5th International/11th Construction Specialty Conference
5e International/11e Conférence spécialisée sur la construction

Vancouver, British Columbia
June 8 to June 10, 2015 / *8 juin au 10 juin 2015*

DEVELOPMENT AND VALIDATION OF REGRESSION MODELS TO PREDICT ANNUAL ENERGY CONSUMPTION OF OFFICE BUILDINGS IN DIFFERENT CLIMATE REGIONS IN THE UNITED STATES

Shideh Shams Amiri[1], Mohammad Mottahedi [2], Somayeh Asadi[1, 3], David Riley[1]
[1] Pennsylvania State University, United State
[2] University of Texas at San Antonio, United State
[3]asadi@engr.psu.edu

Abstract: Energy consumption in commercial buildings has been growing substantially in recent years. Recently, building energy consumption estimation tools have been used to calculate energy savings and emissions reduction. Energy performance of building is complicated since it depends on multiple variables associated to building characteristics, equipment and systems, weather, occupants, and sociological influences. Therefore, the objective of this study is to develop the multi-linear regression models to predict energy consumption of an office building in five different climate regions in the United States. In order to achieve this objective, a typical commercial building was selected and the effect of 17 key building design parameters on its energy performance was investigated. To quantify building energy consumption, eQuest and DOE-2, which are building energy simulation software programs, were used to develop the building profile and perform annual energy simulation. In addition, Monte Carlo simulation technique was used to create a ten thousands comprehensive dataset covering the full range of design parameters for each studied climate region. An in-house computer program was developed to implement the Monte Carlo simulation. Statistical analysis was performed using R statistical analysis program to develop a set of linear regression equations predicting energy consumption of each design scenario. The difference between obtained results from regression model and DOE-2 are largely within 5%. In addition, standardized regression coefficient was calculated to assess the sensitivity of heating, cooling, and total energy loads to different building design parameters across five climate zones. It is believed that the developed regression models can be used to estimate the energy consumption of office buildings in different climate regions when designers and engineers consider various building envelope designs in the early stages of the design.

1 INTRODUCTION

In recent years, the contribution of modern world to energy consumption has been increased significantly. World energy consumption has increased from 524 quadrillion Btu in 2010 to 630 quadrillion Btu in 2010 and 820 quadrillion Btu in 2040, a 30 year increase of 56% (EIA 2010). The same trend was seen in the United States in which total energy consumption was approximately 97.9 quadrillion Btu in 2010 with an increase rate of 8.3% (EIA 2010). According to the EIA (2010), building sector in the U.S. consume 40% of total energy which is higher than transportation and industrial sectors. Therefore, proper tools are needed to estimate and optimize energy consumption in buildings.

Several studies have been conducted to study building energy performance. In addition, there are different methods including simple regression analysis and dynamic simulation software programs (e.g.

EnergyPlus and DOE-2 (Repice 2011) to model building energy performance (Lam et al. 2010, Broun et al. 2014, Catalina et al. 2013, Asadi et al. 2012). In a study conducted by Hygh et al. (2012), EnergyPlus software was used to perform energy simulation and calculate annual building energy consumption of a commercial building in four different climate zones. Mohammadpour et al.(2014) employed EnergyPlus to model energy consumption of three retrofit projects and compared energy consumption before and after the retrofit. Asadi et al. (2014) developed multiple linear regression models to predict building energy consumption for a typical residential building in the hot and humid climate. The effect of 7 buildings shapes as well as 17 building design parameters including HVAC schedule, orientation, building envelop, etc. on building energy performance were investigated. Results of their study showed that there is a good agreement between results of the DOE-2 and regression equations and the error was less than 5% in most cases. In another study, Catalina et al. (2013) developed regression models to investigate monthly heating load in residential building in France. The inputs of the regression model include the window to wall ratio, building envelope U-value, and building shape factor. Their analysis indicated that there is a strong relationship between building shape and energy consumption. Later, Lam et al. (2010) developed regression models using DOE-2 simulation results to determine the impact of 12 building design sensitive variables on building energy performance. The authors reported that there is a strong correlation between annual building energy consumption and design parameters in the warm climates. This paper proposes a simple and realistic approach to estimate energy consumption of a typical office building in five different climate zones. The primary objective of this study is to develop a multi-linear regression model to predict and quantify energy consumption of a commercial building in the early stages of building design.

