ENERGY MODEL CALIBRATION FOR CAMPUS OFFICE BUILDINGS

Bo Lin[1], Zhao Chen[2]
[1]SmithGroup, Washington, DC
[2]Fudan University, Shanghai, China

ABSTRACT

In order to achieve energy reduction, many buildings implement energy conservation measures (ECMs). Building energy model (BEM) is a useful tool for quantitative analysis of ECMs. However, uncalibrated model will generate misleading results. The BEM has to be calibrated to the as-operated condition based on metered data and building information. In this research, we applied variable selection method to prioritize calibration sequence and strategy. Critical variables that are significant to building energy use are identified through big data learning approaches. In the case study, we applied the methodology to campus office buildings in the U.S.

INTRODUCTION

Buildings energy consumption is one of top national economic and environmental focuses, for example in the U.S, the buildings consume 40% of total energy use (DOE, 2011). According to EIA, U.S building sector is very old (EIA, 2017). Only 12% of commercial buildings were built since 2003, more than half of all buildings were built before 1980. The median building age was 32 years old (EIA, 2017). With growing awareness of decreasing energy use in buildings, retrofit and upgrade existing building subsystems have emerged as a critical research field especially for commercial buildings.

Office building is the largest component in commercial building sector with around 17% floor space is office building (EIA, 2017). In this paper, we focused on existing office buildings in the northeast region in the U.S as our case study buildings. The median age in the northeast was 46 years while the counterpart in the south is only 29 years old (EIA, 2017). The northeast region also has the largest commercial building size. The northeastern building average size is 4,000 to 5,000 square feet larger than other regions (EIA, 2017). This is mainly because the northeastern cities have longer history and larger population than others.

Building energy model (BEM) is a versatile, multipurpose analytical tool that calculates energy flows and thermal comforts of building. For existing building analysis, BEM is an important tool to evaluate various energy conservation measures (ECMs) for energy reduction. It takes building's descriptive inputs such as ambient weather condition, geometry model, envelop construction data, interior plug and lighting data, HVAC systems, control strategies and occupants' behavior to predict building full year energy consumption and thermal comfort.

The accurate prediction is not easy, especially for commercial buildings. This is because energy use is an outcome of interactions among mechanical equipment, system controls, ambient conditions, human behavior and so on. The BEM requires many inputs and profound knowledge about the building and system. The error or inaccuracy of BEM is considered as the difference between simulated result and the actual energy data of the building. It is reported that an uncalibrated model gives 7% to 75% of difference from actual metered data (Ahmad et al. 2006). This discrepancy can result in massive energy waste and financial lose or maybe legal case. Therefore, the energy model developer must make sure the model is calibrated to the as-built condition of the building.

Sensitivity analysis is widely used for energy model calibration. For example, in (Westphal and Lamberts. 2005), they developed 6-step methodology to calibrate the model. The first step is to fine tune the near-constant loads such as lighting and plug. Then run the BEM at design days to get the thermal load data. Perform the sensitivity analysis on variables that are significantly related to the heat gains and losses. Adjust the input values for those variables that are found important in the sensitivity analysis and run the full-year simulation. It should be noted that the sensitivity analysis is only performed for variables at load calculation stage and can have different results at annual simulation. (Song et al. 2008) calibrated their model in a 6-story building with sub-metered data. The result showed that their model was well calibrated. (O'Donnell 2009) concluded that the fundamental problem for energy model calibration is lack of comprehensive, consistent measured data and information to develop the model. Raftery et al. (2009) mentioned that the main issue for model calibration is lack of well accepted calibration methodology and insufficient metered data.

ASHRAE guideline 14 is widely used to help calibration. The guideline clearly defines statistic indexes as threshold for calibrated model. From the guideline, a model should have normalized mean bias error (NMBE) at 5% or lower and coefficient of variation of root mean square error (CVRMSE) of 15% or lower if actual monthly data is used. For hourly data calibration, those two indexes are 10% and 30%

The rest of the paper is organized as follows. In the second section, we introduce two high dimensional big data variable selection methods. Both methods can help us find critical variables to building energy use thus guide the sequence of calibration. In the third section, two methods are applied to an office building campus energy model. In the last section, we summarized the findings and conclusions.

