Building Energy Forecasting Model for On-line Operation Using System Identification

Xiwang Li  Jin Wen  Teresa Wu  Er-Wei Bai
Student Member ASHRAE  Member ASHRAE

ABSTRACT

Model based control has become a promising solution for building optimal control and operation to save energy. High fidelity building energy forecasting models, as the basis of on-line model based control, have strict requirements for forecasting accuracy and computationally efficiency. Currently, most of the existing energy forecasting models are in three different categories, namely white box models, black box models and grey box models, among which black box models and grey box models are commonly used in on-line building control. However, pure data-driven black box models often require long training period and are bounded to building operation conditions. On the other hand, creating even a simplified grey box model is often time consuming and need expert knowledge. Therefore, this study proposes to develop a building energy forecasting methodology for building on-line building control and operation using system identification approach. In this study, optimal experiment design theory has been applied to analyze the nonlinearity of building energy system. Based on the system nonlinearity characteristics, system identification model structure will be determined to guarantee the model accuracy and robustness. What’s more, different building energy system excitation strategies will also be applied to generate enough model training data based on data information theories. In the end, the proposed building energy forecasting methodology is validated against EnergyPlus simulation model, and compared with other common modeling approaches, such as Resistance and Capacitance model, Support Vector Machine for Regression model, and Artificial Neural Networks model, in terms of model accuracy, calculation speed, and model uncertainty.

INTRODUCTION

Buildings consume about 41.1% of primary energy and 74% of the electricity in the U.S (DOE 2013). Moreover, it is estimated by the National Energy Technology Laboratory that more than 1/4 of the 713 GW of U.S. electricity demand in 2010 could be dispatchable if only buildings could respond to that dispatch through advanced building energy control and operation strategies and smart grid infrastructure (NETL, 2011). Although batter building design and appliance choosing are capable of reducing energy consumption, it is not a cost-effective way to be applied in every building, especially in existing buildings. Therefore, it is worthwhile and urgent to improve the building control and operation strategies to improve energy efficiency.

In the different building control and operation strategy determination, model based control has been proven to be a most promising approach. The effectiveness of model based control process is significantly affected by the performance building energy form casting models. How to develop accurate, robust, and cost-effective building energy forecasting model is focus of this study. Generally speaking, there are three categorizes of building energy modeling approaches, namely, white box models, black box models, and grey box models. Each of these three modeling approaches has its own advantages and disadvantages. Due to the high computation demanding, few studies have used white box models for building control and operation. Resistance and Capacitance (RC) network model, however, is a widely used grey box model for building operation, based on which numerous studies applied model based control methodology to improve the building energy efficiency and reduce the energy demand by fully utilizing building passive thermal storage (Oldewurtel, et al, 2012).
Different parameter determination methods based on global searching, genetic algorithm, and etc. have been implemented to determination the Rs and Cs. Unfortunately, determining the parameter of RC model is computational demanding and the model structure and parameter are different from buildings to buildings. It is impossible to utilize a RC model for one building to another building. Black box models, such as autoregressive exogenous (ARX), artificial neural networks (ANN), and Support Vector Machine for Regression (SVR) are also widely used in model based building control and operation, because of the high calculation speed. ARX has been used forecast the energy consumption in different model based control studies to reduce energy and demand cost, such as (Avci et al., 2013). ANN and SVR are also selected in a lot of studies, such as (Moon, 2012, Xi et al., 2007). However, data driven models often require long training period and are bounded to building operating conditions that they are trained for. The training data requirement is always a huge obstacle for grey box model and black box model development, due to the sensor quality and sensor availability.

On the contrary, system identification is a process of developing or improving a mathematical representation of a physical system using data that are collected from a designed operation or experiment, in an active manner. Although system identification techniques have been used in other engineering applications, there are only limited applications of system identification techniques in building modeling. Privara et al. (2011) proposed an approach combining the EnergyPlus model and a subspace system identification model to forecast building performance. A MATLAB-BCVTB-EnergyPlus testbed was developed for building excitation, system identification and building performance forecasting. Then a subspace model in MATLAB N4SID toolbox was used for system identification and operation forecasting. An active system excitation based on sum of sinusoids was applied in (Li and Wen, 2014). Besides the model structure selection, system order is another critical parameter to model accuracy and calculation speed, which is determinate by the rank of the Hankel matrix and obtained through singular value decomposition (SVD) decomposition.