2 MATERIAL AND METHOD

Building energy simulation models are commonly used to predict energy performance. They are powerful computational tools helping users to model a building as a system and to identify potential opportunities to reduce building energy consumption. In the present study, a comprehensive set of inputs such as internal loads, mechanical and electrical system, orientation and occupancy schedule was considered to calculate energy consumption. Also five major climate zones including cold dry, cool dry, mixed humid, warm marine and hot humid were considered in this study (Table 1).

Table 1: Five selected cities in each climate region.

Climate	Representative city	HDD	CDD
Cold dry	Billings	>7000	<2000
Cool dry	Salt lake City	<5500-7000	<2000
Mixed humid	Washington DC	<4000-5499	<2000
Warm marine	San Jose	<4000	<2000
Hot humid	Houston	<4000	≥2000

HDD=Average heating degree-days, CDD= Average cooling degree-days (EIA, Noaa 2012)

Monte Carlo simulation was performed by randomly selecting 17 variables based on uniform distribution to generate a new input file for the simulation software. This process was repeated 10000 time to effectively examine the configuration space. The eQuest and DOE-2 software programs were utilized to calculate the annual heating and cooling consumption for each design scenario based on Monte Carlo simulation. eQUEST software, which adds an additional graphical wizards capability to DOE-2, facilitates creation of building envelope and climate zones. Using DOE-2 avoids imprecisions introduced by simplifying algorithms, and since it is a configurable tool, it can be utilized for detailed design. Based on the Monte Carlo simulation, 10,000 simulation runs were defined for each of the five climate zones, covering a complete range of design parameters. In addition, a code was written in Python’s programming language to help extracting required data from DOE-2. Then, these data were used to develop the multiple linear regression equations and investigate the relationship between different parameter and annual energy consumption. Figure 1 illustrates the framework of the analysis.

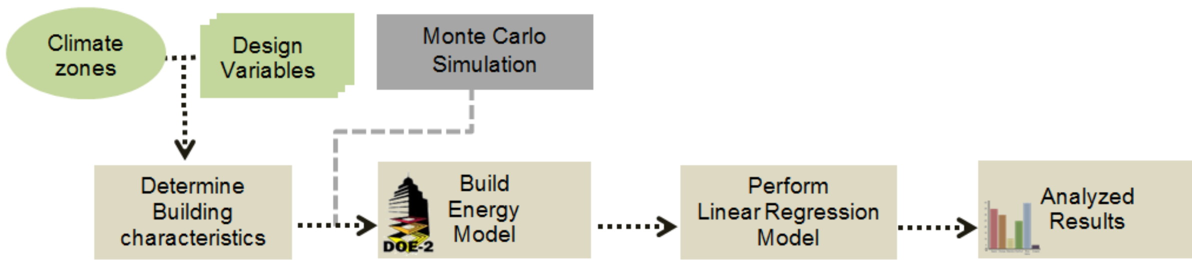


Figure 1: Framework of present study

2.1 Base Case Model Description and Design Variables

Table 2 shows the list of parameters that were used to build the office building model using eQuest and Table 3 represents the implemented variables in the Monte Carlo Simulation. As it can be seen in Table 2, 17 design parameters including building envelop, orientation and occupant schedules were considered. The properties of all building components including wall, roof, ceiling, foundation, and floors were defined in this study. For each parameter, set of values and ranges are selected based on AHRAE 90.1 (ASHRAE 2007). In addition, a comprehensive data set was generated based on random distribution to examine all possible configuration of building envelope. Uniform distribution was applied to each parameter ensuring that all values within the specified range are investigated equally for each design choice.

