Proceedings of BS2013:
13th Conference of International Building Performance Simulation Association, Chambéry, France, August 26-28

A NEW METHOD FOR PREDICTING MIXED-USE BUILDING ENERGY: THE USE OF SIMULATION TO DEVELOP STATISTICAL MODELS

Xuefeng Gao, Ali M. Malkawi, Yun Kyu Yi
T. C. Chan Center for Building Simulation and Energy Studies
University of Pennsylvania, Philadelphia, USA

ABSTRACT

Many building energy prediction models have been developed during the past decades. While popular tools such as Energy Star target single-use buildings, few have focused on mixed-use buildings due to its complexity. In practice, most non-residential buildings are mixed-use buildings supporting various functions such as office, cafeteria, public area, etc. The prediction models developed by Energy Star are based on the building categories defined in Commercial Building Energy Consumption Survey (CBECS), which consider only primary building activities instead of all activity types. This limitation compromises the model's accuracy. This paper aims to tackle this challenge of energy prediction in mixed-use buildings.

By applying simulation and statistical techniques, the proposed method reflects on both modeling and empirical approaches to diminish the difficulty in predicting mixed-use buildings energy consumption. Validation of this method was conducted by comparing the predicted enrgy use to the actual energy consumption data. It was shown that the proposed method was effective for energy prediction in mixed-use buildings. In addition, this approach inherits the complexity and yet efficiency from both simulation and statistical approaches.

INTRODUCTION

Buildings are usually classified based on their primary indoor activities. Many empirical energy prediction methods were developed based on this classification concept. Matson and Piette provided a review of energy studies for commercial buildings (Matson and Piette 2005). Sharp used stepwise regression to model electric energy use per square foot as a function of the CBECS variables, which has been widely applied as an empirical approach in building energy prediction (Sharp 1996 and 1998). The Energy Star building energy prediction models developed by the Environmental Protection Agency (EPA) are regression-based tools which provide prediction models for different building types separately (Energy Star 2011). For example, the Energy Star program has developed different models for offices, K-12 schools, retails, etc. Buildings have to be categorized before it can be used in energy prediction with these methods. However, the process of properly categorizing buildings has its own challenges since there are few cases in which only one activity takes place in a building, especially for the non-residential case.

Kinney and Piette demonstrated this difficulty in defining building categories when conducting energy surveys (Kinney and Piette 2002). For example, Building A may have 30% of its total area as office, 30% as classroom, 20% as lab, and another 20% as miscellaneous use. Thus, this building cannot be clearly categorized into one specific type. Additionally, the number of subdivisions of each building category appeared to have increased over the years. Kinney and Piette argued that much of this diversity in building categories was due to differences in the number of variables recorded and the volume of data collected. Eventually, the same building may fall in different categories under different classification definitions. This could lead to very different energy predictions for the same building when using existing prediction models which were designed for single-use buildings.

Similar to Building A, many commericial buildings are mixed-use, which includes multiple activity zones. Though many empirical energy prediction models are available, few have focused on mixed-use buildings due to the difficulty of categorizing and providing reasonable predictive results for them. The proposed method aims to tackle this challenge in two main steps. The first step is to use simulation modeling to decompose energy use from the whole building level to the zone level. The second step is to apply statistical techniques to predict end-use energy of the activity zones, and then recompose them to retrieve the composite energy of the whole building. The simulation process takes inputs from local typical buildings to best represent local scenarios. The statistical process provides energy prediction models for local buildings as outputs.

Based on Meersche et al.'s research (Meersche et al. 2009), the overall procedure conducted in this paper is shown in Figure 1.

1). Observation (data space): simulation aided sensitivity analysis is conducted to observe the impacts of building features on energy use;

2). Inverse modeling (from data space to model space): the observed energy consumption is used to replicate the prediction model parameters; and

3). Forward modeling (model space): the prediction model is developed based on the relationship between building energy use and model parameters through statistical analysis.

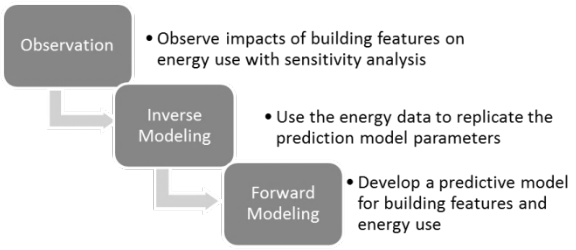

Figure 1 Overall Research Procedure

Given sufficient building samples and feature data, statistical analysis can be conducted on each activity zone and prediction models can be generated based on the relationship between zone energy use and building features. The building level end-use energy can then be obtained through proportional aggregation.