RESEARCH METHODOLOGY

Traditionally, the calibration process is iterative and labor intensive, such as sensitivity analysis may need to run several thousand of simulations to complete computation. In this research, we proposed high dimensional big data analysis approaches to identify the critical variables that are important to the building energy use. Those selected variables are prioritized in the calibration and simulated to match the actual data. The proposed two variable selection methods are called "least absolute shrinkage and selection operator (LASSO)" and "smoothly clipped absolute deviation penalty (SCAD)".

In practice, before the calibration process, engineers will collect building information and monitor various data. For example, various energy or flow meters are installed on the building to collect data such as lighting energy use. Typically, the metered data are grouped into weekdays and weekends due to the distinct difference between weekday and weekend operation. After the data collection is finished, the model calibration utilizes collected information and data to develop as-built energy model. However, there are some constrains in this procedure.

As aforementioned, most of the buildings in the U.S are very old. The original building design parameters' might differ significantly from the drawings or design values because of degradation or damage. Secondly, in practice, it is very common to see meter reading errors for example enormously large reading values, data gaps or overlaps. The metering system upgrades, power outrage and so on are typical causes. If the energy model is calibrated based on the erroneous metered data or out-of-date design values, the accuracy is not guaranteed. On the other hand, the building can have more than a thousand variables, it is very time consuming to calibrate every variable in the actual project. Therefore, we recommend 3 steps calibration in general. The first step is to develop the model geometry and construction based on available information. The second step emphasizes on calibration for the most critical energy related variables. Those variables are key building energy drivers and their variations drive building energy change. The last step is to calibrate the model for less important variables based on collected information, assumptions or experience.

Typical linear regression variable selection methods such as stepwise methods, best subsets methods are not appropriate for variables selection of building data. This is because in practice, the number of metered variables is approximately close to or more than the sample size. The situation is very common in practice because of limited time and budget for data collection, lack of qualified meter reading etc. Therefore, the selected variables are biased and very sensitive to data outliers. To avoid those problems, we adopt variable selection methods from high dimensional big data analysis that are called LASSO and SCAD. Our methods are capable to avoid aforementioned issues and robust to data outliers. The LASSO method is briefly introduced in below.

The original high dimensional linear model includes all candidate variables with observations of n weekdays. It can be expressed as

$$Y = X\beta + \varepsilon, \tag{1}$$

where $\mathrm{Y} = (\mathrm{y}_1, \mathrm{y}_2, \cdots, \mathrm{y}_\mathrm{n})^\mathrm{T}$ are the total energy (kWh) collected in target building for n weekdays, and $\mathrm{X} = \left(\mathrm{x}_1, \cdots, \mathrm{x}_\mathrm{p}\right)^\mathrm{T}$ are the observations of p independent variables. Each vector x_i contains observations of each variable over n days. For example, n-days observation of lighting energy use. The vector norms are defined by

$$\|u\|_1 = \sum_{j=1}^{p} |u_j|, \tag{2}$$

$$\|u\|_2 = \sqrt{\sum_{j=1}^{p} u_j^2}\,,\ \ for\ all\ u \in R^p \tag{3}$$

The LASSO estimator is defined by

$$\hat{\beta}_{LASSO} = \operatorname{argmin}_{\beta\in\mathbb{R}^p} \frac{1}{2}\|Y - X\,\beta\|_2^2 + \lambda\|\beta\|_1 \tag{4}$$

where λ is a tuning parameter to control the model complexity (the number of nonzero coefficients). When $\lambda = 0$, $\hat{\beta}_{\mathrm{LASSO}}$ is the ordinary least squares estimator (OLSE); when $\lambda = \infty$, all coefficients are zero. The estimator is also called the penalized least squares (PLS) estimator.

An important property of the LASSO estimator is sparsity. It means several elements of $\hat{\beta}_{\mathrm{LASSO}}$ exactly equal to zero for a fixed $\lambda > 0$. To those zero coefficients, the corresponding independent variables don't have any contributions to response variable, and automatically screened. The nonzero coefficients are estimated at the same time. Another great advantage of LASSO approach is that its convexity guarantees global optimal results. For any given tuning parameter $\lambda > 0$, the LASSO estimator is unique.