Table 2: Description of eQuest inputs.

Constant parameters	
Building type	Office bldg., two story
Jurisdiction	ASHRAE 90.1
Building Area	2322.6 m2
Cooling Equip	Chilled water coils
Heating Equip	Hot water coils
Analysis Year	2013
Day Light Control	Daylight control
Usage details	Hourly end use profile
Zoning Pattern	One per floor
Floor-To-Floor	2.74 m
Floor-To-Ceiling	2.43 m
Door type	Opaque
Door Construction	Wood, hollow core flush, 0.02-0.096m
Windows Area method	Present of Gross wall area
Floor to Floor Window ratio	40%
Net Floor to Ceiling Window ratio	53.3%
Window High	1.59 m
Cooling source (HVAC)	Evaporate resistance
Heating Source	Furnace
System type	Direct
Number of Occupants	105
People Activity	0.131 kw/hr.

Table 3: Implemented variables in the Monte Carlo Simulation

Variable	Range	U –value (W/m^2k)	Variable	Range	U –value (W/m^2k)
Top floor ceiling interior finish	• Acoustic Tile • Drywall Finish • Plaster Finish	•4.50 •12.6 •10.1	Floor Construction	• 0.05m Concrete • 0.10m Concrete • 0.15m Concrete • 0.20m Concrete	•20.0 •17.8 •11.7 •8.84
Top floor ceiling exterior insulation	• No Board Insulation • Polyurethane (R-6) • polyurethane (R-9)	•0.90 •0.72	Exterior wall absorbance	• light • Medium • Dark	N/A N/A N/A
Top floor batt insulation	• R-30 • R-45 • R-11 • R-19 •No Batt	•0.17 •0.7 •0.47 •0.27	Roof absorbance	• light • Medium • Dark	N/A N/A N/A
Ceiling Interior finish	• Acoustic Tile • Drywall Finish • Plaster Finish	•4.50 •12.6 •10.1	Roof Construction	•ASHRAE Roof # 2,9,11, 16,20 26,28,33,35	•min =0.35 •max=0.74
Ceiling Insulation Parameters	• Wool Batt (R11) • Wool Batt (R19) • Wool Batt (R30)	•0.47 •0.27 •0.17	Interior Wall	•ASHRAE Wall # 3,10,11,17 23,27,31,32,34,38, 39,41,43,35,47	•min=0.17 •max=3.3
Ground Floor Construction	Concrete • 0.1m •0.3m • 0.15m •0.2m	•17.8 •7.69 •17.7 •8.84	Exterior Wall	•ASHRAE Wall #1, 3,6,11,12,19 25,27,29,30,32	•Min=0.19 •max=2.65
Ground Floor Interior Finish	•Carpet (No Pad) •Vinyl Tile •Ceramic/Stone Tile	•0.21 •0.007 •0.004	Glass Category	• Single Low-e • Double Low-e • Triple Low-e	•2.0 •1.0 •7.0
Floor Interior Finish	•No Surface Finish • Carpet (No Pad) • Vinyl Tile/Stone	•4.7 •11.1	Building Orientation	•360 ° •180° •90 ° •270°	
Occupant Schedule	• 08:00:00 AM to 05:00:00 PM (Monday-Friday) +HVAC [1] • 08:00:00 AM to 06:00:00 PM (Monday-Thursday) +HVAC [1] • 07:00:00 AM to 05:00:00 PM (Monday-Thursday) +HVAC [1] • 07:00:00 AM to 04:00:00 PM (Monday-Friday) +HVAC[1] • 08:00:00 AM to 05:00:00 PM (Monday-Friday) +HVAC[2] .... •07:00:00 AM to 04:00:00 PM (Monday-Friday) +HVAC[3]				

[1]HVAC system turns on 1 hour before working hours and turn off 1 hour after working hours.
[2]HVAC system is on 24/7.
[3]HVAC system is on only during workings hours.