METHODOLOGY

Developing the prediction models for zone level energy use is essential in this research. The detailed procedures are listed in Figure 2. In the first step, a database of local sample buildings with their feature information should be collected and analyzed. The building features should include input parameters for simulation modeling, which fit into four categories: geometry, envelope, internal gains, and HVAC system. In the second step, the simulation model takes inputs from the local building database to best represent local scenarios. The features in each category are manipulated in order to generate sufficient simulation scenarios. Then, the zone level energy use data and feature information are collected.

In the third step, statistical techniques are used to analyze zone level data information and generate prediction models. Sensitivity analysis is used to identify important factors influencing energy use for each activity zone. Correlation analysis is applied to test the null hypothesis that the selected features are uncorrelated (independent). Principal Components Analysis (PCA) is conducted to transform those correlated features into independent components before applying regression analysis. Multiple regression analysis is then performed to analyze the dataset and generate prediction models for zone level energy use.

In the fourth step, the prediction models produce end-use energy for each activity zone. The intermediate energy is calculated by taking the product of the operation time and the power capacity of the pumps and fans. The plant waste, known as heat loss through boilers and heaters, can be determined by plant efficiency. Thus, the building level composite energy can be predicted by aggregation of end-use energy, intermediate energy, and plant waste. Following these steps, the simulation and statistical approaches are integrated to deliver reliable and yet efficient predictive results for local mixed-use buildings energy consumption.

To better understand how these procedures are integrated, assume each activity zone of Building A has a base Energy Use Intensity (EUI) of E kWh/m^2.a and k features (X_1, X_2, X_3, …, X_k) which are identified as the influential factors impacting energy use. Since regression analysis requires independent variables, assume these features are independent from each other. Then, the end-use energy can be calculated as:

