. K. and Chang, W. L., 1993. Development of an energy-estimating equation for large commercial buildings, *International Journal of Energy Research*, 17 (1993): 759-773.

Chou, S. K., Chang, W. L. and Wong, Y. W., 1993. Effects of multi-parameter changes on energy use of large buildings, *International Journal of Energy Research*, 17 (1993): 885-903.

Crawley, 1995. Development of a database of building envelope simulations for evaluating building envelope retrofits, In *Proc. of the Building Simulation '95 Conference*, August 14-16, 1995, Madison, Wisconsin, pp. 467-474.

Draper, N. R. and Smith, H., 1981. *Applied Regression Analysis*, Second Edition, John Wiley & Sons.

Hui, S. C. M. and Lam, J. C., 1992. Test reference year for comparative energy study, *Hong Kong Engineer*, 20 (2): 13-16.

Hui, S. C. M. and Lam, J. C., 1995. *Computer Programs for Automated Computer Simulation Using DOE-2 and BLAST*, Research Report BC/95/04, Department of Building and Construction, City University of Hong Kong.

Lam, J. C. and Hui, S. C. M., 1993. Computer simulation of energy performance of commercial buildings in Hong Kong, In *Proc. of the Building Simulation '93 Conference*, 16-18 August 1991, Adelaide, Australia, pp. 129-135.

Lam, J. C. and Hui, S. C. M., 1996. Sensitivity analysis of energy performance of office buildings, *Building and Environment*, 31 (1): 27-39.

Leslie, N. P., Aveta, G. A. and Sliwinski, B. J., 1986. Regression based process energy analysis system, *ASHRAE Transactions*, 92 (Part 1A): 93-102.

Milton, J. S. and Arnold, J. C., 1990. *Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computer Sciences*, Second Edition, McGraw-Hill, Singapore.

Misuriello, H. and Fireovid, J., 1989. ASEAM-2.1 applications using parametric studies and multivariate regression techniques, In *Proc. of the ASEAN Special Sessions of the ASHRAE Far East Conference on Air Conditioning in Hot Climates*, Kuala Lumpur, Malaysia, October 26-28, 1989, pp. 154-170.

NCC, 1981. *Typical Meteorological Year User's Manual TD-9734: Hourly Solar Radiation - Surface Meteorological Observations*, National

Climatic Center (NCC), Asheville, North Carolina, May.

Norusis, M. J., 1993a. *SPSS for Windows, Base System User's Guide*, Release 6.0, SPSS Inc.

Norusis, M. J., 1993b. *SPSS for Windows, Advanced Statistics*, Release 6.0, SPSS Inc.

O'Neill, P. J., Crawley, D. B. and Schliesing, J. S., 1991. Using regression equations to determine the relative importance of inputs to energy simulation tools, In *Proc. of the Building Simulation '91 Conference*, August 20-22, 1991, Sophia-Antipolis, Nice, France, pp. 283-289.

Reddy, T. A. and Claridge, D. E., 1994. Using synthetic data to evaluate multiple regression and principal component analyses for statistical modeling of daily building energy consumption, *Energy and Buildings*, 21 (1994): 35-44.

Sander, D. M., et al., 1993. Development of a simple model to relate heating and cooling energy to building envelope thermal characteristics, In *Proc. of the Building Simulation '93 Conference*, 16-18 August 1993, Adelaide, Australia, pp. 223-230.

Sullivan, R., et al., 1985. Commercial building energy performance analysis using multiple regression, *ASHRAE Transactions*, 91 (Part 2A): 337-353.

Sullivan, R., et al., 1992. A residential fenestration performance design tool, *ASHRAE Transactions*, 98 (Part 1): 832-840.

Sullivan, R. T. and Nozaki, S. A., 1984. Multiple regression techniques applied to fenestration effects on commercial building energy performance, *ASHRAE Transactions*, 90 (Part 1A): 116-123.

Turiel, I., et al., 1984. Simplified energy analysis methodology for commercial buildings, *Energy and Buildings*, 6 (1984): 67-83.

Wilcox, B. A., 1991. Development of the envelope load equation for ASHRAE Standard 90.1, *ASHRAE Transactions*, 97 (Part 2): 913-927.