2.2 Regression Analysis

The aim of regression analysis in this study is to develop simple and accurate models to predict energy consumption in commercial buildings. A multiple regression model with more than one explanatory variable may be written as:

[1] $Y = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \ldots + \beta_n\chi_n$

Where y is the output, β_i is the regression parameters and χ_i is the input variables. The least-squares method is generally used for estimation purposes in the multiple-regression model. Once regression coefficients are identified, a prediction equation can then be used to estimate the value of a continuous output as a linear function of one or more independent inputs. A comprehensive dataset was developed based on the randomly generated building parameters using energy simulation model. Eighty percent (80%) of the simulation runs were selected randomly and used to develop the regression equations. Remaining twenty percent (20%) of the runs were used to validate the developed model. The generated dataset was used to develop regression equations predicting annual building energy consumption.

3 RESULTS AND DISCUSSION

3.1 Interaction between parameters

Analysis of the Interaction between parameters represents the combined effects of the independent parameters on the dependent variable. When an interaction effect is present, the impact of one factor depends on the level of the other factor. One of the methods to determine the interaction between parameter is to identify multicollinearity. Multicollinearity is a statistical phenomenon in which two or more predictor variables in a multiple regression model are strongly correlated. It arises when two or more predictors in the model are correlated and provide redundant information about the response. Generalized variance-inflation factor (GVIF) can be used to detect multicollinearity in the regression equation. The GVIF indicates the degree to which the confidence interval for that variable regression parameter is expanded relative to a model with uncorrelated predictors. As a general rule, GVIF>4 indicates a multicollinearity problem. The GVIF results are presented in Table 4. As it can be seen in this table, the GVIF values in all cases are less than 1.3 indicating that there is no correlation between predictor variables in the multiple regression models.

Table 4: Generalized variance-inflation coefficients

	Billings	Houston	Washington, D.C.	San Jose	Salt Lake City
Building Orientation	1.038285	1.035407	1.028915	1.034806	1.030121
Top Floor Batt Insulation	1.03594	1.035997	1.034708	1.038009	1.036116
Ceiling Interior Finish	1.023482	1.021158	1.020389	1.02523	1.024162
Ceiling Insulation	1.026014	1.02497	1.023632	1.02252	1.024083
Floor Construction	1.038519	1.034395	1.037726	1.03494	1.026742
Top Floor Ceiling Exterior Insulation	1.039947	1.034008	1.037932	1.031834	1.034814
Top Floor Ceiling Interior finish	1.048875	1.044126	1.045476	1.047932	1.047745
Ground Floor Construction	1.022871	1.024229	1.018244	1.022585	1.024278
Ground Floor Interior Finish	1.038625	1.032505	1.034519	1.028883	1.032269
Floor Interior Finish	1.025238	1.025422	1.022385	1.018312	1.02167

Interior Wall	1.230111	1.215584	1.214586	1.220215	1.200708
Exterior Wall	1.197334	1.210685	1.204963	1.199288	1.18698
Roof Construction	1.093883	1.110539	1.101251	1.094528	1.099383
Exterior Wall Absorbance	1.021308	1.02599	1.028086	1.01917	1.025674
Roof Absorbance	1.025294	1.025288	1.023411	1.020919	1.022567
Occupant Schedule	1.129032	1.118156	1.117025	1.115387	1.118505

3.2 Regression Results and Discussion

Table 5 shows the regression equations associated with each climate zones. Five different regression equations were developed for each climate zone. The R^2, root mean square error (RMSE) and F-Test values are shown in this table. R^2 measures how close the data are to the fitted regression line. As it can be seen, the R^2 value is more than 0.94 in all cases which indicates that the model fits with the data.