In this study, a novel building energy forecasting model is developed based on system identification approach. The performance of this is compared with typical existing models, such RC model, SVR model, and ANN model. The energy forecasting accuracy, uncertainty and engineering effort are compared. A validated detailed mid-size commercial building EnergyPlus model was used to provide building operation data in lieu of a real building.

**BUILDING ENERGY SYSTEM IDENTIFICATION MODEL DEVELOPMENT**

As introduced in last section, a small-size office building EnergyPlus simulation model provided by U.S. Department of Energy (DOE) was used in lieu of a real building for model development and validation. The procedure of building prior information collection, model structure determining, model development and verification will be introduced in detail in this section.

**Building Prior Information Collection**

It is believed that a system’s nonlinearity affects the system identification structure selection. Therefore a system nonlinearity test is designed and applied here to obtain necessary information for system identification. Generally, the nonlinearity of building energy system is dependent on, but not limited to, building size, floor are, HVAC system type, and HVAC system size. The physics theories for building energy system are very complicated and with a lot of disturbances. It is very difficult to determine the system characteristics of building energy system directly from theoretical analysis. Therefore, the building energy system nonlinearity is evaluated through an active test, by injecting different testing input signal and analyzing the relationship of input and output signals.

A nonlinearity index, based on the spectral density method is then developed using coherence between the system input and output:

\[ C_{xy} = \frac{|G_{xy}|^2}{G_{xx}G_{yy}} \]  

Where, \( G_{xy} \) is the cross-spectral density between system input \( x \) and system output \( y \), and \( G_{xx} \) and \( G_{yy} \) the auto-spectral density of \( x \) and \( y \) respectively. If \( C_{xy} = 1 \), then the system is a linear system, and if \( 0 < C_{xy} < 1 \), then the system is a
nonlinear system. Therefore, the results of $C_{xy}$ will provide a critical prior information for system identification model development.

**Energy Forecasting Model Structure Selection**

As stated above, the objective of this step is to develop an on-line building energy model using system identification method. Model structure plays an important role in model forecasting accuracy. From the results of nonlinearity test, even though building energy systems are nonlinear systems, the nonlinearity behaves in a relative long time interval manner. A lot of linear models have been successfully applied in building energy consumption forecasting. Based on these characteristics, frequency response function approach is chosen in this study due to its excellent performance in capture system dynamics in frequency domain and computation efficiency. Fundamentally a frequency response function is a mathematical representation of the relationship between the input and the output of a system in frequency domain, which can simplify the time domain transfer function and still capture the useful information of the system dynamic responses, as demonstrated in Eq. 1:

$$H(j\omega) = \frac{Y(j\omega)}{U(j\omega)} = \frac{S_{yu}(j\omega)}{S_{uu}(j\omega)}$$  \hspace{1cm} (1)

Where $Y(j\omega)$ is the Fourier transform of system output $y(t)$, and $U(j\omega)$ is the Fourier transform of system input $u(t)$. However, better results can be obtained in practice by computing the frequency response function $(S_{yu})$ as the ratio of cross-spectrum between input and output to the power spectrum of the input $(S_{uu})$ (Ljung, 1999).

**Energy Forecasting Model Input and Output Determination**

The inputs and outputs of the whole building system identification model are selected based on the building physics and are tabulated in Table 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable Name</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ebldg</td>
<td>Building cooling energy (J)</td>
<td>Output</td>
</tr>
<tr>
<td>Tout</td>
<td>Outdoor air temperature (C)</td>
<td>Input</td>
</tr>
<tr>
<td>Tzone, i</td>
<td>Zone i temperature (C)</td>
<td>Input</td>
</tr>
<tr>
<td>Rin, i</td>
<td>Equipment /occupancy schedule in zone i (-)</td>
<td>Input</td>
</tr>
<tr>
<td>Tsol-air</td>
<td>Solar air temperature (C)</td>
<td>Input</td>
</tr>
<tr>
<td>Qdir</td>
<td>Direct solar radiation (W/m²)</td>
<td>Input</td>
</tr>
<tr>
<td>Qdif</td>
<td>Diffuse solar radiation (W/m²)</td>
<td>Input</td>
</tr>
<tr>
<td>Voa</td>
<td>Ventilation rate (m³/s)</td>
<td>Input</td>
</tr>
<tr>
<td>Qfan</td>
<td>Supply fan power (W)</td>
<td>Input</td>
</tr>
</tbody>
</table>

In the real application, building cooling energy (Ebldg), outdoor air temperature (Tout), zone temperature (Tzone), occupancy schedule (Rin) and fan heat gains (Qfan) are usually available in building automation system or easy to measure. All the solar information can be obtained from National Solar Radiation Database (Wilcox, 2007). Unfortunately, buildings usually operate within a very narrow range of temperature setpoints and internal equipment schedules. Therefore, the zone temperature, and equipment and occupancy schedule ratio are excited inputs, which will be excited during a model training period in the EnergyPlus simulation testbed.

