BUILDING PERFORMANCE METRICS CALCULATION AND VISUALIZATION FOR K-12 SCHOOL BUILDINGS IN NEW YORK CITY

Young M. Lee[1], Fei Liu[1], Estepan Meliksetian[1], Jane Snowdon[1] and Michael Bobker[2]
[1]IBM T.J. Watson Research Center, Yorktown Heights, New York 10598, U.S.A.
[2]Building Performance Lab, CUNY Institute for Urban Systems, NY, NY 10010, U.S.A.

ABSTRACT

An integrated statistical method for computing energy performance metrics of buildings and visualizing the energy performance through spatial analysis is described. This technique combines the hierarchical multivariate regression models and the Variable Base Degree Day (VBDD) model to compute the overall energy performance indicator, and the energy performance indices of base load, heating load and cooling load for each building in a portfolio. Then the various energy performance metrics in the portfolio are selectively visualized on the associated Geographical Information System (GIS) based heat map. The method provides useful information for accurately assessing energy performance, benchmarking and identifying energy performance improvement opportunities in a large portfolio of buildings. The method has been successfully applied and deployed for K-12 school buildings in New York City.

INTRODUCTION

In the United States, 40% of the nation's total energy consumption occurs in commercial and residential buildings (U.S. Dept of Energy, 2006). The buildings also contribute to 45% of the country's greenhouse gas (GHG) emissions, which are linked directly to global climate change (EPA ESPM Greenhouse Gas Inventory 2009). Global warming, caused by gradually increasing concentrations of GHG in the atmosphere, is one of the most significant crises that human beings face today. Already, it affects our lives, for even small changes in the Earth's average temperature can increase the occurrences of severe weather conditions such as storms, floods and droughts, stimulate drastic climatic changes in all ecosystems, and initiate adverse impact on the health and lives of humans and other species. The majority of greenhouse gases are emitted when fossil fuels are burned to produce heat and electricity. The majority of the world's population either lives or works in a building; therefore, everybody has a responsibility and a role to play in reducing energy consumption, controlling GHG emissions, and confronting climate changes and its potential impacts. End-use energy efficiency can contribute to more than 50% of total global energy conservation (World Energy Outlook 2009).

Saving energy, improving energy efficiency and reducing greenhouse gas (GHG) emissions are key initiatives in many cities and municipalities and for building owners and operators. For example, New York City (NYC)'s government spends over $1 billion a year on energy on their approximately 4,000 buildings (e.g. public schools, prisons, court houses, administrative buildings, waste water treatment plants, etc.), and is committed to reducing the City government's energy consumption and CO2 emissions by 30% by 2030 from the 2005 level through an initiative called PlaNYC (PlaNYC 2007). NYC plans to invest, each year, an amount equal to 10% of its energy expenses in energy-saving measures over the next 10 years. The largest segment of NYC government buildings are the 1,200 K-12 public schools serving 1.1 million students and covering about 150 million square feet. The NYC public school system is one of the largest public school systems in the world. The NYC Department of Education is interested in understanding how energy efficient their buildings are, what factors contribute to inefficiencies, what are the opportunities for improvements given budget constraints, and how much they can contribute to saving energy and reducing GHG emissions toward NYC's PlaNYC initiative. Developed along the initiatives is the IBM Building Energy and Emission analytics (i-BEE™) Toolset (Lee et al. 2011), a new analytical tool which assesses, benchmarks, diagnoses, tracks, forecasts, simulates and optimizes energy consumption in building portfolios. Our focus in this paper is to introduce the computational method of energy performance metrics of buildings, along with the visualization and spatial analysis capability.

