BUILDING ENERGY CONSUMPTION ON-LINE FORECASTING USING SYSTEM IDENTIFICATION AND DATA FUSION

Xi wang Li  
Department of Civil, Architectural and Environmental Engineering  
Drexel University  
Philadelphia, PA, USA  
Email: xlew3@drexel.edu

Jin Wen  
Department of Civil, Architectural and Environmental Engineering  
Drexel University  
Philadelphia, PA, USA  
Email: jinwen@drexel.edu

ABSTRACT

Model based control has been proven to have significant building energy saving potentials through operation optimization. Accurate and computationally efficient, and cost-effective building energy model are essential for model based control. Existing studies in this area have mostly been focusing on reducing computation burden using simplified physics based modeling approach. However, creating and identification the simplified physics based model is often challenging and requires significant engineering efforts. Therefore, this study proposes a novel methodology to develop building energy estimation models for on-line building control and optimization using an integrated system identification and data fusion approach. System identification model has been developed based on frequency domain spectral density analysis. Eigensystem realization algorithm is used to generate the state space model from the Markov parameters. Kalman filter based data fusion technique has also been implemented to improve the accuracy and robustness of the model by incorporating with real measurements. A systematic analysis of system structure, system excitation selection as well as data fusion implementation is also demonstrated. The developed strategies are evaluated using a simulated testing building (simulated in EnergyPlus environment). The overall building energy estimation accuracy from this proposed model can reach to above 95% within 2 minutes calculation time, when compared against detailed physics based simulation results from the EnergyPlus model.

The electricity consumption of the US grew 1.7% annually from 1996 to 2006, and the total growth will reach 26% until 2030 [1]. Among that consumption, buildings are responsible for over 70% of electricity consumption in the US [2]. Studies have shown that most of the commercial and residential buildings have equipment and operational problems that reduce the comfort and waste more energy. Around 4% to 20% of energy used in HVAC and lighting system [3] was wasted due to equipment and operation problems. Moreover, it is estimated by the National Energy Technology Laboratory that more than one-fourth 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 [4]. Therefore the quality of building control and operation is significant economically and environmentally.

The quality of building control and operation is significantly affected by the building energy forecasting models. How to develop accurate, robust, and cost-effective building energy forecasting models is the focus of this study. One of the most comprehensive white box model in the existing building energy modeling tools is EnergyPlus, which is a whole building energy simulation program that engineers, architects, and researchers use to model energy and water use in buildings [5]. Moreover, In order to apply EnergyPlus in modeled based building control and operation, a Building Controls Virtual Test Bed (BCVTB) was developed by Wetter and Haves to link the building models (EnergyPlus) with real control systems [6]. BCVTB can be sued as a middleware tool that allows to data sharing among different simulation programs, such as EnergyPlus, Matlab, Modelica, and etc., for distributed simulation. Therefore, through this test bed different user defined building control and optimization strategies can be applied into different building simulation models. For example, Ma et al. proposed and demonstrated an economic MPC technique to reduce energy and demand cost using EnergyPlus and BCVTB [7]. Even though these elaborate simulation tools are very effective and accurate, they require detailed information and parameters of buildings, energy system and outside weather conditions. Identifying these parameters,
is implemented to capture the dynamics of building energy system and forecast the energy consumption.

Besides using simulation alone, Kalman filter based data fusion framework is also proposed to be used in this study, in cooperate with simulation and real measurements. Building energy estimation based on data fusion has recently received attention for its real time on-line updating capabilities which is able to capture the time varying building dynamics. An integrated 3R2C and EKF (Extended Kalman Filter) model was developed to estimate the building energy consumption in [17]. In this work, the EKF was used to estimate the state vector $X$ using real sensor measurement data. The estimated load matched the EnergyPlus results within 10% at 93% of the time. This RC-EKF approach has also been tried in [18]. In this study, a self-adaptive thermal building model was developed based on a 1R1C model and an EKF. Existing Kalman filter studies show good potential for this technique, when combined with other techniques, to further improve the building energy model accuracy and robustness.