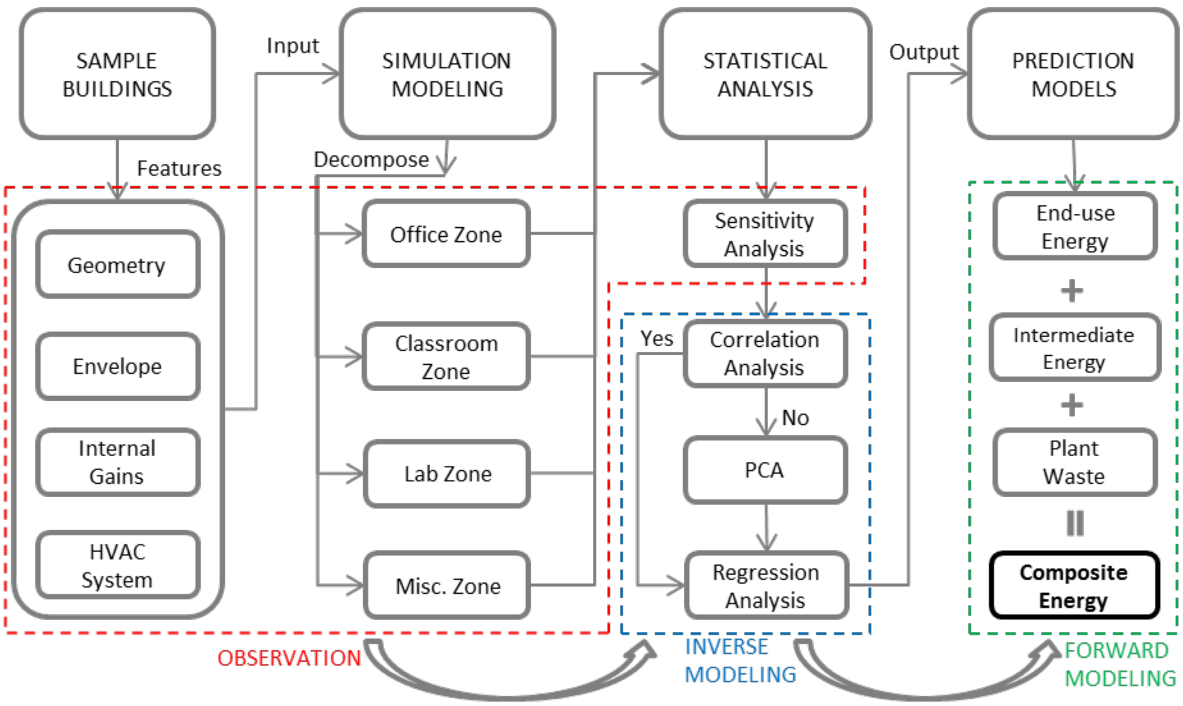

Figure 2 Detailed Procedures of Model Development

$$EUI_{end} = \mathrm{E} + \mathrm{a}_1 \times X_1 + \mathrm{a}_2 \times X_2 + \mathrm{a}_3 \times X_3 + \ldots + \mathrm{a}_k \times X_k \quad (1)$$

Cumulatively, the composite energy use can be calculated as:

$$EUI_{cmpst} = EUI_{end(o)} + EUI_{end(c)} + EUI_{end(l)} + EUI_{end(m)} + EUI_{int} + EUI_{plt} \quad (2)$$

where,

$a_1, a_2, a_3, \ldots, a_k$ = coefficients of the k features,

EUI_{cmpst} = the whole building composite energy use,

$EUI_{end(i=o,c,l,m)}$ = the end-use energy intensity for office, classroom, lab and miscellaneous,

EUI_{int} = intermediate energy use, i.e. energy used by pumps and fans, and

EUI_{plt} = plant waste, i.e. heat loss generated by boilers and coolers.

This calculation assumes the features are independent from each other. However, they may be correlated. Thus, the correlation analysis would be necessary to test whether or not these features are independent. Correlation coefficients are usually used as indicators of correlation between features. Assume matrix **Y** ($n \times k$), where rows (n) represent building observations and columns (k) represent features,

$$\mathbf{Y} = [Y_1\ Y_2\ Y_3\ \ldots\ Y_k] \quad (3)$$

where,

$Y_1, Y_2, Y_3, \ldots Y_k$ are $n \times 1$ vectors, i.e. n building observations for each feature.

The covariance for the ith and jth features $C(i, j)$ is calculated as

$$C(i, j) = \mathrm{cov}(Y_i, Y_j) = \mathrm{E}[(Y_i - \mathrm{E}(Y_i))(Y_j - \mathrm{E}(Y_j))] \quad (4)$$

Then, its covariance matrix **C** can be expressed as

$$\mathbf{C} = \mathrm{E}[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^{\mathrm{T}}] \quad (5)$$

The correlation coefficient for the ith and jth features $R(i, j)$ is thus calculated as

$$R(i,j) = \frac{C(i,j)}{\sqrt{C(i,i)C(j,j)}} \quad (6)$$

The P Hypothesis test can then be conducted to support the hypothesis of no correlation. If $P(i,j)$ is small, for example, less than.05, then the correlation $R(i,j)$ is considered significant and the PCA could be conducted to transform the correlated features into independent components so that the assumption for Equation (1) remains valid. If the correlation is insignificant, PCA would not be required and regression analysis can be conducted as the next step.

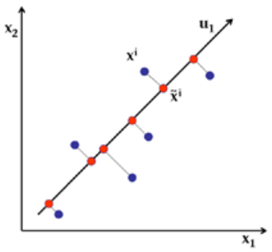

Figure 3 PCA Transformation

PCA Transformation for Correlated Variables

PCA is a mathematical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. As illustrated in Figure 3, PCA projects each building observation x^i (i = 1, …, n) from the features domain ($X_1, X_2, \ldots$) onto the new orthogonal components domain ($u_1, \ldots$). This transformation means that the first principal component accounts for the largest possible variability in the data and each succeeding component, in turn, has the highest variance possible, with the constraint that it be independent from the preceding components.