We develop a multi-step statistical analysis procedure, which combines the hierarchical multivariate regression models and the Variable Base Degree Day (VBDD) regression model (Kissock et al., 2003) to compute energy performance indices. The energy performance indices are further visualized through a GIS based heat map, which provides valuable insights for benchmarking and identifying buildings that have energy saving opportunities for large portfolios of buildings. In this approach, we first build a multivariate regression model which correlates the overall energy consumption with building characteristics. The energy related building characteristics are identified through the stepwise variable selection technique. The results are valuable in providing building energy performance scores for each building in the whole portfolio. Next, in order to accommodate building heterogeneity, we build VBDD regression models separately for each building. These models are used to separate the base load energy consumption from the weather dependent usage such as heating load and cooling load. The base load, being separated by the statistical model, typically includes energy consumption related to lighting, which is one of the largest source of consumption in commercial building. Other typical base load usage includes hot water, cooking and other plug loads such as appliances and computers. The results in this step consist of estimating the balance temperature, base load consumption, as well as calculating coefficients for HDD (Heating Degree Day) and CDD (Cooling Degree Day) for all buildings. Finally, by generating multivariate regression models of the resultant coefficients from the VBDD model, we further breakdown the total energy use of each building into energy load types (e.g., heating, cooling and base load), leading to the respective performance scores for heating, cooling and base load energy consumption.

The proposed technique provides an integrated statistical analysis for building energy performance metrics calculation and visualization for building portfolios. The overall energy performance index for each building with respect to other buildings in the portfolio is displayed using the probability distribution curve, and can be compared with other peer buildings in the portfolio. The energy performance indices for each energy load type (e.g., heating, cooling and base load) are also displayed using the corresponding probability distribution curves. The energy performance indices are also visualized on the GIS based heat maps, where the varying shades of color represent the relative ranking of the performance indices. This GIS heat map capability allows building managers to geographically characteristize the energy performance of buildings in a portfolio, and screen out potential buildings that may be good candidates for energy improvement projects such as retrofits. In the remainder of this paper, we will first describe the statistical approach, followed by the visualization techniques for K-12 school buildings in the NYC.

STATISTICAL MODELING

We develop a multi-step statistical analysis procedure, which combines the multivariate regression models in multiple hierarchy and the VBDD regression model to assess energy use and identify energy efficiency improvement opportunities for large portfolios of buildings.

In the first step, we build a regression model which correlates the energy consumption with building characteristics. The energy related building characteristics are then identified through the stepwise variable selection technique. The results are valuable in providing building energy performance scores for the whole portfolio. Additionally, the energy consumption of buildings with similar characteristics can be predicted. For large building portfolios like the NYC K-12 public school system, the amount of data is usually huge. With the energy data being collected cumulatively over time for each building, the capability to quickly access the energy performance for the whole portfolio is very important for successful energy management. The U.S. Environmental Protection Agency (EPA) Energy Star Performance Rating (EPA 2009) provides a valuable management tool to assess the building energy efficiency, relative to similar buildings and climate zones nationwide. In addition to the nationwide comparison, the assessment of the energy efficiency of a building, relative to its peer buildings within the portfolio, which provides local and more detailed information, is also of great value for the building managers. This calls for the development of a local performance indicator, solely based on the data within the portfolio. In this section, we describe the development of multi-step, multivariate regression models, which lead to a building performance indicator for local portfolio assessment.

We first describe the general method utilized to derive the building performance indicator. Let y_i be the quantity of our interest (in this case energy consumption of a building), typically referred to as the response variable, and $x_{i1},\ldots,x_{ip}$ be the p predictor variables. The multivariate regression model then takes the form

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ip}\beta_p + \varepsilon_i, \quad (1)$$

where $\beta_1,\ldots\beta_p$ are the regression coefficients and ε_i is the error term. As in EPA (2009), we utilize the multivariate regression model in Equation 1 to normalize the response variable according to the predictor variables. In the building portfolio context, the predictor variables consist of the building characteristics and operational activities, for example, gross floor area (GFA), number of windows, number of computers, and number of operating hours, and so on. Some of the predictors are not directly energy related; therefore, are not statistically significant. We thus perform a stepwise variable selection procedure, which removes the redundant variables from the model.

The multivariate regression model in Equation 1 also serves as a reference to the energy use for the general population in the building portfolio. As a result, while fitting the model, we remove data that cannot represent the general population (e.g., outliers). In practice, a data point is identified as an outlier and therefore is removed from the subsequent analysis if its absolute value of the standardized residual is larger than 2. After removing the outliers, we fit the model again and use the resultant estimates for further analysis.