During the last decade, statisticians and mathematicians spend great effort on improving the algorithms of LASSO. Nowadays, it can be conveniently applied in C, C++, Python, Matlab and R language. The technical proof of sparse property is beyond the scope of this paper. Readers are referred to (Tibshiriani 1996) for intuitive explanation. More detailed discussion can be found in (Fan and Li 2001), (Fan and Li 2006).

Another popular variable selection method is named "smoothly clipped absolute deviation penalty (SCAD)". SCAD approach is a member of regularization regression family and an improvement of LASSO method. It utilizes a special penalty function to identify significant variables. Similarly, we still consider the linear regression model in equation (1)

$$y = X\beta + \varepsilon \qquad (1)$$

Where y is a $n \times 1$ vector and X is $\left(\mathrm{x}_1, \cdots, \mathrm{x}_\mathrm{p}\right)^\mathrm{T}$, observations of independent variables. The dimension, the number of elements of β, can be larger than the number of observations. The ordinary least squares estimator is obtained via minimizing the quadratic loss function with respect to β.Some constrains are added on parameters to control the model complexity. These methods are called regularization regression or penalized least square estimate, and are defined as below

$$\hat{\beta}_{SCAD} = arg \min_{\beta \in R^p} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + n \sum_{j=1}^{p} P_\lambda(|\beta_j|) \qquad (5)$$

Where i stands for the observation of the i-th day and j stands for the j-th predictor. P_λ is designed penalty function and λ is the tuning parameter to control the model complexity. When λ is zero, the model is originally full model, when λ equals infinity, no variables are selected. (Fan and Li 2001) suggests three criteria to evaluate penalty function for coefficients estimation.

1. Unbiasedness: Modeling bias should be avoided when number of true coefficients is large.

2. Sparsity: Model complexity is reduced when P is large, non-significant coefficients are automatically shrink to zero.

3. Continuity: Continuity should be guaranteed in estimators to keep statistical stability of estimates.

The penalty function formulation is suitable to address three properties. Stepwise and best subsets selection methods are popular in past applications, but stochastic errors are accumulated in each step. Ridge regression encounters difficulty in shrinking nonsignificant coefficients into zero, derived model is hard to interpret. LASSO is a member of regularization regression methods, but it can't meet the unbiasedness criteria. It guarantees global optimal results because of its convexity properties. However, it might cause unpredictable bias of variables' coefficients estimators. The final selected model may contain redundant variables. Although redundant variables won't excessively affect prediction ability of the selected model, it can harm the explanation ability of important variables.

To overcome these drawbacks, (Fan and Li 2001) proposed Smoothly Clipped Absolute Deviation SCAD penalty functions. The mathematical description of penalty functions P_λ are defined in the (Fan and Li 2001). For more details, readers are suggested refer to that paper. There are several main reasons to apply LASSO and SCAD. First, the data demonstrate strong linear relationship, so a linear model with very high R square value will be suitable. Second, when the linear model works well and other nonlinear relationship improves the model marginally, a simpler model is always preferred due to the better interpretation. Third, nonlinear relationship can also be coped with by LASSO and SCAD by transforming the independent variables before plugging into the linear model. For example, the quadratic relationship can be achieved through linear model by first square all the independent variables. Last but not the least, most nonlinear relationship can be viewed as the limit of linear approximation. So eventually many statistical models are still setting up the linear model by adding higher order polynomials. Unlike LASSO method, SCAD enjoys the so-called "oracle" property. The property means that SCAD can exactly select all important variables and accurately estimate the nonzero coefficients, for proper tuning parameter. It's as if we knew the latent true model in advance like an oracle. The nonzero estimators are the same as ordinary least squares estimators only with the selected variables. Detailed information can be found in (Fan and Li 2001). The computation of SCAD is also a big challenge due to the non-convexity. (Li and Zou 2008) proposed the local linear approximation algorithm (LLA) to use LASSO algorithm to obtain the SCAD estimator. (Wang et al. 2013) proposed a high dimensional Bayesian information criterion (HBIC) to choose optimal tuning parameter and a calibrated LLA (Li and Zou 2008) to calculate the estimator.

OFFICE BUILDING CAMPUS CASE STUDY

In this section, we will demonstrate how to use LASSO and SCAD method to facilitate the energy model calibration process.