Table 5: Regression equations associated with each climate zones

Regression Coefficient: $y = = \beta_1 x_1 + \beta_2\, x_2 + \beta_3\, x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6\, x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9\, x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12}\, x_{12} + \beta_{13} x_{13} + \beta_{14} x_{14} + \beta_{15}\, x_{15} + \beta_{16}\, x_{16} + \beta_{17} x_{17}$

	San Jose	Washington, DC	Houston	Billings	Salt Lake City
β_1	-2.20	-0.68	1.68	-2.10	-0.41
β_2	-1.35	0.70	1.19	-3.52	-0.87
β_3	-2.07	-3.16	0.49	-5.02	1.80
β_4	0.19	-4.14	-1.62	-8.93	-7.37
β_5	1.30	1.02	0.73	0.49	-0.03
β_6	0.04	-0.10	0.31	-0.42	-0.12
β_7	0.57	-3.32	-1.78	-2.94	-2.15
β_8	1.40	1.41	2.56	-2.78	-2.64
β_9	-0.47	1.91	-7.04	-1.57	-2.37
β_{10}	-0.97	-2.83	2.65	2.26	-3.01
β_{11}	0.48	0.42	-0.10	0.21	0.28
β_{12}	-0.40	-0.89	-0.33	-1.33	-1.70
β_{13}	-0.42	0.09	1.30	-0.21	0.44

β_{14}	1.64	-0.01	-0.98	-4.45	1.10
β_{15}	0.43	-1.48	1.51	3.25	0.60
β_{16}	6.29	12.65	9.80	15.66	12.53
β_{17}	-0.31	0.27	0.15	0.37	-0.32
R^2	0.95	0.94	0.95	0.94	0.95
RMSE	226.2	221.3	218.4	-3.31	222.0
F-Test	950	868	918	862	952

x_1 = building orientation, x_2 = top floor batt insulation, x_3 = ceiling interior finish, x_4 = ceiling insulation, x_5 = floor construction, x_6 = top floor ceiling exterior insulation, x_7 = top floor ceiling interior finish, x_8 = ground floor construction, x_9 = ground floor Interior finish, x_{10} = floor interior finish, x_{11} = interior wall, x_{12} = exterior wall, x_{13} = roof absorbance, x_{14} = exterior wall absorbance, x_{15} = roof absorbance, x_{16} = occupant schedule, x_{17} = glass category.

The objective of multiple-linear regression analysis is to predict the single dependent variable (energy consumption) by a set of independent variables (building orientation, wall insulation, glass type, occupancy schedule, etc.). Multiple regression shares all the assumptions of correlation including normality, independence, linearity, and homoscedasticity. Figure 2 shows the scatter plots of residuals which allows visual assessment of the distance of each observation from the fitted line. The residuals from a fitted model are the differences between the responses observed at each combination values of the explanatory variables and the corresponding prediction of the response computed using the regression function. As it can be seen, the residuals are randomly scattered in a constant width band about the zero line and no discernable pattern, without any relationship to the value of the independent variable is observed.