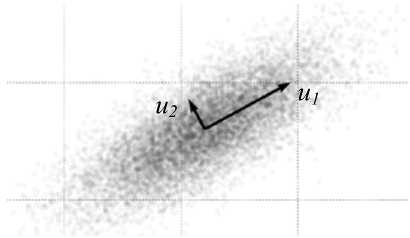

Figure 4 Orthogonal Components from PCA

PCA is mostly used as a tool in exploratory data analysis and for making prediction models. PCA transform correlated variables into independent components so that regression analysis can be conducted to generate prediction models. For regression model $EUI_{end} \sim (X_1, X_2, X_3, \ldots)$ in Equation (1), the first step is to apply PCA to the building database. As illustrated in Figure 4, assume it produces independent components as follows,

$$u_1 = \alpha \times X_1 - \beta \times X_2 + \gamma \times X_3 + \ldots \quad (7)$$

$$u_2 = \alpha' \times X_1 + \beta' \times X_2 - \gamma' \times X_3 + \ldots \quad (8)$$

$$\ldots \ldots$$

where,

$\alpha, \beta, \gamma, \alpha', \beta', \gamma'$ = variables coefficients.

Then, the second step is to conduct regression analysis on the dependent variable EUI_{end} and the independent components produced in the first step. The result may be written in the equation as follows,

$$EUI_{end} = \mathrm{c} + \mathrm{a} \times (\alpha \times X_1 - \beta \times X_2 + \gamma \times X_3 + \ldots) + \mathrm{b} \times (\alpha' \times X_1 + \beta' \times X_2 - \gamma' \times X_3 + \ldots) + \ldots \quad (9)$$

where,

a and b are the component coefficients and c is constant.

The third step is to use mathematical transformation to obtain the best relationship of EUI_{end} and (X_1, X_2, X_3). In this process, the distributive, commutative and associative law would be used to eventually arrive at the following equation:

$$EUI_{end} = \mathrm{c} + (\mathrm{a} \times \alpha + \mathrm{b} \times \alpha') \times X_1 + (\mathrm{b} \times \beta' - \mathrm{a} \times \beta) \times X_2 + (\mathrm{a} \times \gamma - \mathrm{b} \times \gamma') \times X_3 + \ldots \quad (10)$$

Through these three steps, the independence of variables is guaranteed by using PCA, and the prediction models for end-use energy are generated

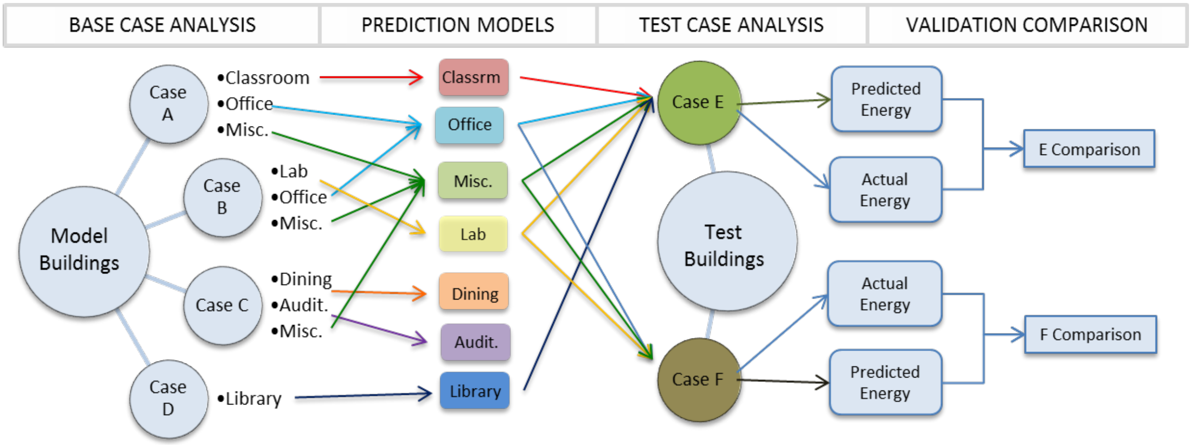

Figure 5 Procedures of Validation Analysis

by applying regression analysis and mathematical transformation.

EXPERIMENT AND VALIDATION

As illustrated in Figure 5, the proposed method was tested with four local mixed-use educational buildings, which cover various activity types. The information of selected building features is listed in Table 1.

