2022

Hybrid Modeling Approach For Better Identification Of Building Thermal Network Model And Improved Prediction

Sang woo Ham

Donghun Kim

Hybrid modeling approach for better identification of building thermal network model and improved prediction

Sang woo Ham[1*], Donghun Kim[2]

[1] Building Technology & Urban System Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States of America
sham@lbl.gov

[2] Building Technology & Urban System Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States of America
donghunkim@lbl.gov

* Corresponding Author

ABSTRACT

The gray-box modeling approach (i.e., semi-physical thermal network model) has been widely used for prediction applications for buildings such as a model predictive control (MPC). However, applying the modeling approach for practical buildings is still challenging due to unmeasured disturbances such as occupants, lighting, appliances, and in/exfiltration loads. To overcome this problem, several system identification approaches have been proposed by considering the dynamics of unmeasured disturbance. However, the performance of long-term (e.g., one day) zone temperature or load predictions could still be very poor, and this is an important research topic for enabling grid-interactive buildings. In this study, we propose a hybrid modeling approach to improve long-term temperature or load predictions. Several system identification approaches for gray-box models are compared using simulations to understand the limitations. A neural network model that accounts for unmeasured disturbance is developed by considering the limitation of the gray-box model and is combined with the gray-box model. This hybrid model approach shows 0.24°C root mean squared error (RMSE) for 1-day ahead temperature prediction on average, while the conventional gray-box model shows 1.1°C RMSE on average.

1. INTRODUCTION

The gray-box modeling approach (i.e., semi-physical thermal network model) has been widely used for prediction applications for buildings such as model predictive control (MPC) because it has a flexible structure representing a complex building with a low order R-C (thermal resistance and capacitance) network model from measured data while keeping its physical principles. However, in real buildings, some information for the gray-box model such as internal heat gains from occupants, lighting, appliances, and in/exfiltration is generally not available due to the lack of sensors, and the retrofit of these sensors for the unmeasured disturbance is costly prohibited especially in small and medium-sized commercial buildings (SMCBs).

Accordingly, a conventional system identification for a real-building, which minimizes the sum of squared error between *n*-step ahead simulation and measurement (Braun and Chaturvedi, 2002), often results in significant bias in parameters, especially the unmeasured disturbance is not negligible (Kim et al., 2016). It could yield a critical problem in predictive application. For example, the unmeasured disturbance could be represented as unrealistic large thermal capacitance and solar heat gains in the conventional system identification, so the predictive controller's performance (e.g., load shift) may be wrongly executed.

Several system identification approaches (Coffman and Barooah, 2018; Kim et al., 2016) have been proposed to correctly estimate model parameters of a gray-box model by handling the internal heat gains as a lumped unmeasured disturbance. Although they showed better identification performance, the success of the identification highly depends on the quality of data (i.e., uncorrelated control inputs and measured/unmeasured disturbances)(Kim et al., 2018). Furthermore, the unmeasured disturbance information is not given for the future prediction, so the predictive performance is poor regardless of the accuracy of identified system parameters.

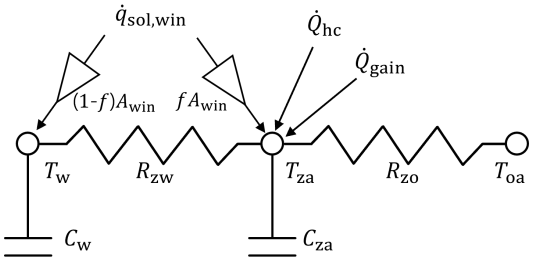


Figure 1: A simple RC network model for a case study building.