**Energy Forecasting Model Training: System Excitation**

As introduced previously, buildings are usually operating in a very narrow range of temperature setpoints and internal equipment schedules. Therefore, in this active system identification, the building energy system will be excited to provide enough high quality model training data. Sum of sinusoids (SINE) model is used to generate the exciting signals (Eq.2),
because sinusoids signals are versatile periodic and can adjust signal shape and character of the power spectrum by adjusting their parameters.

\[ U_{τ+1} = U_τ + \sqrt{2α_τ} \sin(ω_τ T + φ_τ) \]  

(2)

Where \( U_{τ+1} \) is the excitation signal; \( \sqrt{2α_τ} \) is a magnitude scale parameter from 0 to 1; \( T \) is the sampling time, and \( φ \) is the phase lag parameter from 0 to \( 2\pi \), which do not affect the signal spectrum; \( ω \) is periodic frequency parameter from 0 to \( 2\pi \), which is determined by Eq.3:

\[ \frac{1}{β_{dom}^H} ≤ ω_τ ≤ \frac{α_{S}}{τ_{dom}^H} \]  

(3)

where, \( τ_{dom}^H \) and \( τ_{dom}^L \) correspond to the high and low estimates of the dominant time constant of the system (denote the slowest and the fastest systems time constants (Rivera, 2009). \( α_{S} \) and \( β_{S} \) are user-decisions on high and low frequency content.

A parametric case study has been conducted to determine the sampling length and signals injection interval, which are 6 hours and 30 minutes in this study. Following the procedure discussed above, the exciting signals for temperature setpoints and equipment schedules used in this study are generated, where all the excitation signals are injected into the building model every 30 minutes, and a moving window containing 6 hours training data will be processed simultaneously to develop the energy forecasting model.

**Building Energy Forecasting Model Realization**

Based on the training data generated from the excited system, a system identification model based on spectral density model for frequency response function is then developed. Figure 2 shows the model development procedure upon the operation data. In this figure, \( \text{U} \) is training inputs, \( \text{Y} \) is training outputs data, \( \text{PSD} \) is power spectral density model for inputs, and \( \text{CPSD} \) is cross power spectral density model for input and output. \( S_{uu} \) and \( S_{yu} \) are the results of \( \text{PSD} \) and \( \text{CPSD} \), respectively. \( G(z) \) is the transfer function in frequency domain, which can be transferred to time domain transfer function \( G(t) \) using inverse Fourier function transformation, and saved as a set of Markov parameters. \( \hat{Y} \) is the forecasting results.

![Figure 2 System identification model](image)

**BUILDING ENERGY FORECASTING MODEL VALIDATION AND PERFORMANCE COMPARISON**

In this section, the system identification (SID) model developed previously is utilized to forecast the building cooling energy consumption in summer for the small-size office building. Meanwhile, different models such as RC & chiller model, SVR model and ANN model are also applied to compare with the performance of the SID model. On the other hand, in order to assess the uncertainty of those energy forecasting models, Gaussian distributed white noise is added into the inputs of each model, and then the uncertainty is evaluated in a Monte Carlo (MC) simulation process.
System Identification Case Study: Energy Forecasting

The mid-size commercial building studied in this project is a single story office building, which has six zones, five conditioned zones and an unconditioned attic zone, and the total floor area is 510 m². The window-to-wall ratio of this building’s facades is approximately 21.2%, and the windows are equally distributed. The overall U-factor of these single pane windows is 3.4 W/m²K and the solar heat gain factor is 0.36. The solar absorptivity, transmissivity and reflectivity are 0.06, 0.69 and 0.24, respectively. The roof insulation has an R-value of 15. The roof is covered in an asphalt membrane, with a solar absorptivity value of 0.9. The overall U-factor for the walls is 0.68 W/m²K. The building location is selected as in Philadelphia, PA, USA for this study. The HVAC systems used in this building are constant-air-volume (CAV) air handling units (AHUs) with direct expansion (DX) coils.