Table 1 Feature Information of Four Base Models

#	Activity Zone	Area (m^2)	*H*	*L*	*E*	*P*	*WP*	*ACR*
A	Classroom	6,209	3.7	12	12	0.50	40%	10
	Office	5,825	3.7	12	26	0.07	40%	10
	Misc.	13,342	3.7	12	12	0.07	40%	10
B	Lab	2,343	4.9	18	45	0.05	80%	8
	Office	291	4.9	11	20	0.07	80%	8
	Misc.	1,056	4.9	5	2	0.11	80%	8
C	Dining	1,192	4.3	10	20	0.20	25%	6
	Auditorium	520	12.8	8	2	0.20	25%	6
	Misc.	5220	4.3	5	2	0.11	25%	6
D	Library	17,917	3.7	13	19	0.03	50%	4

where,

H = Floor height, m

L = Lighting density, W/m^2

E = Equipment density, W/m^2

P = People density, persons/m^2

WP = Window percent

ACR = Air Change Ratio, r/h

In order to produce a reasonably large database, these four building models were used as base cases to generate more local scenarios by conducting simulations with manipulated feature values.

The simulation base models were first calibrated with the buildings' actual energy consumption data. The goal of the calibration was to build a robust simulation model that produces well-matched with real energy consumption of the building and is capable to be utilized to make energy performance analysis in this experiment. The four base buildings were modeled with the advanced dynamic energy simulation tool, EnergyPlus, to imitate complex building systems. The results of whole building energy models were validated through the calibration process using a dynamic method to assess the accuracy of the model, i.e. comparing the simulated data with the existing on-site metered data (Yoon et al. 2003).

The calibration method used in the research consists of identifying and making sure that correct input parameters are used in the simulation and the sensitivity analysis is used to improve the accuracy of the simulation model (Pedrini et al. 2002; Westphal et al. 2005). Figure 6 shows the occurrences of changes made during the calibration process of all four base models. As expected, schedule and internal load changes account for 60% of the calibration process. In particular, adjustment of schedules, along with the HVAC settings, shows the highest occurrences in the four base models.

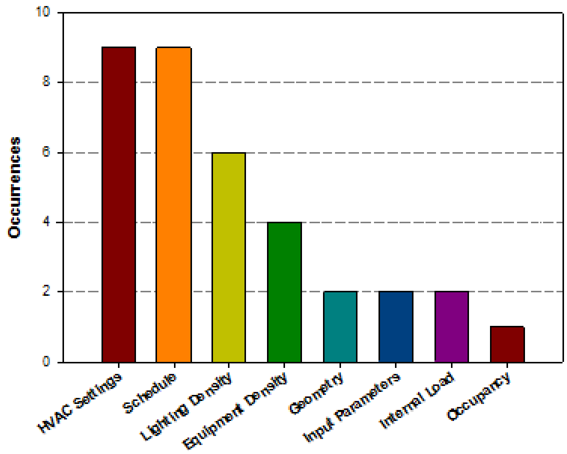

Figure 6 Changes by Category for All Calibration Process

After completing iterations of the calibration process, the simulation models of the four buildings yielded simulation results showing the level of semblance to the actual metered energy consumption data.

From there, 230 simulation variations were generated and 6 features were selected according to their high sensitivity to energy consumption. These six features include height, lighting density, equipment density, people density, window percent and ventilation air change ratio. The simulation variations were generated with manipulated values of four out of the six selected features, lighting, equipment, people density and window percent, as well as an additional common building property, wall U value. The variations of these four base models are summarized in Table 2. For example, five simulation variations were generated based on the lighting density of Model A, with its value varying every 4 W/m^2 from 8 W/m^2 to 24 W/m^2.

Table 2 Variations of Four Base Models

#	*L*	*E*	*P*	*U*	*WP*
A	8 -24	9 -29	0.07 -0.27	0.1 -0.5	0 -80%
B	10 -26	15 -55	0.07 -0.27	0.1 -0.5	10% -90%
C	11 -27	10 -50	0.07 -0.35	0.1 -0.5	0 -80%
D	9 -25	9 -29	0.03 -0.19	0.1 -0.5	50%

where,

U = Wall U-value, W/(m^2×k).

Including all features, Model A has 90 variations, B has 46 variations, C has 75 variations and D has 19 variations. In total, there are 230 simulation variations generated from the four base models.

With these 6 features (k = 6) and 230 variations (n = 230) representing the local scenarios, correlation analysis was conducted. Table 3 lists the calculated correlation coefficients for the 6 selected features used in this paper.

Table 3 Calculated Correlation Coefficients

Coefficients	*H*	*L*	*E*	*P*	*WP*	*ACR*
H	1.0	-0.1	-0.7	0.5	-0.9	-1.0
L	-0.1	1.0	-0.4	0.2	0.3	0.3
E	-0.7	-0.4	1.0	-0.5	0.5	0.6
P	0.5	0.2	-0.5	1.0	-0.3	-0.4
WP	-0.9	0.3	0.5	-0.3	1.0	1.0
ACR	-1.0	0.3	0.6	-0.4	1.0	1.0

Table 4 shows that most features are correlated, since $P(i,j)$ is very small, or less than 0.05.