The expected value of the response variable, for a specific building with given characteristics, can be immediately calculated as

$$\hat{y}_i = E(y_i \mid x_{i1},\ldots,x_{ip}) = x_{i1}\hat{\beta}_1 + x_{i2}\hat{\beta}_2 + \ldots + x_{ip}\hat{\beta}_p. \tag{2}$$

The standardized residual, $\hat{z}_i$, can be then calculated as

$$\hat{z}_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}},$$

where $\hat{\sigma}$ is the standard error. We note that for a portfolio with many buildings, the $\hat{z}_i$ approximately follows the standard normal distribution. Matching z_i to the standard normal curve, we can calculate the probability that buildings with the same characteristics consume more energy than building i as

$$P(Z > \hat{z}_i) = 1 - \Phi(\hat{z}_i),$$

where Z is the standard normal random variable and $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. We further define the performance indicator for building i as

$$PI_i = 100 \times P(Z > \hat{z}_i) = 100 - 100\Phi(\hat{z}_i). \tag{3}$$

The resulting performance indicator is a number between 0 and 100, with larger values indicating better energy efficiency. In addition, the interpretation of the performance indicator is very intuitive. By definition, the performance indicator is the percentage of similar buildings that consume more energy than building i in the portfolio. The performance indicator is graphically displayed to users using the probability distribution curve according to Equation 3 as shown in Figure 1. For instance, the building of interest has the performance indicator of 56, implying that the relative performance of this building is about in the middle among other buildings in the portfolio. The performance indicator can also be compared with peer buildings in the portfolio using selection drop down menus as shown the bottom half in Figure 1. In this example, the user compares the energy performance scores by picking the top 10 performance peer buildings that are built between 1900 and 1960, have over 100,000 square feet of GFA, are high schools, and are located in a specified NYC borough.

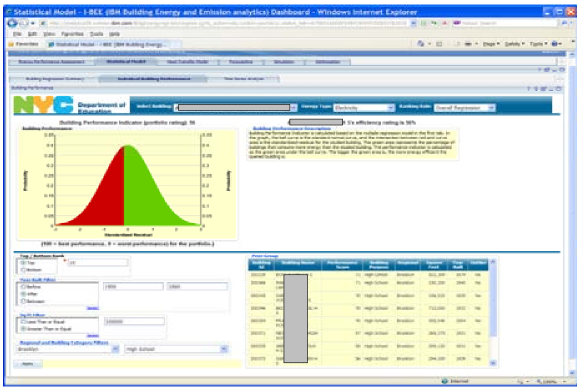

Figure 1 Overall Performance Indicator

In the second step, to accommodate building heterogeneity, we build the VBDD regression models separately for each building. For the VBDD model, the total monthly energy usage data for building i, in period t is denoted y_{it}, and is modeled as:

$$y_{it} = b_i + c_i CDD_t(T_i^{(b)}) + h_i HDD_t(T_i^{(b)}) + \varepsilon_{it} \tag{4}$$

where b_i is the base load usage, c_i is the cooling coefficient, h_i is the heating coefficient, and ε_{it} are the error terms reflecting the month-to-month variations that can not be explained by base load, heating or cooling usage. We further restrict that $b_i > 0$, $c_i > 0$

and $h_i > 0$. The heating degree day (HDD) and the cooling degree day (CDD) for building i in month t are defined as

$$HDD_t(T_i^{(b)}) = \sum_{d=1}^{d_t}(T_i^{(b)} - T_{itd})^+ \text{, and}$$

$$CDD_t(T_i^{(b)}) = \sum_{d=1}^{d_t}(T_{itd} - T_i^{(b)})^+ .$$

Here, T_{itd} is the outdoor temperature for building i on day d of month t, $i \in \{1,\ldots,n\}$, $t \in \{1,\ldots,m\}$, $d \in \{1,\ldots,d_t\}$, and $T_i^{(b)}$ is the balance-point temperature for building *i*. $HDD_t(T_i^{(b)})$ and $CDD_t(T_i^{(b)})$ are the cumulative heating and cooling energy usage for month t when the balance point temperature is $T_i^{(b)}$. These models are used to separate the base load energy consumption from the weather dependent usage. The results in this step consist of the base temperature estimates, base load estimates, as well as the estimated coefficients for HDD and CDD for all buildings.