In this study, a detailed discussion about system identification model development, state space model reformation, and data fusion implementation, used to develop a cost-effective building energy model are presented. A validated detailed mid-size commercial building energy simulation model is used to provide building operation data in lieu of a real building.

**BUILDING ENERGY SYSTEM IDENTIFICATION MODEL DEVELOPMENT**

In this study, an on-line building energy forecasting model is developed for a mid-size commercial building. In lieu of real building operational data, a EnergyPlus simulation test bed, provided by U.S. Department of Energy (DOE) [19], is used to generate operational data to train and validate the developed on-line building energy forecasting model. The procedure of model structure determining, input and output selection, system exciting and model training and validation will be introduced in detail in this section.

**System Identification Model Development**

**Model 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. In order to develop a relative simple system model, frequency response function approach is applied in this study due to its excellent performance in handling system nonlinearity [15]. 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. (2):

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

(2)
Where $Y(j\omega)$ 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 ($S_{uu}$) [15]. Then by applying the Inverse Fourier Transform, the Impulse Response Functions (IRF) per measurement channel are obtained.

**Determining Input and Output**

The inputs and outputs of the whole building system identification model are determined 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>$E_{\text{zone}}$</td>
<td>Building cooling energy (J)</td>
<td>Output</td>
</tr>
<tr>
<td>$T_{\text{out}}$</td>
<td>Outdoor air temperature (C)</td>
<td>Input</td>
</tr>
<tr>
<td>$T_{\text{zone,i}}$</td>
<td>Zone temperature (C)</td>
<td>Input</td>
</tr>
<tr>
<td>$R_{\text{in},i}$</td>
<td>Equipment/occupancy schedule in zone $i$</td>
<td>Input</td>
</tr>
<tr>
<td>$T_{\text{sol-air}}$</td>
<td>Solar air temperature (C)</td>
<td>Input</td>
</tr>
<tr>
<td>$Q_{\text{dir}}$</td>
<td>Direct solar radiation (W/m$^2$)</td>
<td>Input</td>
</tr>
<tr>
<td>$Q_{\text{dif}}$</td>
<td>Diffuse solar radiation (W/m$^2$)</td>
<td>Input</td>
</tr>
<tr>
<td>$Q_{\text{fan}}$</td>
<td>Supply fan heat gain (J)</td>
<td>Input</td>
</tr>
</tbody>
</table>

Solar air temperature ($T_{\text{sol-air}}$) is a variable used to determine the total heat gain through opaque exterior surfaces to calculate cooling load of a building. It is not a direct measurement from any weather station but need to be computed from outside temperature using eq.(3) for façade with different orientations [20]:

$$T_{\text{sol-air}} = T_a + \frac{\alpha I - \Delta Q_{IR}}{h_a} \quad (3)$$

Where, $\alpha$ is absorptivity of an opaque wall; $I$ is the global solar irradiance (W/m$^2$); $\Delta Q_{IR}$ is extra infrared radiation due to difference between the external air temperature and the apparent sky temperature (W/m$^2$); and $h_a$ is the heat transfer coefficient for radiation (long wave) and convection (W/m$^2$K).

Direct solar radiation and diffuse solar radiation are used to estimate the building heat gain due to the solar transmission through windows. They can be either obtained from weather forecasting information or calculated from global solar irradiance. In the real application, building cooling energy ($E_{\text{zone}}$), outdoor air temperature ($T_{\text{out}}$), zone temperature ($T_{\text{zone,i}}$), occupancy schedule ($R_{\text{in},i}$), and fan heat gains ($Q_{\text{fan}}$) are usually available in building automation system or easy to measure. All the solar information can be obtained from National Solar Radiation Database [21].