In building simulation, the unmeasured disturbances (i.e., occupant, appliance, and infiltration heat gains) are generated by time-based scheduled values in deterministic (ASHRAE, 2017) or stochastic ways (Hong et al., 2020). This approach is valid in a predictive application for white-box simulation because the goal is to estimate typical energy consumption based on typical unmeasured disturbance schedules for building design. However, the unmeasured disturbance needs to be estimated from the data in a real building in operation. While it can be easily obtained via a Kalman (Coffman and Barooah, 2018; O'Neill et al., 2010) or a particle filter (Ham et al., 2021) from the measured data, these methods do not provide a future prediction. Some studies (Dong et al., 2016; Lee and Zhang, 2021) developed a black-box model to predict non-HVAC energy consumption from historical data and use it to quantify unmeasured disturbances. One study models unmeasured disturbances as a function of time-related features (i.e., time of the day, day of the week, and day of the year) via feed-forward neural network (Ellis, 2021). The model is trained after estimating the disturbance via a Kalman filter or together with a building gray-box model. Though previous studies show good prediction results on their data, it is necessary to investigate more in-depth investigation on the impact of the gray-box model quality and the proper structure of the black-box model because the gray-box model quality is not guaranteed in the real building due to the lack of sensors and low data quality.

In this research, we propose a hybrid approach that integrates a neural network-based black-box model for unmeasured disturbances with a gray-box model for a predictive application. First, we provide a simulation case study to show how the conventional identification algorithm poorly behaves under unmeasured disturbances and propose and compare alternative identification algorithms to overcome this issue. After investigating the limitation of alternative identification algorithms in a predictive application, we propose a neural network model by considering the gray-box model structure. Finally, the impact of input features and model types on the performance is investigated with actual prediction performances.

2. COMPARISONS OF IDENTIFICATION APPROACHES UNDER UNMEASURED DISTURBANCE AND EFFECT ON PREDICTION: SIMULATION CASE STUDY

For real buildings, the various heat gain information (e.g., occupants, lighting, appliances, and in/exfiltration loads) is often not available. and it often leads to failure of the conventional system identification approach (CONV) (Braun and Chaturvedi, 2002), which minimizes a norm of errors between simulation and measurement, despite its popularity in the literature (Kim et al., 2016; Kim et al., 2018). Different identification approaches are needed to overcome this problem. We provide a simulation case study to show how the conventional identification algorithm poorly behaves under unmeasured disturbances and propose and compare alternative identification algorithms to overcome this issue. The proposed algorithms model not only the thermal dynamics of a building envelope system but also the dynamics of unmeasured disturbances in the form of either a lumped input disturbance (ID) (Coffman and Barooah, 2018) or output disturbance (OD) (Kim et al., 2016; Kim et al., 2018).

2.1 True system description and data generation

It is very difficult to compare the performance of identification approaches for a real building since we do not know the true system dynamics or parameters. Instead, we created a hypothetical building envelope model and treated it as the *True model* for this study. The True model was tuned with a data set that was generated from a laboratory environment (FLEXLAB) (Lawrence Berkeley National Laboratory, 2021). The laboratory represents a single-zone office space. The True model has the RC network shown in Fig. 1 and the state-space form of Eq. 1 (state transition process and measurement process).

$$\underbrace{\begin{bmatrix}\dot{T}_{\mathrm{w}}\\ \dot{T}_{\mathrm{za}}\end{bmatrix}}_{\dot{\mathbf{x}}} = \underbrace{\begin{bmatrix}\frac{-1}{C_{\mathrm{w}}R_{\mathrm{zw}}} & \frac{1}{C_{\mathrm{w}}R_{\mathrm{zw}}}\\ \frac{-1}{C_{\mathrm{za}}R_{\mathrm{zw}}} & \frac{-1}{C_{\mathrm{za}}R_{\mathrm{zw}}}+\frac{-1}{C_{\mathrm{za}}R_{\mathrm{zo}}}\end{bmatrix}}_{\mathbf{A}}\underbrace{\begin{bmatrix}T_{\mathrm{w}}\\ T_{\mathrm{za}}\end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix}0 & \frac{(1-f)A_{\mathrm{win}}}{C_{\mathrm{w}}} & 0 & 0\\ \frac{1}{C_{\mathrm{za}}R_{\mathrm{zo}}} & \frac{fA_{\mathrm{win}}}{C_{\mathrm{za}}} & \frac{1}{C_{\mathrm{za}}} & \frac{1}{C_{\mathrm{za}}}\end{bmatrix}}_{\mathbf{B}}\underbrace{\begin{bmatrix}T_{\mathrm{oa}}\\ \dot{q}_{\mathrm{sol,win}}\\ \dot{Q}_{\mathrm{hc}}\\ \dot{Q}_{\mathrm{gain}}\end{bmatrix}}_{\mathbf{u}} \tag{1}$$