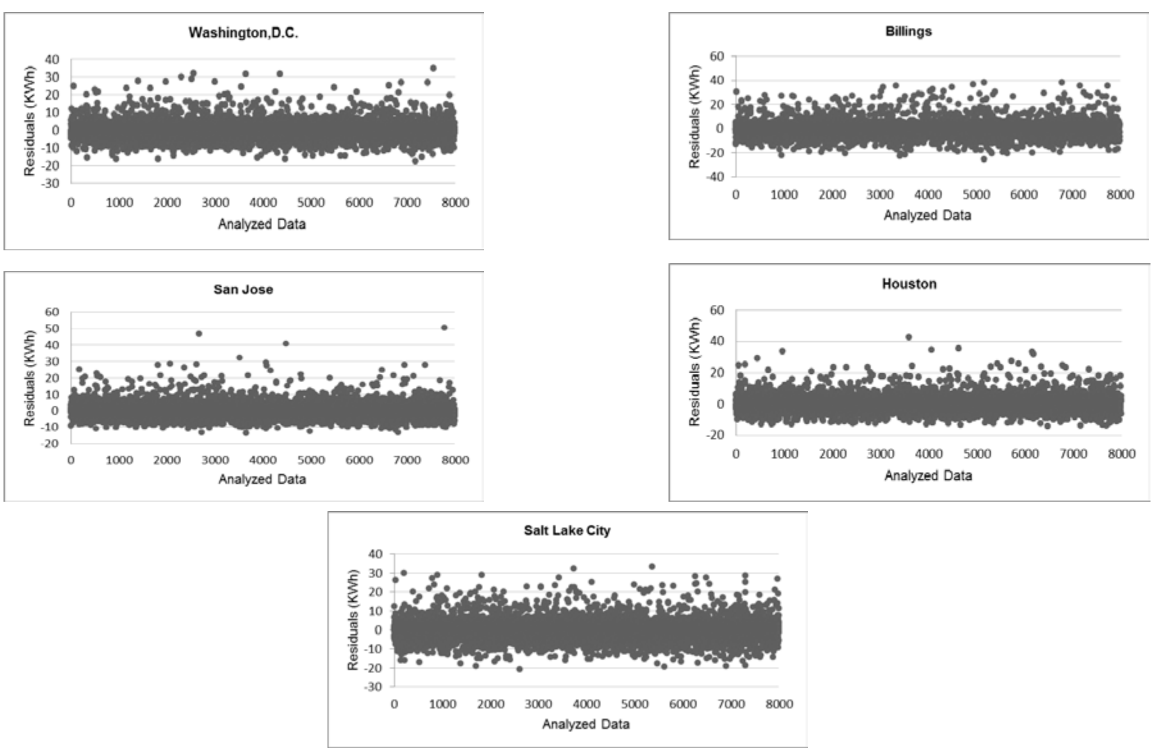


Figure 2: Residual scatter for the five locations.

To demonstrate the variation in energy consumption for each climate zone, identical parameter samples were used for all locations, but the observed range and variability of total energy were unique in different locations (Fig 3). It can be seen that number of outliers is highest in Billings where the first and third quartiles makes up less than half of the range between the minimum and maximum observed. The high variability in Billings is driven by the cold winters and hot summers, which exhibits wide variation depending on the combination of the values for the design parameters.

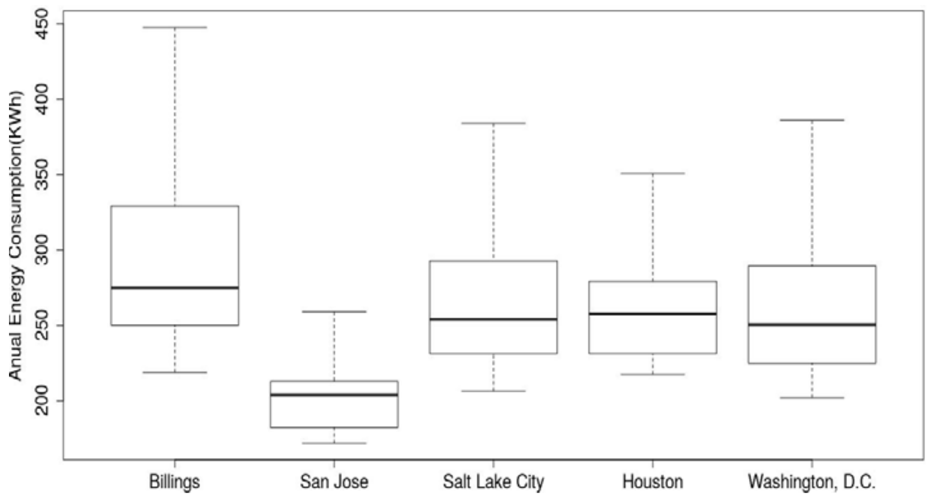
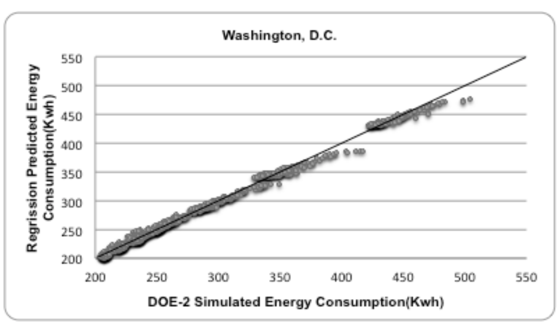
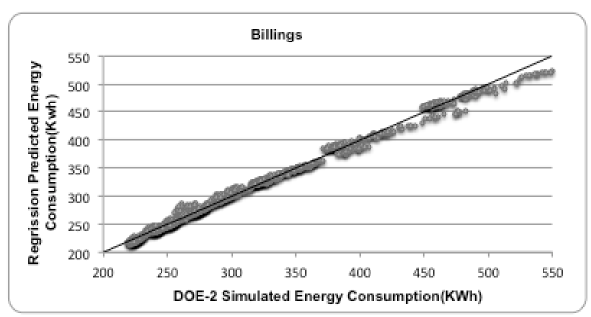