Table 4 P Hypothesis Test Result

P Test	*H*	*L*	*E*	*P*	*WP*	*ACR*
H	1.00	**0.10**	0.00	0.00	0.00	0.00
L	**0.10**	1.00	0.00	0.01	0.00	0.00
E	0.00	0.00	1.00	0.00	0.00	0.00
P	0.00	0.01	0.00	1.00	0.00	0.00
WP	0.00	0.00	0.00	0.00	1.00	0.00
ACR	0.00	0.00	0.00	0.00	0.00	1.00

Due to the correlation of features, PCA was applied to ensure the transformed variables are independent from each other. Then, multiple regression analysis was used to generate the end-use energy prediction model for each activity zone. The submodels of electricity, heating and cooling energy prediction for the office zone are listed below in Equation (11, 12, 13). The master model of end-use energy prediction for the office zone is shown in Equation (14). Due to space limitation, this paper only shows the master models for zones of classroom, lab, library and auditorium as in Equation (15, 16, 17, 18). These equations all have similar forms as Equation (1).

For office electricity use energy prediction:

$$EUI_{end(o,e)} = 1.75 + 0.48\times L + 0.54\times E - 3.80\times P + 0.10\times ACR \quad (11)$$

For office heating use energy prediction:

$$EUI_{end(o,h)} = 4.43 + 0.44\times L + 3.34\times E - 356.89\times P - 15.48\times WP + 3.66\times ACR \quad (12)$$

For office cooling use energy prediction:

$$EUI_{end(o,c)} = 10.46 + 0.31\times L + 0.07\times E + 51.08\times P + 12.89\times WP + 0.12\times ACR \quad (13)$$

From Equation (11), (12), (13), Equation (1) was resolved. The prediction model for office zone end-use energy thus can be calculated as follows,

$$EUI_{end(o)} = 16.65 + 1.23\times L + 3.95\times E - 309.60\times P - 2.59\times WP + 3.88\times ACR \quad (14)$$

Using the same procedure, the equations for the other zones end-use energy prediction are as follows,

$$EUI_{end(c)} = 92.7 + 0.66\times L - 0.02\times E + 33.81\times P + 3.76\times WP + 2.21\times ACR \quad (15)$$

$$EUI_{end(l)} = 193.1 + 0.44\times L + 1.29\times E - 21.7\times P + 4.82\times WP \quad (16)$$

$$EUI_{end(lib)} = 70.2 + 0.11\times L + 0.12\times E + 33.27\times P \quad (17)$$

$$EUI_{end(a)} = 27.0 + 0.72\times L - 2.15\times P + 15.88\times WP \quad (18)$$

With these prediction models available for different zones, the composite energy as calculated in Equation (2) could be obtained. The intermediate energy EUI_{int} could be calculated as the product of operation time and power capacity of the fans and pumps. The plant waste EUI_{plt} can be determined by plant efficiency.

The result validation procedure was conducted on two test buildings with surveyed data, as illustrated in Figure 5. The test buildings are also local mixed-use educational buildings. Table 5 contains the required inputs of these two test cases for the prediction models.

Table 5 Inputs of Test Buildings

#	**Activity Zone**	**Area (m^2)**	**L**	**E**	**P**	**WP**	**ACR**
E	Lab	165	19	28	0.07	13%	16
	Library	369	12	12	0.03	13%	5.4
	Classroom	613	12	12	0.03	13%	5.4
	Office	6,652	12	14	0.07	13%	5.4
	Misc.	3,584	5	5	0.03	13%	2.4
F	Office	416	14	11	0.07	0	4
	Lab	3,304	14	117	0.07	0	10
	Misc.	5,756	5	5	0.03	0	4

By applying the prediction models developed above, the predicted composite energy uses of two test buildings were calculated and compared to their actual data. Table 6 shows the result.