$$y_{\mathrm{za}} = \underbrace{\begin{bmatrix}0 & 1\end{bmatrix}}_{\mathbf{C}}\mathbf{x}$$

where T_{w}, T_{za}, and T_{oa} are temperature nodes of large thermal mass (w), zone air (za), and outdoor air (oa), R_{zw} and R_{zo} are thermal resistances between temperature nodes [K/kW], C_{w} and C_{za} are thermal capacitances [kWh/K], $\dot{q}_{\mathrm{sol,win}}$ is incident solar radiation per area on windows [kW/m^2], A_{win} is effective window area [m^2], f is convective fraction of the incident solar radiation, $\dot{Q}_{\mathrm{hc}}$ is heating/cooling rate from maximum heating/cooling rate ($\dot{Q}_{\mathrm{heat,max}}/\dot{Q}_{\mathrm{cool,max}}$) times signal ($i_{\mathrm{heat}}/i_{\mathrm{cool}}$) [kW], $\dot{Q}_{\mathrm{gain}}$ is internal heat gain (e.g., summation of plug, lighting, and occupancy loads) [kW], and y_{za} is measured zone air temperature [°C].

The parameters to be estimated through the system identification are $\theta = [C_{\mathrm{w}}, C_{\mathrm{za}}, R_{\mathrm{zw}}, R_{\mathrm{zo}}, f, A_{\mathrm{win}}]$ and were tuned with the complete measurements including $\dot{Q}_{\mathrm{gain}}$. The values of the True model were set to C_{w}: 4.0 kWh/K, C_{za}: 1.0kWh/K, R_{zw}: 1.2 K/kW, R_{zo}: 9 K/kW, f: 0.3, and A_{win}: 3.0 m^2. The maximum heating and cooling rates ($\dot{Q}_{\mathrm{heat,max}}$ and $\dot{Q}_{\mathrm{cool,max}}$) are set to 6kW and −6kW, respectively. Many SMCBs are controlled by a on/off thermostat controller with its own thermostat setting such as minimum cycling time. To simulate this behavior, an ideal on-off heating/cooling controller that sending heating and cooling operation signal (i_{heat} and i_{cool}, respectively) with minimum cycling time of 5-minute is implemented. Then, the data is resampled 15-minute interval through centered moving average.

To test the performance of the identification algorithms of the CONV, ID and OD under a practical scenario, $\dot{Q}_{\mathrm{gain}}$ is assumed to be unknown. Instead, it is assumed that we can design experiments and actively control indoor temperature setpoint (within an acceptable room air temperature band) for the purpose of better system identification. Two weeks of data were generated from the True model with Oakland, CA, weather data. During the weekdays, cooling and heating setpoints are set to 25°C and 21°C for occupied times (6 – 18 hour) and 30°C and 16°C for the unoccupied time. However, the setpoint of the first weekend (2 days) was perturbed according to a pseudo-binary random signal (PRBS) with 2-hour of time-scale and 4th order. The binary signal is mapped to sampled setpoints between 18 – 26°C. The generated data set is visualized in Fig. 2.