In the third step, we further conduct root cause analysis by building the multivariate regression models for the results from the VBDD model, from which the performance scores can be derived for base load energy consuption and energy consumption for heating and cooling. The multivariate regression model in the level takes the form

$$b_i = x_{i1}^b \beta_1^b + x_{i2}^b \beta_2^b + \ldots + x_{ip}^b \beta_p^b + \varepsilon_i^{b,p}$$
$$c_i = x_{i1}^c \beta_1^c + x_{i2}^c \beta_2^c + \ldots + x_{ip}^c \beta_p^c + \varepsilon_i^{c,p}$$

$$c_i = x_{i1}^h \beta_1^h + x_{i2}^h \beta_2^h + \ldots + x_{ip}^h \beta_p^h + \varepsilon_i^{h,p}$$

where b_i, c_i, h_i are the base load usage, cooling coefficient, heating coefficient, i.e., the response variables, and ε_{it} are the error terms reflecting the month-to-month variations that can not be explained by base, heating or cooling usage. $x_{i1},\ldots,x_{ip}$ are the p predictor variables, and $\beta_1,\ldots\beta_p$ are the regression coefficients.

The performance indicators for base load, heating load and cooling load are calculated using Equation 1 and 3 as shown earlier. The performance scores are also displayed graphically using cumulative distribution curve defined in Equation 3 as shown in Figure 2 below. In this example, a building in the portfolio has fairly good performance indicator for heating (i.e., 73), but medium performances for cooling (performance score 43) and base load (performance score 50), respectively.

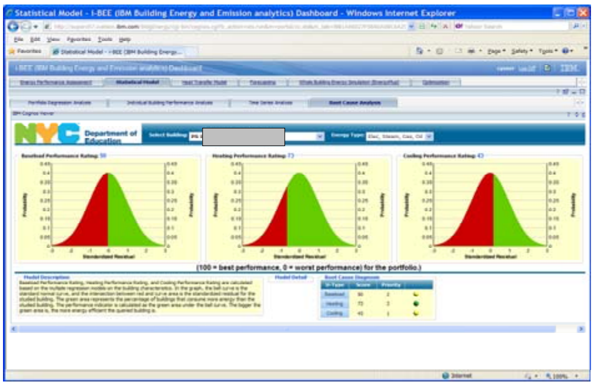

Figure 2 Performance Indicators for Base Load, Heating Load and Cooling Load

The multi-step statistical analysis procedure is described by the schematic given in Figure 3. The details of the technique are explained in another paper (Liu et al. 2011).

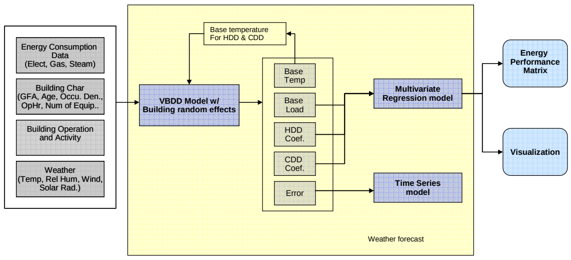

Figure 3 Integrated Statistical Approaches

VISUALIZATION OF ENERGY PERFORMANCE

The computed energy performance indicators explained in the earlier section are visualized graphically in a portfolio level so that building managers, operators and other related personnel can easily see the performance of buildings in a portfolio and their geographical distribution, select a subset of the buildings and make appropriate decisions. As shown in Figure 4, energy performance is displayed using a Graphical Information System (GIS) based heat map.

Each building in a portfolio is represented by a circle in a geographical map. The size of a circle typically represents the GFA (gross floor area) of the building. However, it can represent other characteristics of buildings if needed. The color of a circle represents the relative performance of a certain energy usage of a building. In Figure 4, we use the colors to represent the overall energy performance indicator (Equation 3) for all the energy types together (including electricity, natural gas, steam and fuel oil). The green color indicates the high performance indicators, the yellow color indicates the medium performance, the red color represent low performance, and color gradients between red, yellow and green represent performance scores between them (see Figure 5 for color scheme). We call this visualization a "heat map". Other performance indicators such as the ones for different energy type, load type, energy intensity (site and source) and even greenhouse gas emission can be represented by various colors at user's command.