All of these input variables are categorized into two groups: unexcited and excited inputs. The zone temperature, and equipment and occupancy schedule ratio are excited inputs, which will be excited during a model training period, and the other variables are unexcited inputs, which values are changed naturally. Using the simulation test bed, the zone temperature is excited by changing the zone temperature setpoints. And the equipment and occupancy schedules are excited by updating their on/off schedules in the simulation test bed. The reason for adding excitements is that the common building operation range is too narrow to provide enough training data.

**System Excitation Signal Generation**

Sum of sinusoids (SINE) model is used to generate the exciting signals (eq. (4)), because sinusoids signals are versatile periodic and can adjust signal shape and character of the power spectrum by adjusting their parameters.

$$U_{t+1} = U_t + \sqrt{2a_s} \sin(\omega_t t + \varphi_t) \quad (4)$$

Where $U_{t+1}$ is the excitation signal; $\sqrt{2a_s}$ is a magnitude scale parameter from 0 to 1; $\omega$ is periodic frequency parameter from 0 to $2\pi$; $T$ is the sampling time, and $\varphi$ is the phase lag parameter from 0 to $2\pi$, which do not affect the signal spectrum. Lowering $T$ will result in a higher frequency bandwidth [22]. Notice that

$$\frac{1}{\beta_s \tau_{\text{dom}}} \leq \omega \leq \frac{\alpha_s}{\tau_{\text{dom}}} \quad (5)$$

where, $\tau_{\text{dom}}$ and $\tau_{\text{dom}}$ correspond to the high and low estimates of the dominant time constant of the system (denote the slowest and the fastest systems time constants) [22]. $\alpha_s$ and $\beta_s$ are user-decisions on high and low frequency content based on identification requirement. The procedure of excitation signals design is summarized in Figure 1.

![Figure 1. Excitation signal generation procedure](image)

The response time constant of a dynamic system is a measure of how quickly the system responds to an input change. It is usually determined by experiments. For example, the impulse response of a dynamic system can be expressed as:

$$x(t) = (\alpha/T)e^{-t/T} \quad (4)$$

where, $T$ is the response time constant, $\alpha$ is a state parameter. The response time for the system output, $x(t)$, to reach 95% of its final steady state value after an input change, is defined as $T_{95}$. For the building in this studied project, $\tau_{\text{dom}} = 360$ minutes, and $\tau_{\text{dom}} = 30$ minutes; $\alpha$ determines the high frequency content in the excitation signal and represents the response speed. $\beta_s$ specifies low frequency information.
corresponding to the system settling time. \( \alpha_s \) and \( \beta_s \) are chosen to be 2 and 3 for the 95\% of the system settling time, respectively. For the temperature setpoint excitation signals in this project, \( T_{max} = 32^\circ C (90^\circ F) \) and \( T_{min} = 10^\circ C (50^\circ F) \); while for the schedule ratio excitation signals, \( R_{max} = 1 \) and \( R_{min} = 0 \).

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 and shown in Figure 2, where all the excitation signals are injected into the building model every 30 minutes.

![Figure 2. Building operation excitement](image)

**Building Energy System Identification Model**

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 3 shows the model development process from building operation data. In this figure, \( U \) is training inputs, \( h \) is a reference signal to analyze the input data, \( Y \) is training outputs data, \( PSD \) is power spectral density model for inputs data and the reference signal, and in this study \( h \) is Welch spectrum object, \( CPSD \) is cross power spectral density model for input and output. \( Suu \) is the result of \( PSD \), \( Suu \) is the result of \( CPSD \), \( S_{uu} \) and \( S_{yu} \) estimate the correlation between input and output. \( G(z) \) is the transfer function in frequency domain, which can be transferred to time domain transfer function \( G(t) \) using inverse Fourier function transformation, and \( \hat{y} \) is the output estimation. \( G(t) \) will be saved as a set of Markov parameters to capture the relationship between each input and output variables.