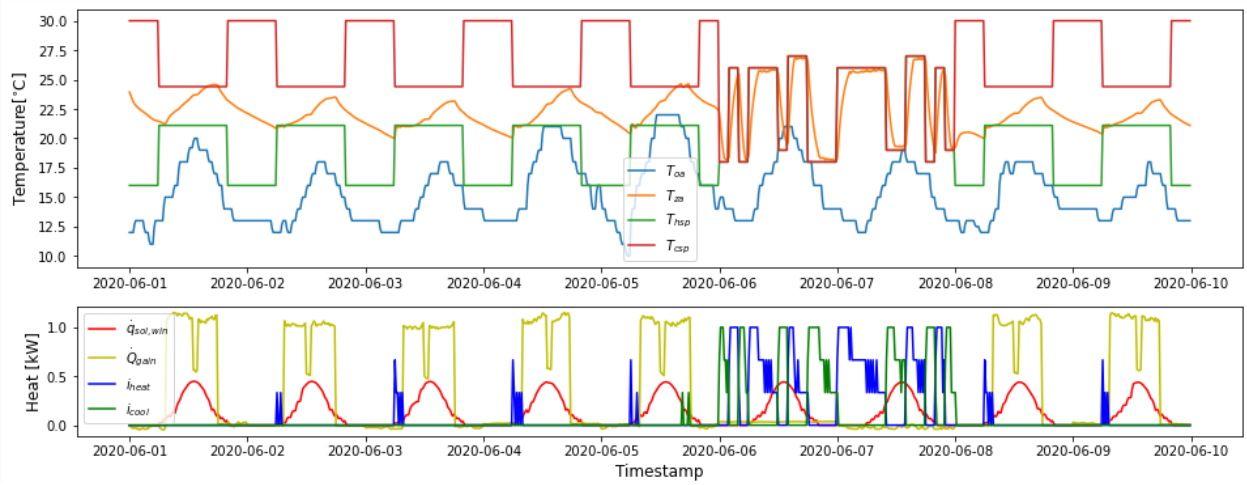


Figure 2: Temperatures and measured/unmeasured disturbances/resampled control signals of simulated data.

2.2 Summary of system identification approaches

Eq. 1 was discretized in a 15-min sampling time by using the zero-order hold (Rouchier et al., 2019). Three different system identification approaches (CONV, ID, and OD) were tested with the test data set aiming at identifying the True

model. They have the same model structure for the building envelope dynamics, although their disturbance model structures differ. Note that the unmeasured disturbance ($\dot{Q}_{\text{gain}}$) was not provided to the identification algorithms. In other words, the identified model G maps only the measured disturbances (T_{oa}, $\dot{q}_{\text{sol,win}}$) and control input ($\dot{Q}_{\text{hc}}$) to the indoor air temperature (y_{za}). In the following paragraphs, we will introduce the basic concept of each identification approach because of the lack of space.

Conventional simulation error minimization approach: The CONV approach (Braun and Chaturvedi, 2002) assumes that all unmeasured disturbances can be expressed as white noise (e_{conv}) in the measurement process. The discretized system can be written as Eq. (2);

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}_{\text{d}}\mathbf{x}(k) + \mathbf{B}_{\text{d}}\mathbf{u}(k) \\ y(k) &= \mathbf{C}_{\text{d}}\mathbf{x}(k) + e_{\text{conv}}(k) \end{aligned} \tag{2}$$

where $\mathbf{A}_{\text{d}}$, $\mathbf{B}_{\text{d}}$, and $\mathbf{C}_{\text{d}}$ from $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in Eq. (1).

The set of parameters (θ^*_{conv}) is estimated by minimizing the sum of squared errors between simulation ($\hat{y}(k;\theta) = \mathbf{C}_{\text{d}}\hat{\mathbf{x}}(k;\theta)$) and measurement ($y(k)$) via the nonlinear optimization (Eq. 3);

$$\begin{aligned} \hat{\mathbf{x}}(k+1;\theta) &= \mathbf{A}_{\text{d}}(\theta)\hat{\mathbf{x}}(k;\theta) + \mathbf{B}_{\text{d}}(\theta)\mathbf{u}(k) \\ y(k) &= \mathbf{C}_{\text{d}}\hat{\mathbf{x}}(k;\theta) + \varepsilon_{\text{conv}}(k;\theta) \\ \theta^*_{\text{conv}} &= \arg\min_{\theta} \sum_{k=1}^{N} \left(\varepsilon_{\text{conv}}(k;\theta)\right)^2, \end{aligned} \tag{3}$$

where subscript d indicates a discretized system and, k is a discretized time index.