The case buildings are located in an office campus in Maryland, US. There are 7 buildings in the campus (Figure 1) with more than 2 million square feet. As shown in the figure, building 1 was built in 1959 and accommodates 1,350,000 gross square feet of space composed of five floors. Major renovations were completed between 2006 and 2008. The scope of the work included façade rebuilt, new lighting and power distribution, a new HVAC system and building automation system (BAS). The air distribution network is comprised of six central station air-handling units (AHU). Each AHU contains 3 supply fans, 3 return fans, a chilled water coil, pre-heat coil and main heating coil. The chilled and hot water is served from the campus main chiller plant and boiler plant.

The building 2 is a ten-story office building and the attached one-story auditorium was constructed in 1957. The building occupies approximately 312,000 gross square feet. The campus main chilled water plant and boiler plant are located in the basement of this building. The boiler plant has 3 boilers and have been renovated in 2011. The chiller plant has 4 chillers and a primary/secondary distribution system. Each chiller has its own dedicated primary pump and condenser pump.

The building 3 was built in 1964 and accommodates approximately 460,000 gross square feet of space composed of five floors, including a basement. A major renovation to the building was completed in 2002. The building has 9 central AHUs. It is a low temperature VAV system with perimeter reheat.

The building 4 & 5 were built in 1967. The 4 & 5 have approximately 124,000 and 15,000 gross square feet respectively. The major renovation was completed in April of 1999. The mechanical system is centralized AHU with VAV with reheat. In the perimeter spaces, they have fan coil units (FCUs) to provide additional heating. The building 4 & 5 chilled water is provided by the main campus chilled water and hot water plant.

The building 6 & 7 were constructed in 1973. The buildings combined occupy approximately 397,515 gross square feet. Two buildings are structurally independent from one another but are connected by a common interior corridor at the first and second floors. The building 7 is 5 stories tall with two levels of undergrounding parking and building 6 is two stories tall with one underground level as mechanical space. Both buildings are connected to its own chilled water plant. The plant has three 400-ton centrifugal chillers. The chilled water is primary/secondary loop and is supplied to air handlers located in both buildings. Both buildings use dual duct system. The heating is mainly provided by the electrical heating coil in the AHU. Perimeter spaces have unitary electrical reheat. The electrical heating system causes high energy use and cost in both buildings.

Figure 1 Case Study Campus Buildings

The campus has completed an advanced metering project in 2015. The installation of metering devices has enabled collection of usage data information for various energy sources throughout the campus. Electricity, water, natural gas and fuel oil data is gathered and reported through a central meter monitoring system. More than 1000 meters in total are installed on campus. A third-party energy management system (EMS) is installed on campus. The EMS is capable to visualize the data within the software and report and download the data set as queried.

In this advanced metering project, the energy model is intended to test energy saving and cost estimates for various energy conservation measures (ECMs) proposed by the energy consultant.

Model Development and Calibration Process

We collected various information about the buildings such as drawings, on site visit, metered data, etc. The energy model is developed in Openstudio and Energyplus with collected information. Energyplus is a public domain (funded by U.S Department of Energy), modular structured software tool which is very powerful to model HVAC system and simulate building energy use and thermal comfort. After model development, each building is calibrated individually with the steps mentioned below. In the past several years, the metering system collected more than 1000 data points. Each data is metered at 15-minute interval and stored in on site computer. The overall calibration workflow is a bottom-up process. The main idea is to tune the sub-system energy use to match the actual energy data prior to compare with the total energy use.

The calibration process is divided into three main steps which are (1) building envelop and ambient conditions, (2) calibrate key variables identified by LASSO and

SCAD method, (3) fine tune the rest of the variables. Each step is discussed in the following section.

Step (1): Building envelop and ambient condition. The as-built building floor plans and geometry is created according to design drawings and site visit. Since metered energy and flow devices doesn't record building geometry information, hence they are not applied in this step. The floor plans and space programs are retrieved from Revit model and drawings. Construction thermal properties and glazing parameters are set to match the reported value. When the information is not available, it is derived from typical values of similar buildings in the same region. Instead of running the energy model with typical meteorological weather file, the on-site weather data is set to have the same start and end dates of the calibration period (1/1/2016 to 12/31/2016).