Power spectral density (PSD and CPSD) describes how the power of a signal or time series is distributed over the frequency spectrum, which is a property of the system signal and very useful in frequency domain system identification [15]:

\[
S_{uu}(k) = \frac{1}{T} \sum_{t=1}^{T} R_{uu}(\tau) e^{-j2\pi k \frac{\tau}{T}}
\]

(4)

\[
S_{yu}(k) = \frac{1}{T} \sum_{t=1}^{T} R_{yu}(\tau) e^{-j2\pi k \frac{\tau}{T}}
\]

(5)

Where, \( R_{uu} \) is the auto-correlation between the inputs and \( R_{yu} \) is the cross-correlation between input and output, and \( l \) is the length of the sampling data. \( l \) is a very important parameter which affect the estimation accuracy and speed, because within one data sample, the power density is calculated simultaneously.

![Figure 3. System identification model](image)

**Markov Parameters from System Identification Model**

The building studied in this project is a light mass building, whose system response time is relatively short. A parametric experiment test has been conducted to find out the best injection frequency. Finally, 30 minutes with 6 hours has been chosen for excitation frequency and sampling length, respectively. The time step for state updating is 15 minutes in the system identification model. Figure 4 shows the Markov parameters for each input variable, using the excitation and response time-history data during the sampling window. All the Markov parameters are close to 0 at the end of the sampling period, which means the sampling length is longer enough to capture the influence on the output from all the inputs. These ten series of Markov parameters is then used in the state space model formation and Kalman filter implantation.
Building Energy State Space Model Development

Recommended by study [23] that the system identification model developed previously is stored as a set of Markov parameters. In order to utilize Kalman filter into this model, however, state space model format is needed. The general linear discrete-time system state space model can be expressed as stochastic difference equations:

\[
x_k = A_{k-1}x_{k-1} + B_{k-1}u_{k-1} + w_{k-1} \\
y_k = C_kx_k + v_k
\]

(6)

(7)

Where \( x \) is the state, \( u \) is the input, and \( y \) is the measurement of the system. Process noise \( w \) and measurement \( v \) are white, zero-mean, uncorrelated random noises.

![System identification model (Markov parameters)](Image)

Figure 5. State space model through ERA

In order to transit Markov parameters into a state space format dynamic model, an eigensystem realization algorithm (ERA) developed by Juang and Pappa [24] is applied. The overall procedure of state space model formation through ERA is illustrated in Figure 5. The details of Hankel matrix generation, state matrix formation and system order determining, etc. are introduced as follows.

From eq. (6) and eq. (7), the Markov parameters can be expressed as:

\[
MP_k = CA^{k-1}B
\]

All these Markov parameters \( MP_1, MP_2, MP_3, \ldots \) are calculated in the frequency response model (Fig. 4) without explicit knowledge of the system matrices \( A, B, C, \) and \( D \). ERA is started from a system realization matrix: Hankel matrix, which is composed of the Markov parameters:

\[
H(k-1) = \begin{bmatrix}
MP_k & MP_{k+1} & \cdots & MP_{k+\beta-1} \\
MP_{k+1} & MP_{k+2} & \cdots & MP_{k+\beta} \\
\vdots & \vdots & \ddots & \vdots \\
MP_{k+\alpha-1} & MP_{k+\alpha} & \cdots & MP_{k+\alpha+\beta-2}
\end{bmatrix}_{\alpha \times \beta r}
\]

Where \( \alpha \) and \( \beta \) are determined by the number of Markov parameters from the system identification model, which should be greater than system order \( n \).

Since building energy model is a dynamic model, whose system order could be changing according to the system operation situation (for example during starting up and shutting down period). Therefore, singular value decomposition (SVD) was then applied on \( H(k) \) to determine the system order and state space model parameters:

\[
H(k) = R\Sigma S^T
\]

Where \( R \) and \( S \) are orthonormal matrices and the rank of \( \Sigma \) is the order of this system model.

\[
\Sigma = \begin{bmatrix}
\Sigma_n & 0 \\
0 & 0
\end{bmatrix} \text{ with } \Sigma_n = \text{diag} \{\sigma_1, \sigma_2 \ldots \sigma_n\}
\]