Performance Comparison: SID, RC, SVR, ANN

In the energy forecasting study, all these different models are trained using the operation measurements during Aug.1 to Aug. 14 (SID model uses operation data from Aug.1 to Aug 7) from a same building model and under the same conditions.

The parameters in the RC model, such as Rs, Cs and parameters for the chiller models are determined through pattern searching based optimization methods by minimizing the difference between the forecasting and measured cooling energy consumption (Eq. 4).

\[
J(R, C, \rho) = \sqrt{\frac{\sum_{j=1}^{N}(Q_{RC,j} - Q_{Act,j})^2}{N-1}}
\]

(4)

Where, \(Q_{RC}\) and \(Q_{Act}\) are the cooling energy consumption from RC model and EnergyPlus model, respectively; \(N\) is the total time step of this whole simulation. \(R\), \(C\), and \(\rho\) are parameters in RC & Chiller model.

In the SVR model, radial basis function kernel (Eq. 5) was used in LibSVM (Chang and Lin, 2011).

\[
K(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)
\]

(5)

Where, \(x\) and \(x'\) are real and forecasted energy consumption, respectively; \(\sigma\) is user-defined parameter.

An ANN model with 10 sublayers was developed using a Neural Network Toolbox (Beale, 2010) in the forecasting comparison study.
In order to evaluate the forecasting accuracy, two indexes are employed, namely, Coefficient of Determination ($R^2$) and Normalized Root Mean Square Error (NRMSE):

$$R^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\hat{x}_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(\hat{x}_i - \bar{x})^2}$$

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} / \left(\frac{x_{\max} - x_{\min}}{n}\right)$$

Where $x_i$ and $\hat{x}_i$ is the true and forecasting value; $\bar{x}$ and $\bar{\hat{x}}$ are the average of true and forecasting value, respectively.

The energy forecasting results from these four different models are illustrated in Figure 3 and Table 2. The results show that the SID model achieved the highest forecasting accuracy. Even though it requires longer calculation time than SVR and ANN, the calculation time for one-week energy consumption is still acceptable.

### Table 2. Cooling Energy Forecasting Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Period</th>
<th>Time</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>NRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SID</td>
<td>0801-0807</td>
<td>21s</td>
<td>0.96</td>
<td>0.48 kW</td>
<td>7%</td>
</tr>
<tr>
<td>RC &amp; Chiller</td>
<td>0801-0814</td>
<td>451s</td>
<td>0.87</td>
<td>0.83 kW</td>
<td>11.70%</td>
</tr>
<tr>
<td>SVR</td>
<td>0801-0814</td>
<td>0.02s</td>
<td>0.69</td>
<td>1.04 kW</td>
<td>14.59%</td>
</tr>
<tr>
<td>ANN</td>
<td>0801-0814</td>
<td>2.78s</td>
<td>0.93</td>
<td>0.68 kW</td>
<td>9.60%</td>
</tr>
</tbody>
</table>

Similarly, more comparison study was then conducted under different conditions. Five scenarios are designed using the same model trained previously to forecast the cooling energy consumption of the same building under different conditions. The results, summarized in Table 3, shows that SID model is able to achieve better result than RC & Chiller model when the forecasting condition is close to the training condition, and to achieve better result than SVR and ANN models when the forecasting condition is far from the training condition.

In summary, the SID model has more extendibility and higher accuracy than SVR and ANN models when forecasting condition changes, and has higher accuracy and higher speed than RC & chiller model when forecasting condition is similar to training condition.

### Table 3. Cooling Energy Forecasting Performance

<table>
<thead>
<tr>
<th>Model Training</th>
<th>Cooling Energy Forecasting</th>
<th>Forecasting accuracy, $R^2$</th>
<th>Forecasting accuracy, NRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SID</td>
<td>0816-0818</td>
<td>Sid</td>
<td>RC &amp; Chiller</td>
</tr>
<tr>
<td>RC &amp; Chiller</td>
<td>0826-0828</td>
<td>0.94</td>
<td>0.91</td>
</tr>
<tr>
<td>SVR</td>
<td>0716-0718</td>
<td>0.85</td>
<td>0.81</td>
</tr>
<tr>
<td>ANN</td>
<td>0701-0730</td>
<td>0.86</td>
<td>0.81</td>
</tr>
<tr>
<td>0801-0814</td>
<td>0901-0930</td>
<td>0.89</td>
<td>0.75</td>
</tr>
</tbody>
</table>

**Model Uncertainty Analysis: RC, SVR, ANN**

In real cases, most of the measurements are contaminated with noises. Therefore, the building energy forecasting models are tested in the noisy conditions, where Gaussian distributed white noise was added to each input.