Step (2): Calibrate key variables identified by LASSO or SCAD method. We use the collected data from 1/1/2016 to 12/31/2016 to run the LASSO or SCAD procedures. The data is grouped by each building. The daily total energy use (kWh) is the response variable Y in equation (1), collected various metered data and ambient conditions are represented as X in equation (1). In LASSO and SCAD method, the estimated coefficients corresponding to the insignificant variables are automatically shrunk to 0. Key variables remain on the final model are with nonzero estimators of coefficients. Before running the LASSO or SCAD procedures for each building, the observations of independent variables are standardized. This eliminates the effects of different data units and allows a fair comparison.

Variables selected by LASSO and SCAD method

After running the LASSO method for each building on campus, 6 variables are mutually selected which are ambient dry bulb temperature, interior lighting power density, plug load power density, space cooling and heating set point and supply air temperature.

The SCAD method result shows that only 2 variables, interior lighting and plug load power density, are the most critical for those office buildings on campus. Theoretically, SCAD method is more accurate than LASSO method and thus selected less key variables. Interior lighting and plug load power density is selected by LASSO as well. This shows that two variable selection methods are complementary to each other, and indicates that among selected key variables, interior lighting and plug loads are more important than others and should be calibrated with emphasis in the model.

After LASSO and SCAD identifies critical variables to case study building energy use, the next step is to generate inputs for those variables in the model. The following paragraphs discusses the logic of parsing 15-minute interval data to the model inputs.

Ambient dry bulb temperature: weather is commonly considered as a key driver of energy use and is selected by LASSO method in the project. In the simulation, we use on site weather data to formulate weather files for simulation, the start date is 1/1/2016, end date is 12/31/2016.

Interior lighting/ plug load: LASSO and SCAD method both identified that lighting and plug load are key drivers for buildings' energy use on campus. This agrees with our expectation. Nowadays, office building is primarily internal load driven, with large number of lights, numerous computers, printers and some data centers. The core spaces require 24 hours cooling for an entire year. After building renovation, our building envelop is well insulated, so they are less responsive to the weather change. The mechanical system operation is driven by the heat generated by lighting and plug load as well.

In order to post-process the data, metered interior lighting and plug interval data is divided into the following categories:

1. Interior lighting energy use during weekdays and weekends.
2. Interior plug energy use during weekdays and weekends.
3. Data center energy use.
4. Miscellaneous internal energy use.

The grouping of interval data into four categories above was driven by the scheduling of the operation in building. As in our buildings, weekdays have distinct operational difference from the weekend.

A logical stream was developed to generate inputs for lighting and plug inputs based on interval metered data. The selection of the method was based on the availability of interval energy use data and whether irregular building usage was apparent. The logic is shown as following:

- If interval data is available and length of problematic period is less than a month: Monthly weekday and weekend load profiles are generated to develop load power density and schedules. Problematic data points are extrapolated and estimated from normal typical operational periods.
- If interval data is available and length of problematic period expands more than a month: since the difficulty of disaggregating problematic data from one month to another, extrapolation will cause additional uncertainty into the model. Therefore, those data points are excluded from the analysis. Monthly weekday and weekend load

profiles and power densities are generated only through the available and qualified periods.

- If interval data is available but shows strong irregular operation or erroneous data values: The data visualization will indicate obvious contradiction of typical operation condition or abnormal values. The inputs to those metered data will be generated from the correct data periods or values derived from the site visit.

In case study buildings, some have lighting and plug load data metered in each zone, some only available per floor or at building levels. The inputs are generated based on data availablility. When the zone level data is unavailable, averaged whole building load inputs are applied. For example, the 15-minute interval monthly weekday interior lighting profile in building 7 is shown in figure 2. Corresponding lighting power density and schedule inputs (figure 3) are derived accordingly from the graph. The maximum power desntiy is calcualted as maximum hourly energy use divided by building area. Same method is applied for plug load data inputs. Simulated and acutal daily lighting energy use comparison is shown in figure 4.