Therefore, the parameter \( s \) of the state space mode can be obtained from the following equations:

\[
\hat{A}_k = \Sigma_n^{-\frac{1}{2}}R_n^T H(k) S_n \Sigma_n^{-\frac{1}{2}} \\
\hat{B}_k = \Sigma_n^{-\frac{1}{2}}S_n E_r \\
\hat{C}_k = E_n^T R_n \Sigma_n^{-\frac{1}{2}}
\]

Where \( E_r = [I_r \ O_r \ \ldots \ O_r] \) and \( E_n = [I_m \ O_m \ \ldots \ O_m] \), where \( I \) is identical matrix, \( r \) is the number of inputs and \( m \) is the number of outputs. \( R_n \) and \( S_n \) are the first \( n \) columns if matrices \( R \) and \( S \), respectively. \( \hat{A}_k, \hat{B}_k \) and \( \hat{C}_k \) are the estimation of state space model parameters: \( A_k, B_k \) and \( C_k \).

The order of model is changing according to the energy system operation condition changing. System at different partial loads will have different characteristics and nonlinearity, and especially, during the starting up and shutting down periods the system nonlinearity changes very fast.

![Kalman Filter Based Data Fusion Implantation](Image)

Figure 6. Kalman filter state estimation

Kalman Filter Based Data Fusion Implantation

Recall the state space equations in eq. (6) and eq. (7). Noises \( w_k \) and \( v_k \) have known covariance matrices \( Q_k \) and \( R_k \), respectively. \( Q_k \) is determined by the discrepancy of the forecasted and real energy consumption. \( R_k \) is determined by the measurement noise. In this study, simulation results from the validated EnergyPlus model were used as the ground truth (real energy consumption), and measurement error stand deviation was chosen as 10%:

\[
w_k \sim N(0, Q_k)
\]
\[ v_k \sim N(0, R_k) \]

The concept of Kalman filter is to estimate the state of a dynamic system by using a feedback process: a state space filter estimates the process and receives feedback from noise measurements (Figure 6). The state space model will forecast forward (discrete time) from the current state and process covariance \( Q_k \) to get a priori estimation of the state at next time step. The noisy measurements will be incorporated in the system model to update the posterior forecasting (Figure 7).

In this study, a system order auto checking model has been developed before state space model generation and Kalman filter implementation. In the state space model, state vector \( X_k = [E_k E_{k-1} \ldots E_{k-n}] \) (\( n \) is system order, \( E_k \) is the cooling energy consumption at time \( k \)), measurement vector \( Y_k = [E_k] \), and input vector contains all the input variable in the system identification model, as described in Table 1.

Follow the Kalman filter implementation procedure discussed before, the process and measurement covariance are determined as:

\[ Q_k = I(n) \cdot \text{stderr}_k^2 \]
\[ R_k = (E P_k \cdot \sigma \cdot \text{rand}(-1, 1))^2 \]

where \( I(n) \) is a \( n \times n \) identity matrix, \( \text{stderr}_k \) is the standard deviation of the state space model forecasting error, \( E P_k \) is the EnergyPlus cooling energy simulation results at step \( k \), \( \sigma \) is the measurement noise: 10\%, and \( \text{rand}(-1, 1) \) is a random number between -1 and 1. Then Kalman filter is applied to update the state space model and to improve the forecasting accuracy every time step (15 minutes).

**BUILDING ENERGY FORECASTING RESULTS AND DISCUSSION**

**Building Description**

The mid-size commercial building studied in this project is a single story office building (Figure 8), 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.