Input disturbance identification approach: The ID approach assumes that unmeasured disturbances come from the input channel, i.e., the heat gain term, and treats the input disturbance as an additional dynamic state. This can be written as an augmented state space form (Eq. 4) (Coffman and Barooah, 2018).

$$\begin{aligned} \underbrace{\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\zeta}_{\text{ID}} \end{bmatrix}}_{\dot{\mathbf{x}}_{\text{ID}}} &= \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{A}_{\zeta_{\text{ID}}} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\mathbf{A}_{\text{ID}}} \underbrace{\begin{bmatrix} \mathbf{x} \\ \zeta_{\text{ID}} \end{bmatrix}}_{\mathbf{x}_{\text{ID}}} + \mathbf{B}\mathbf{u} + \mathbf{w}_{\text{ID}}, \text{ and } \mathbf{A}_{\zeta_{\text{ID}}} = \begin{bmatrix} \mathbf{0} \\ \frac{1}{C_{\text{za}}} \end{bmatrix} \\ y_{\text{za}} &= \underbrace{\begin{bmatrix} \mathbf{C} & 0 \end{bmatrix}}_{\mathbf{C}_{\text{ID}}} \mathbf{x}_{\text{ID}} + e_{\text{ID}} \end{aligned} \tag{4}$$

where $\mathbf{w}_{\text{ID}}$ and e_{ID} are state and measurement noises. ζ_{ID} represents the lumped input disturbance term.

The key idea in this approach is to treat ζ_{ID} as Wiener process, so therefore , it behaves as the Brownian motion after discretization according to its noise level (i.e., $\zeta_{\text{ID}}(k+1) = \zeta_{\text{ID}}(k) + w_{\zeta_{\text{ID}}}(k)$). For the system identification, the ID approach calculates one-step ahead prediction errors after filtering states via Kalman filter at each time. More details can be found in Coffman and Barooah, 2018.

Output disturbance identification approach: The OD approach (Kim et al., 2016; Kim et al., 2018) models the effect of unmeasured heat gain on the output (i.e., room air temperature) because the output more smoothly changes compared to the input heat gain. The aggregated contribution of the unknown heat sources to the output is called the output disturbance. The OD approach models the output disturbance as a filtered process of white noise ($e_{\text{OD}}(k)$), which is called output disturbance ($v_{\text{OD}}(k)$ in Eq. 5). The output disturbance dynamics can be modeled with two more parameters, ρ_1 and ρ_2 (Eq. 5).

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}_{\text{d}}\mathbf{x}(k) + \mathbf{B}_{\text{d}}\mathbf{u}(k) \\ y(k) &= \mathbf{C}_{\text{d}}\mathbf{x}(k) + v_{\text{OD}}(k) \\ \zeta_{\text{OD}}(k+1) &= \begin{bmatrix}\rho_1\end{bmatrix}\zeta_{\text{OD}}(k) + \begin{bmatrix}\rho_2\end{bmatrix}e_{\text{OD}}(k) \\ v_{\text{OD}}(k) &= \zeta_{\text{OD}}(k) + e_{\text{OD}}(k) \end{aligned} \tag{5}$$

The OD approach estimates a set of parameters by minimizing the square sum of one-step ahead prediction error. More details for the identification algorithm can be found in Kim et al., 2016.