Figure 3: Distribution of total energy consumption for the five locations

3.3 Regression Model Validation

Model validation is one of the most important steps in finding the best fit for the regression model. R^2 and RMSE values are commonly used to validate the model. In this study two thousands of simulations runs were set aside to test the regression model performance and validate the results. Figure 4 shows the validation results for each climate region. It can be observed that the results from the model are well correlated with the data from simulations with acceptable error of less than 5%.





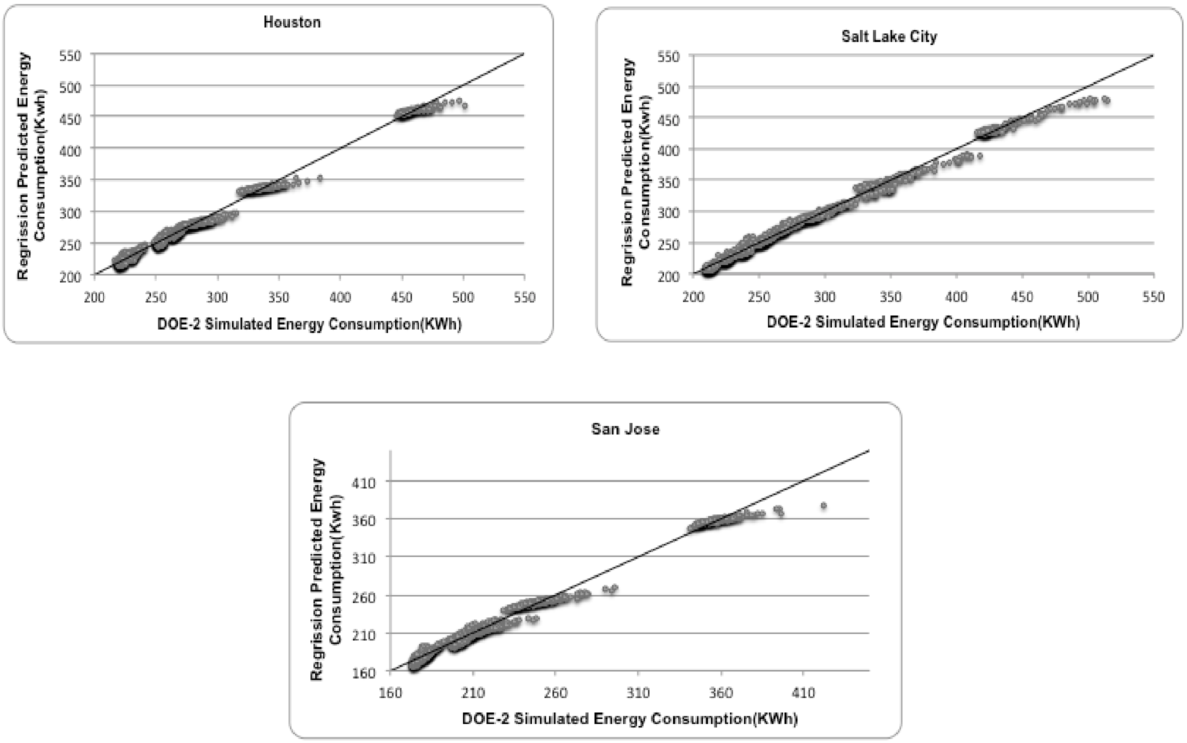


Figure 4: Validation of the total energy consumption models

4 CONCLUSION

The goal of this study was to develop simple regression models for office building in the five major climates including cold dry, cool dry, mixed humid, warm marine and hot humid. A total of 17 key building design variables were identified and considered as inputs in the regression models. The coefficient of determination R^2 varies from 0.94 to 0.95 indicating that 95% of the variation in annual building energy consumption can be explained by change in 17 parameters. The analysis indicates that there is a strong interaction between building location and level of energy consumption. It also shows Billings (cold-dry) with cold winters and hot summers consume the highest amount of energy in comparison with other location. On the other hand, San Jose (warm marine) with the subtropical Mediterranean climate has the least temperature variation and subsequently has the least annual energy consumption. The difference between regression-predicted and DOE-simulated annual building energy use are largely within 5%. Consequently, the developed regression models can be used for comparative energy studies to estimate the potential energy savings during the early stage of design when different building schemes and design concepts are being considered.

Climates with the least temperature variation, subsequently have the least annual energy consumption. The difference between regression-predicted and DOE-simulated annual building energy use are largely within 5%. Consequently, the developed regression models can be used for comparative energy studies to estimate the potential energy savings during the early stage of design when different building schemes and design concepts are being considered.