![Figure 8. Small commercial building view](image)

The HVAC systems used in this building are constant-air-volume (CAV) air handling units (AHUs) with direct expansion (DX) coils. The coefficient of performance (COP) of the cooling system is 3. Heating is provided by electricity with an efficiency of 0.95. The baseline model internal load inputs are summarized in Table 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupant Density</td>
<td>0.005 person/square foot</td>
</tr>
<tr>
<td>Ventilation Requirement</td>
<td>26.5 CFM/person</td>
</tr>
<tr>
<td>Lighting Power Density</td>
<td>1.8 watts/square foot</td>
</tr>
<tr>
<td>Interior Small Plug Loads</td>
<td>1.0 watts/square foot</td>
</tr>
<tr>
<td>Elevator Consumption</td>
<td>32,000 watts</td>
</tr>
<tr>
<td>Exterior Lighting</td>
<td>18,000 watts</td>
</tr>
<tr>
<td>Envelope Infiltration Rate</td>
<td>0.223 CFM/square foot</td>
</tr>
</tbody>
</table>

**Building Energy On-line Estimation**

As introduced early, a validated EnergyPlus simulation test bed is used as a “real” building to provide training data and validation data for both heating and cooling seasons. The
overall procedure of the generation and validation of the on-line energy estimation model is shown in Figure 9. First, frequency response function is applied to create a system identification model to forecast the energy consumption, and then the ERA method is used to reformat the system identification model to be a state space model, which is also used to forecast the energy consumption. At the last step, Kalman filter is implemented to update the energy consumption forecasting based on the “real” measurements. There are two sets of validation data, one is used to validate the system identification model and state space model and the other is used as “real” measurements to update and validate the Kalman filter on-line model. The difference between these two sets of validation data is that “real” measurements are corrupted with measurement noises. The methods used in each step have been introduced in previous sections, as the results will be discussed in following ones.

\[ y_k = EP_k + EP_k \cdot \sigma \cdot \text{rand} \left( -1, 1 \right) \]

Where, \( y_k \) is energy consumption measurement, and \( EP_k \cdot \sigma \cdot \text{rand} \left( -1, 1 \right) \) is measurement noise.

**System Identification and State Space model Estimation Results**

**System Identification Forecasting Results**

The system identification model developed based on the July training data is then used to forecast the whole building cooling energy consumption in the same training period. As shown in Figure 12, the model is able to capture the overall trend of building cooling energy from EnergyPlus simulation test bed. The \( R^2 \) is 0.94 for the entire 10 days of training data.

**Training and Validation Data Generation**

The system excitation signals discussed before was modeled and generated in Matlab. In order to apply the excitation into EnergyPlus model, building controls virtual test bed (BCVTB) is used to inject excited building heating and cooling setpoint, internal load schedules in to EnergyPlus model. BCVTB is a middleware tool that can transfer data between different simulation programs [6]. The Matlab-BCVTB-EnergyPlus connection is illustrated in Figure 10. During the entire study, typical meteorological year (TMY) weather data for Philadelphia is used. Excitation is applied for 10 days, i.e. (July 01 to July 10) with a duration of 24 hours per day to generate training data.

After the training data are generated, 3 days validation data are also generated using typical commercial building control and operation schedules. The validation data include 3 days, i.e., (July 11 to July 13). The typical control and operation schedules used in generating validation data are shown in Figure 11. In this study, in order to represent the real situation and apply data fusion techniques, a white measurement noise is added to the output from Energy Plus test bed:

<table>
<thead>
<tr>
<th>Table 3 System identification model cooling energy results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy (R²)</strong></td>
</tr>
<tr>
<td><strong>Training Period</strong></td>
</tr>
<tr>
<td>Whole Building</td>
</tr>
</tbody>
</table>

Figure 11. Testing control signals setting

Figure 12. Building energy training results
The system identification model is then used to forecast whole building cooling energy consumption for three days (July 11 to 13). The results together with simulated results from the test bed are illustrated in Figure 13a-b. They illustrate the comparison of EnergyPlus simulated whole building energy consumption ($E_{es}$) and System identification model ($SID$) forecasted while building energy consumption ($E_{ep}$). Due to the underestimation of direct solar radiation related cooling energy consumption in the afternoons, when cooling load is high, discrepancy between $E_{es}$ and $E_{ep}$ exists in the afternoons. Even with this underestimation, the overall forecasting accuracy is still acceptable. As Table 3 shows, the forecasting period accuracy ($R^2=0.975$) is higher that of training period ($R^2=0.944$). That is because building energy system has much higher dynamics during the training period when excitation signals are applied.