In this test, 5% noise was chosen, which means random noise from -5% to +5% of each variable at each time step was added into each input. The same cooling energy forecasting models were developed here again upon the noisy measurements, which were then used to forecast the cooling energy consumption based on the noisy measurements at forecasting period. The energy forecasting results during Aug.22 to Aug.28 from all these four models are summarized in
Table 4. As expected, the performance of all these four models decreased comparing to that under no noise condition. The accuracy of SID model, however, is still the highest under the noisy data condition, followed by the RC & chiller model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Period</th>
<th>Time</th>
<th>R²</th>
<th>RMSE</th>
<th>NRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SID</td>
<td>0801-0807</td>
<td>23.8s</td>
<td>0.79</td>
<td>1.02 kW</td>
<td>14.8%</td>
</tr>
<tr>
<td>RC &amp; Chiller</td>
<td>0801-0814</td>
<td>457.3s</td>
<td>0.74</td>
<td>1.19 kW</td>
<td>17%</td>
</tr>
<tr>
<td>SVR</td>
<td>0801-0814</td>
<td>0.03s</td>
<td>0.63</td>
<td>1.09 kW</td>
<td>18%</td>
</tr>
<tr>
<td>ANN</td>
<td>0801-0814</td>
<td>2.82s</td>
<td>0.67</td>
<td>1.13 kW</td>
<td>15.9%</td>
</tr>
</tbody>
</table>

Beyond the energy forecasting case study under noisy conditions, a Monte Carlo (MC) simulation was conducted to analyze the model uncertainty. N=1000 simulations are executed in these four models. In summary, the MC process steps were:

1. Initialize Monte Carlo simulation by defining input noise distributions and injecting them into each model;
2. Perform Monte Carlo: for i=1...N
   - Sample noise values from defined distributions
   - Run each model for energy forecasting
   - Calculate MC output (daily energy consumption) from each model
3. Analyze the performance of each model

*Figure 4. Boxplots of Monte Carlo Daily Energy Consumption Simulation Results*

Boxplots of the MC output (daily energy consumption) during the testing period (Aug.22 to Aug.28) are illustrated in Figure 4. The boxes show the 5% and 95% percentile and the middle lines show the mean of the daily energy consumption (kWh) during the MC simulation. Even though the SVR model and ANN model have achieved lower uncertainty, their
accuracy is much lower than that of SID model and RC & chiller model. It means that SID model and RC & chiller model are more sensitive to the input noise, while they still can maintain better accuracy than SVR model and ANN model (Table 5). Table 5 shows SID model and RC & chiller model achieved the lowest RMSE in 5 days and in 2 days, respectively, in the 1000 times MCMC runs.

Table 5. Energy Forecasting Accuracy in MCMC Simulation

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 6</th>
<th>Day 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SID</td>
<td>6.08</td>
<td>3.88</td>
<td>4.80</td>
<td>4.82</td>
<td>6.51</td>
<td>12.83</td>
<td>10.55</td>
<td></td>
</tr>
<tr>
<td>RC &amp; chiller</td>
<td>10.98</td>
<td>2.65</td>
<td>36.41</td>
<td>27.65</td>
<td>22.53</td>
<td>8.00</td>
<td>10.89</td>
<td></td>
</tr>
<tr>
<td>SVR</td>
<td>13.17</td>
<td>7.40</td>
<td>33.25</td>
<td>25.25</td>
<td>25.57</td>
<td>11.15</td>
<td>12.09</td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>7.87</td>
<td>22.12</td>
<td>15.26</td>
<td>3.74</td>
<td>5.58</td>
<td>11.55</td>
<td>13.23</td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSION AND FUTURE WORK

This study developed a novel system identification model for building energy forecasting based on the system characteristics. Optimal experiment design theory has been applied to collect model training data through active system excitation. The performance, such as accuracy, calculation speed, extendibility, and uncertainty, of the system identification model is compared with RC & chiller, SVR and ANN, three common modeling approaches. In the comparison study, the novel SID model has the capability to achieve high accuracy and extendibility in the noise-free and noisy conditions. Future efforts are needed to shorten the excitation period and to examine the developed methodologies for different types of buildings and HVAC systems, and for real world conditions.

ACKNOWLEDGEMENT

Financial support provided by the U.S. National Science Foundation Award (ID: 1239247) is greatly appreciated.

REFERENCES

Ljung, L., System identification 1999: Wiley Online Library.