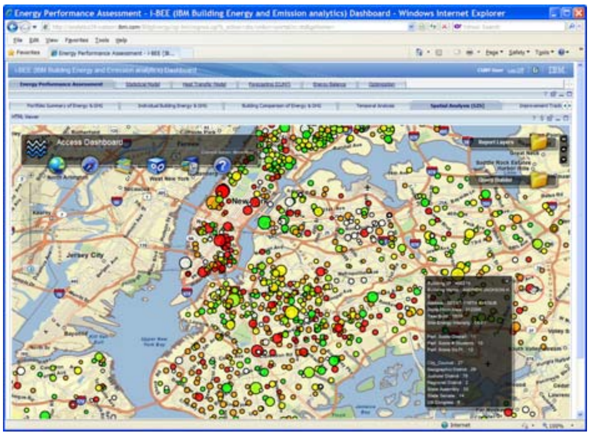

Figure 4 Performance Indicators (Overall Performance)

The heat map can be generated for a subset of buildings in the portfolio. For instance, buildings built in certain time periods, within a certain size, and in a certain performance range can be selectively filtered using the graphical user interface shown in Figure 5 and displayed in the GIS based heat map.

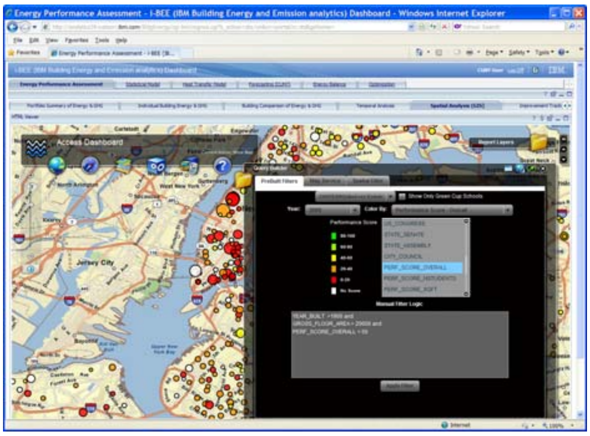

Figure 5 Heat Map Setting and Filtering

Figure 6 shows building energy performance heat map for electricity consumption while Figure 4 is for all types of energy together. There seems to be significant energy performance difference between overall energy, i.e.,electricity, steam, natural gas and fuel oil (Figure 4) and that of electricity only (Figure 6). Energy performance indicators for usage levels i.e., for base load, heating load and cooling load can also be displayed in the heat map.

In addition to energy performance indicators calculated by the multi-level statistical modeling described earlier, simpler and more common energy performance indicators such as energy intensity, which is energy use per unit GFA, can also be displayed for site energy consumption (site energy intensity) and source energy consumption (source energy intensity). Energy use per tenant can also be displayed. Figure 7 shows a heat map for site energy intensity. It is interesting to see that the heat map for overall energy performance indicators (Figure 4) looks quite different from the simple energy intensity heat map (Figure 7).

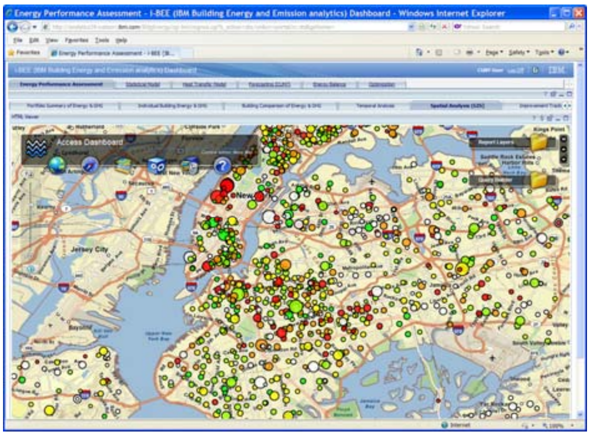

Figure 6 Performance Indicators (Electricity Performance)