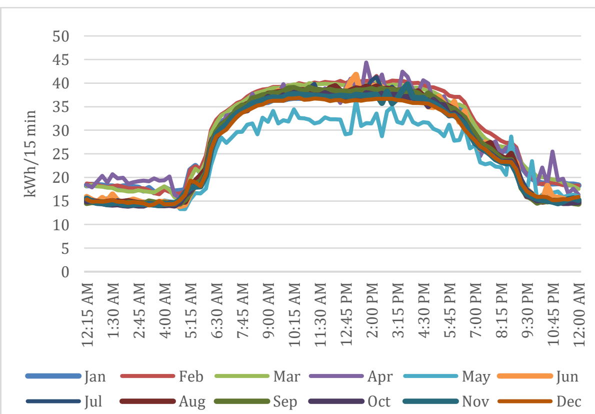

Figure 2 Building 7 Monthly Weekday Interior Lighting Energy Profile

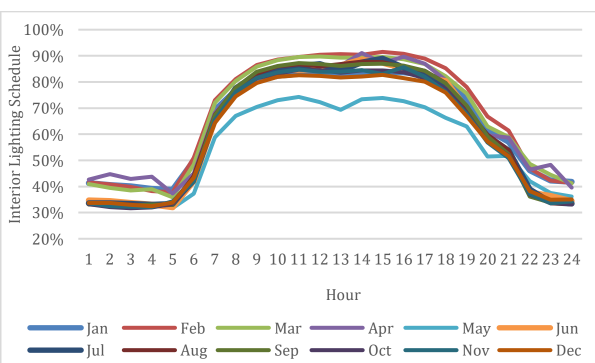

Figure 3 Building 7 Monthly Weekday Interior Lighting Hourly Schedule Profile

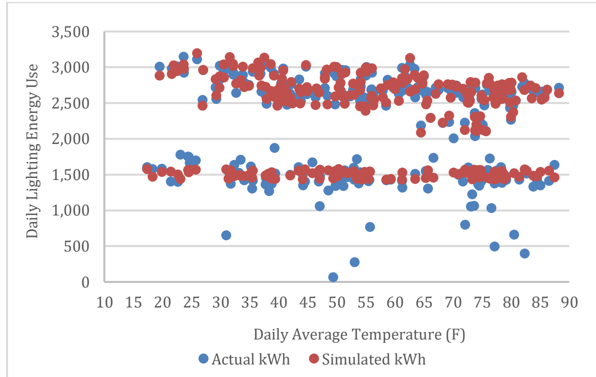

Figure 4 Building 7 Interior Lighting Daily Energy Comparison

Space cooling, heating set point and supply air temperature: Space cooling/ heating set point has reset strategy based on the occupied period. Typical values for cooling is 74°F occupied and 84°F unoccupied period. Heating is 72°F occupied and 65°F unoccupied period. The values are recorded in the building management system (BMS). The set point profiles are developed for each building based on the data retrieved from the BMS. Supply air temperature profile is derived similarly as the spaces cooling and heating set point.

After update key variables selected by LASSO and SCAD, we compared the energy use of lighting, plug, data center. For example, in building 1, the monthly energy end use comparisons are shown in below.