Table 6 Validation Result of Two Test Cases

#	Actual Source Energy (kWh/m^2)	Predicted Source Energy (kWh/m^2)	Difference
E	806	949	18%
F	1,797	2,017	12%

Both buildings, particularly Building F, have high source energy use intensity. This is reasonable since the equipment density of a lab zone is usually much higher than other activity zones due to its special utility characteristics. The difference between predicted and actual energy uses of the two buildings are within twenty percent, which indicates a fairly good prediction. While considering the nature of uncertainty in mixed-use buildings, prediction models generated from the proposed method are considered effective in providing reasonably predictive results. Though uncertainty is inevitable in mixed-use building energy prediction, for future work, researchers can enhance the accuracy of prediction models by further analyzing the diversity of other activity zones.

CONCLUSION

This paper focused on developing a new method for mixed-use building energy prediction. By applying simulation and statistical techniques, the proposed method reflects on both modeling and empirical approaches to diminish the difficulty in predicting mixed-use building energy consumption. The simulation process takes inputs from local typical buildings to best represent local scenarios. The statistical process provides energy prediction models for local buildings as outputs. The proposed method inherits the complexity from simulation modeling and the efficiency from statistical analysis.

Four educational buildings have been used as base models to generate a database of local scenarios. Sensitivity analysis was applied to understand the impacts of building features on energy use. Correlation analysis, PCA transformation and multiple regression analysis were conducted to generate prediction models for various activity zones. Two local buildings were used as test cases to examine the accuracy of the prediction models. In comparison to the actual energy consumption data, the result demonstrated that the prediction models generted from the proposed method were able to provide reasonable energy prediction for mixed-use buildings.

NOMENCLATURE

E = base energy use intensity of building A

$X_1, X_2, X_3, \ldots, X_k$ = k features identified as influential factors on energy use

$a_1, a_2, a_3, \ldots, a_k$ = coefficients of the k features

EUI_{end} = the end-use energy use intensity

EUI_{cmpst} = the composite energy use intensity

$EUI_{end(i=o,c,l,lib,m)}$ = the end-use energy intensity for office, classroom, lab, library and miscellaneous

EUI_{int} = intermediate energy use intensity, i.e. energy used by pumps and fans

EUI_{plt} = plant waste, i.e. heat loss generated by boilers and coolers

n = number of building observations

k = number of features identified as influential factors on energy use

$Y_1, Y_2, Y_3, \ldots Y_k$ are $n\times1$ vectors, i.e. n building observations for each feature

$C(i,j)$ is the covariance for the ith and jth features

$R(i,j)$ is the correlation coefficient for the ith and jth features

$P(i,j)$ is the P Hypothesis test value for the ith and jth features

$\alpha, \beta, \gamma, \alpha', \beta', \gamma'$ = variables coefficients

a, b = components coefficients

c = constant

H = Floor height, m

L = Lighting density, W/m^2

E = Equipment density, W/m^2

P = People density, persons/m^2

WP = Window percent

ACR = Air Change Ratio, r/h

U = Wall U-value, W/(m$^2\times$k)

ACKNOWLEDGEMENT

The authors would like to thank the Facilities & Real Estate Services at University of Pennsylvania for making available the data.

REFERENCES

Kinney, S., and M. A. Piette. 2002. Development of a California commercial building benchmarking database.

Matson, N. E., and M. A. Piette. 2005. Review of California and national methods for energy performance benchmarking of commercial buildings.

Meersche, K. V. d., K. Soetaert, and D. V. Oevelen. 2009. xsample(): an R Function for Sampling Linear Inverse Problems. Journal of Statistical Software, Code Snippets, 30(1), 1-15.

Pedrini, A., F. S. Westphal, and R. Lamberts. 2002. A methodology for building energy modeling and calibration in warm climates. *Building and Environment* 37: 903-912.

Sharp, T. 1996. Energy benchmarking in commercial office buildings. Paper presented at Proceedings of the ACEEE 1996 Summer Study on Energy Efficiency in Buildings.

Sharp, T. 1998. Benchmarking energy use in schools. Paper presented at Proceedings of the ACEEE

1998 Summer Study on Energy Efficiency in Buildings.

Westphal, F. S., and R. Lamberts. 2005. Building simulation calibration using sensitivity analysis. *Ninth International IBPSA Conference, August*.

Yoon, J. H., E. J. Lee, and D. E. Claridge. 2003. Calibration procedure for energy performance simulation of a commercial building. *Journal of Solar Energy Engineering* 125 (3): 251-257.

Energy star performance ratings – technical methodology. 2011. Environmental Protection Agency.