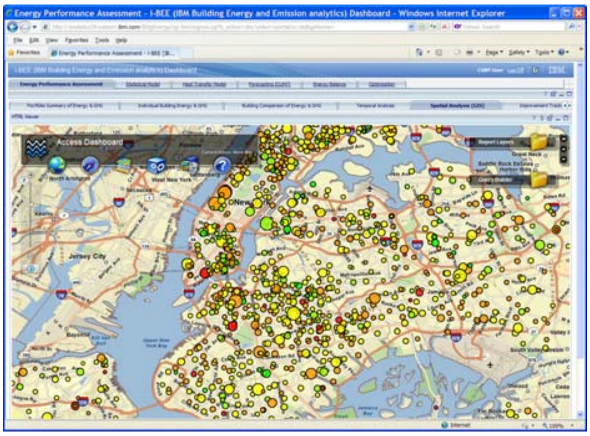

Figure 7 Performance Indicators (Site Energy Intensity)

The GIS based heat map visualization tool can also allow users to identify a set of buildings in a portfolio that are candidates for energy efficiency improvement actions such as retrofits. The heat map in Figure 8 shows a sample set of buildings that have poor energy performance indicators, thus are good candidates for consideration in energy efficiency improvement programs.

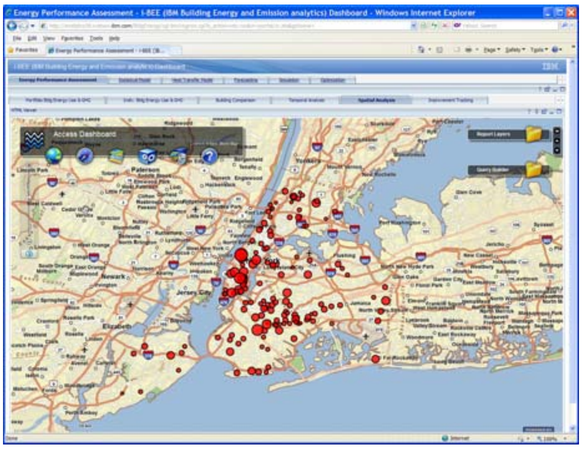

Figure 8 Heat Map Filtering Tool

CONCLUSION

This paper described a multi-step statistical procedure for computing energy performance indicators and visualizing the energy performance in statistical graphs and GIS based heat map for a large portfolio of buildings. The tool was deployed for the K-12 public school buildings in the New York City. The tool is being used for NYC building operator certification (BOC) training classes (Bobker et al. 2011) and has received very positve feedback from both the instructors and the students. We plan to expand the use of the tool to more public school systems in the U.S. and other public building portfolios.

REFERENCES

Bobker, M., Belfast, J., Leifer D., Snowdon J., Lee, Y., Dail, P., Weisner P., Energy-data Dashboard: Designing for Use in NYC School, 2011. ICEBO 2011.

EPA 2009. Energy Perforemance ratings: Technical Methodology. United States Enviromental Protection Agency.

EPA ESPM (EnergyStar Portfolio Manager) 2009. Greenhouse Gas Inventory and Tracking in Portfolio Manager, Accessed February 2010. http://www.energystar.gov/ia/business/evaluate_performance/Emissions_Supporting_Doc.pdf.

Kissock, J. K., Haberl, J., and Claridge, D. E. 2003. Inverse modeling toolkit (1050rp): Numerical Algo-rithms. ASHRAE Transactions, 109:425–434.

Lee, Y.M, et al. 2011. Proceedings of the 2011 Winter Simulation Conference. Jain, S., Creasey, R.R., Himmelspach, J., White, K.P., and Fu, M., eds.

Liu, F., Lee, Y.M., Jiang, H., Snowdon, J., Bobker M., 2011. Proceedings of Building Simulation 2011, 12th Conference of International Building Perfromance Simulation Association, Sydney, 2011.

PLANYC 2007. PlaNYC 2030: A greener, greater New York. http://www.nyc.gov/html/planyc2030/html/home/home.shtml.

U.S. Department of Energy. Buildings Energy Data Book. 2006. Accessed September, 2010. http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=1.1.1.

World Energy Outlook 2009. Accessed September 2010.