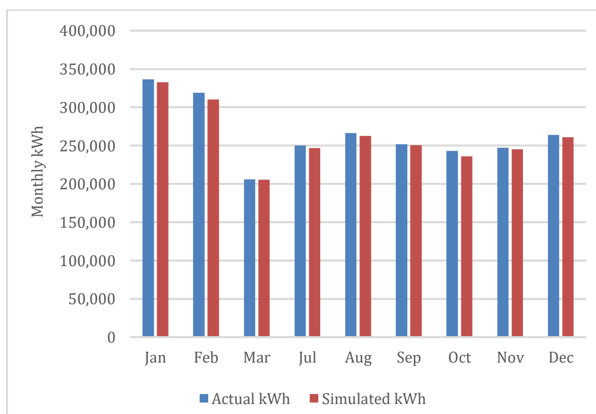

Figure 5 Building 1 Interior Lighting Monthly Energy Comparison

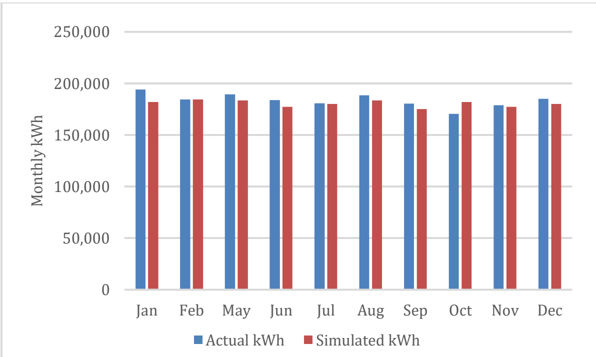

Figure 6 Building 1 Plug Monthly Energy Comparison

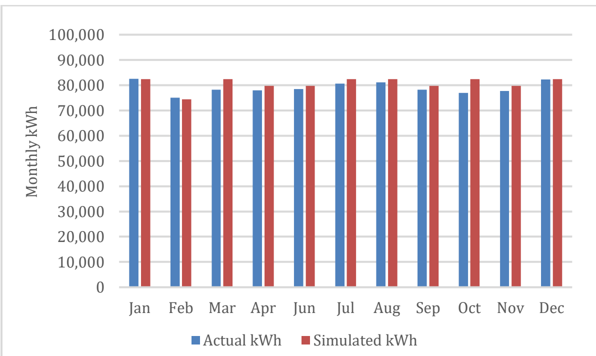

Figure 7 Building 1 Data Center Monthly Energy Comparison

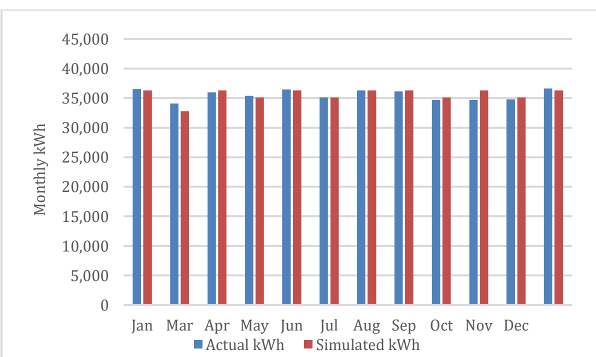

Figure 8 Building 1 Telecommunication Room Monthly Energy Comparison

Step (3): Fine tune the rest of the model. From the above step, we can see key variables related energy uses are simulated accurately. In this step, we updated the model with information collected from site visit, BMS such as fan power, pressure rise, pump head etc. When the data is not available, experienced values or typical assumption data is used.

After finishing steps above, the simulated total energy use of each building is compared with the data. For example, Figure 9 and 10 shows the simulated building 1 daily and monthly total energy use comparison. In this project, our goal is to compare annual ECMs saving, thus daily and monthly data calibration is sufficient.

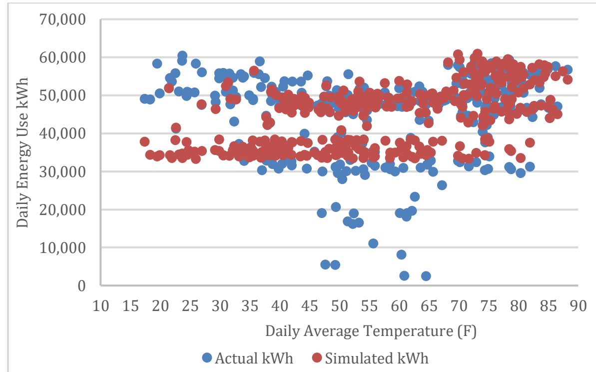

Figure 9 Building 1 Daily Total Energy Use Comparison

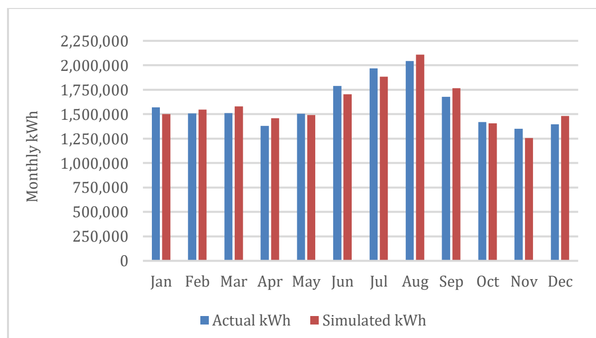

Figure 10 Building 1 Total Energy Comparison

In Figure 10, the coefficient variation of root mean square error (CVRMSE) is 4.76% , normalized mean bias error (NMBE) is -1.36% and meet criteria defined in ASHRAE guideline 14. Then we sum up all building monthly usage to get the whole campus simulated result in Figure 11 (June data is missing because of meter reading error).