The models for energy consumption prediction presented in this study, will be expanded in future and will be validated using case studies on physical commercial building to better estimate the prediction accuracy.

References

Asadi., S., Amiri., S.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*, **85**:246-255.

Asadi, S., Hassan, M., and Beheshti, A., 2012. Development and validation of a simple estimating tool to predict heating and cooling energy demand for attics of residential buildings. *Energy and Buildings*, **54**:12-21.

ASHRAE, ASHRAE 90.1-2007, American Society of Heating, Refrigeration, and Air-Conditioning Engineers, Atlanta, GA, 2007.

Broun, R., Babaizadeh, H., Zakersalehi, A., and Menzies, G.F., 2014. Integrated Life Cycle Energy and Greenhouse Gas Analysis of Exterior Wall Systems for Residential Buildings. *Sustainability*, **6** (12):8592-8603.

Catalina, T., Iordache, V., and Caracaleanu, B., 2013. Multiple regression model for fast prediction of the heating energy demand. *Energy and Buildings,* **57**:302-312.

Department of Energy (DOE) EnergyPlus 7,2011.

Department of Energy (DOE), DOE-2., Building Energy Use and Cost Analysis Tool, 2009, Available from: http://doe2.com/DOE2/index.html.

Energy Information Administration (EIA), Annual Energy Review, DOE/EIA03842010, U.S. Energy Information Administration, 2011.

Hygh, J.S., DeCarolis, J.F., Hill, D.B., Ranji Ranjithan, S., 2012, Multivariate regression asan energy

assessment tool in early building design, *Building and Environment*, 57, 165–175.

Lam, J.C., Wan, K., Liu, D., and Tsang., C.L, 2010. Multiple regression models for energy use in air-conditioned office buildings in different climates. *Energy Conversion and Management,* **51**(12): 2692-2697.

Mohammadpour, A., Alanqar, I., Anumba, C., and Messner, J., 2014, Cross Case Energy Simulation Modeling Analysis in Healthcare Facilities Retrofit, *Computing in Civil and Building Engineering Conference, Orlando, FL, USA*.

Repice, R., 2011. Annual Energy Review 2010. DOE/EIA-0384 (2010), US Energy Information Administration, Washington, DC.