![Figure 13](image.png)

Figure 13. System identification model results: a) Building energy forecasting results; b) Energy forecasting error analysis

State Space Model Estimation Results

By using the ERA method that discussed before, the system identification model is then regenerated to be a state space model based on the reformation of Markov parameters. The order of the state space model varies from 4 to 36 in this study, due to the order of dynamics of the building energy system at different operation situations.

The state space model sacrificed some accuracy due to the truncation of the Markov parameter according to the system order. As it is shown in Figure 14a-b, the error between forecasting results from state space model ($E_{es}$) and EnergyPlus simulation results ($E_{ep}$) are much larger than that in Figure 13. The overall $R^2$ is just 0.71. However, the state space model is still able to follow the trend of the EnergyPlus estimation results, and it does not under estimate the cooling energy at noon of the third day when cooling energy consumption is very high. In the next step, Kalman filter based data fusion techniques will then be applied to improve the forecasting accuracy and robustness.

![Figure 14](image.png)

Figure 14. State space model results: a) Energy forecasting results; b) Energy forecasting error analysis

Kalman Filter Based Data Fusion On-line Estimation Results

Upon the state space model, noisy measurements have been used to update the energy forecasting. The forecasting results from the Kalman filter based on-line energy estimation model are shown in Figure 15a-c. The forecasting days are from July 11 to 13. The Kalman filter is updated on every 15 minutes for the next 15 minutes. The results show that the on-line model achieved better forecasting accuracy than the original system identification model, especially at the HVAC system starting up and turning down periods. The overall $R^2$ is 0.98. What’s more, the on-line model improved the energy
consumption forecasting accuracy at the high cooling load period. In Figure 15b, the green line (Emea) is the real energy consumption measurement with noise.

![Simulation Speed](image)

Simulation speed is another crucial factor for building energy forecasting model application in real field. The forecasting speeds and accuracy of these three models are also summarized in Table 4. The system identification model training time is around 48 second, as show in Table 3. Fortunately, in any on-line estimation, energy forecasting model doesn’t need to be trained every time step. In this study, this model is just trained once for three days’ forecasting, once the Markov parameters were calculated from the training data, they were saved and used by the on-line forecasting every step. Therefore, from calculation point view, this on-line energy forecasting model is much battery than EnergyPlus, which usually takes more than 2 minutes every time step in the forecasting periods.

Table 4 Model simulation accuracy and speed comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>System Identification</th>
<th>State Space</th>
<th>Data Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>0.98</td>
<td>0.71</td>
<td>0.97</td>
</tr>
<tr>
<td>Speed (s)</td>
<td>0.0089</td>
<td>0.03</td>
<td>0.02</td>
</tr>
</tbody>
</table>

CONCLUSION

This study introduced a new systematic methodology for on-line building energy estimation model development and validation, based on frequency response system identification and data fusion. A system excitement scheme is developed for system identification model development. This excitement scheme is able to generate enough data at both high and low operation dynamics around the building operation range, which can be applied to other system identification models for building energy forecasting. Frequency response function model realized by power spectral density model was implanted to develop the system identification model and forecasting building cooling energy consumption. Eigensystem realization algorithm (ERA) was used to reformat the energy forecasting model to be state space model from Markov parameters. Finally, Kalman filter is applied to the state space model to update the energy forecasting with a frequency of 15 minutes and improve the forecasting accuracy and robustness. This on-line building energy model can achieve over 95% forecasting accuracy within two minutes calculation time (training and forecasting) for three days energy forecasting for cooling energy consumption in summer of the mid-size commercial building studied here. Future efforts are needed to shorten the excitation period and to examine the developed methodologies for very different types of buildings and HVAC systems, and for real world conditions.

ACKNOWLEDGMENTS

Financial support provided by the U.S. National Science Foundation Award 1239247 is greatly appreciated.
REFERENCES

ANNEX A