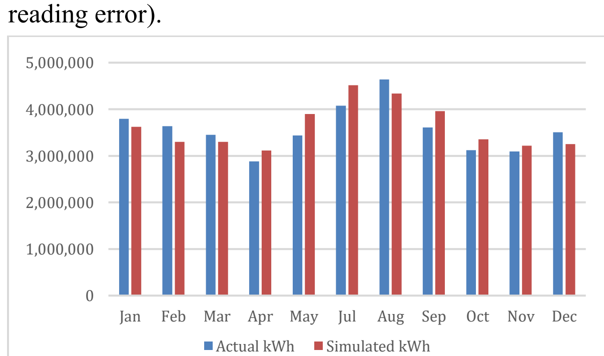

Figure 11 Campus Total Electric Use Comparison

CONCLUSION

The goal of this research is to apply high dimensional big data analysis method to calibrate building energy model. To achieve the goal, this paper briefly discussed background and theory of LASSO and SCAD method. Compared with conventional variable selection approaches, our methods are robust to deal with insufficient data observations, noise and outliers. More importantly, these two methods help us prioritize calibration effort. In the research, we proposed a 3-step calibration strategy and explained the idea and method for each step. Given limited staff resource, time and budget in the actual project, this strategy significantly reduces the time and effort on calibration and achieve high model accuracy. A 2 million square feet office building campus data in the U.S northeast region is used as case study to verify the method in application.

The advantages of this approach over traditional methods are discussed. The calibrated model can be used to analyze savings for different ECMs and applied in measurement and verification. Selected variables from LASSO and SCAD method are key variables that relate to the building energy use. In the energy management system, those variables should be monitored continuously and compared with historical data. This will show the energy flow and trends inside the building. This calibration steps can be followed in other building types and data.

Reference

Ahmad, M. and Culp, C.H., 2006. Uncalibrated building energy simulation modeling results. HVAC&R Research, 12(4), pp.1141-1155.

Department of Energy (DOE). 2011. Building Energy Data Book.

Energy Information Administration (EIA). 2017. Commercial Building Energy Consumption Survey (CBECS)

Fan, J. and Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), pp.1348-1360.

Fan, J. and Li, R., 2006. Statistical challenges with high dimensionality: Feature selection in knowledge discovery. arXiv preprint math/0602133.

O'Donnell, J.T., 2009. Specification of optimum holistic building environmental and energy performance information to support informed decision making.

Raftery, P., Keane, M. and Costa, A., 2009, July. Calibration of a detailed simulation model to energy monitoring system data: a methodology and case study. In Eleventh International IBPSA Conference (pp. 1199-1206).

Song, S. and Haberl, J., 2008. A procedure for the performance evaluation of a new commercial building: Part I-Calibrated as-built simulation.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), pp.267-288.

Wang, L., Kim, Y. and Li, R., 2013. Calibrating non-convex penalized regression in ultra-high dimension. Annals of statistics, 41(5), p.2505.

Westphal, F.S. and Lamberts, R., 2005, August. Building simulation calibration using sensitivity analysis. In Ninth International IBPSA Conference (pp. 1331-1338).

Zou, H. and Li, R., 2008. One-step sparse estimates in nonconcave penalized likelihood models. Annals of statistics, 36(4), p.1509.

Nomenclature

Y observations of response variable (total energy use).

X observations of independent variables.

β the coefficients or parameters of linear regression model (1).

$\hat{\beta}$ the statistical estimator by using the observations.

λ the tuning parameter for LASSO or SCAD method.

n the sample size (observation days)

p the total number of considered independent variables

P_λ the penalty function with tuning parameter λ.

ε random error terms.

u a p-dimensional vector.

R^p p-dimensional real space.

$|.|$ the absolute value.

$\|.\|_2$ the Euclidean norm.

Subscripts

i the i-th days

j the j-th considered independent variable.