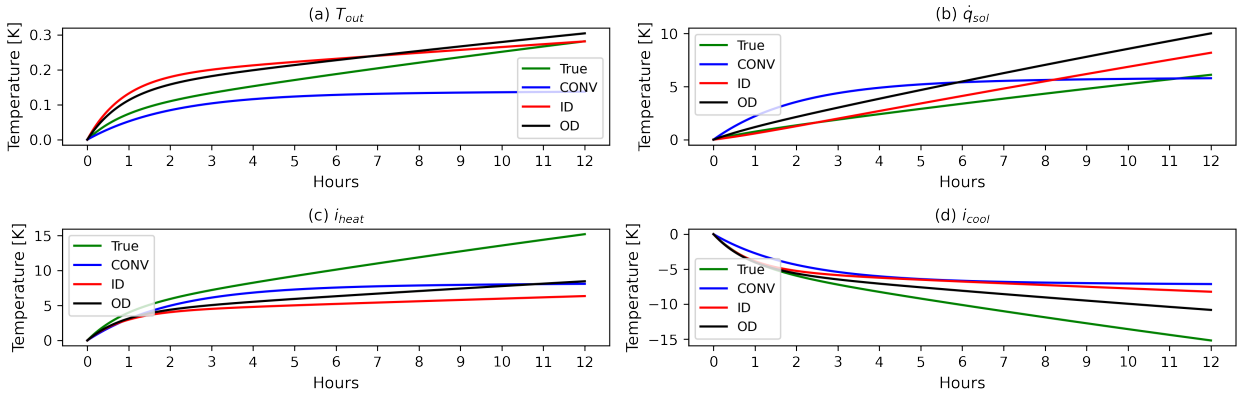


Figure 3: Comparison of step response of measured disturbances (T_{oa} and $\dot{q}_{\text{sol,win}}$) and control input ($\dot{Q}_{\text{hc}}$) for each method.

2.3 System identification results and discussion

For each identification algorithm, the optimization problem is non-convex, and hence to find a better optimal solution, we randomly sampled initial starting points and repeatedly solved the optimization problems 50 times. The estimation results of CONV, ID, OD identification approaches are summarized in Table. 1.

Table 1: Comparison of estimated parameters from system identification approaches

	C_{w} $[\frac{\text{kWh}}{\text{K}}]$	C_{za} $[\frac{\text{kWh}}{\text{K}}]$	R_{zw} $[\frac{\text{K}}{\text{kW}}]$	R_{zo} $[\frac{\text{K}}{\text{kW}}]$	f [-]	A_{win} [m^2]	$\dot{Q}_{\text{heat,max}}$ [kW]	$\dot{Q}_{\text{cool,max}}$ [kW]
True	4.0	1.0	1.2	9.0	0.3	3.0	6.0	-6.0
CONV	40	0.38	6.17	40.0	1.0	1.06	1.48	-1.3
ID	20.63	1.37	0.81	3.37	0.03	20.0	6.7	-8.7
OD	10.58	1.33	0.84	4.14	0.15	12.37	6.72	-8.6

Overall, the estimated R, C, and $\dot{Q}$ values of ID and OD have realistic scale compared to CONV, but all methods wrongly estimate solar-related parameters (i.e., f and A_{win}). A building thermal dynamics is the combination of all the inputs and current state information, so some of wrong parameters are affecting the long-term thermal dynamics of the building. Fig. 3 shows the comparison of step response of measured disturbances and control inputs for each method. While all three methods fail to capture long-term responses, ID and OD show better performance overall, especially for the short-term period. For example, ID and OD capture a short-term dynamics of cooling control input (i_{cool}) for 2 hours.

2.4 Challenges in prediction application

In a predictive application such as model predictive control (MPC), the performance of prediction is critically important to correctly respond to the current grid's status (e.g., price signal). However, it is difficult for a building when the unmeasured disturbances are significant, even if we have a perfect gray-box model. For example, the predicted temperature profiles for daytime are much lower than the measurement if there is no credible internal heat gain prediction. This results in a wrong prediction for the cooling period. This section provides a more detailed simulation study that shows this issue and motivates developing an additional model, namely the hybrid modeling approach described in the following section, to overcome this problem.

In Fig. 4, the 1-day ahead, room air temperature predictions of the True, CONV, ID, and OD models are compared with the measurement. Like before, the heat gain ($\dot{Q}_{\text{gain}}$, marked with the red line in the bottom figure) information was not given to all models. Not surprisingly, all methods show large deviations from the measurement in future prediction, even for the True model. This is because the True model did not receive the heat gains while the physical building did. This wrong prediction is problematic when the predictive application targets daily demand management, such as load shift. Therefore, in addition to the correct parameter estimation, it is also essential to forecast the unmeasured disturbances for the predictive application.

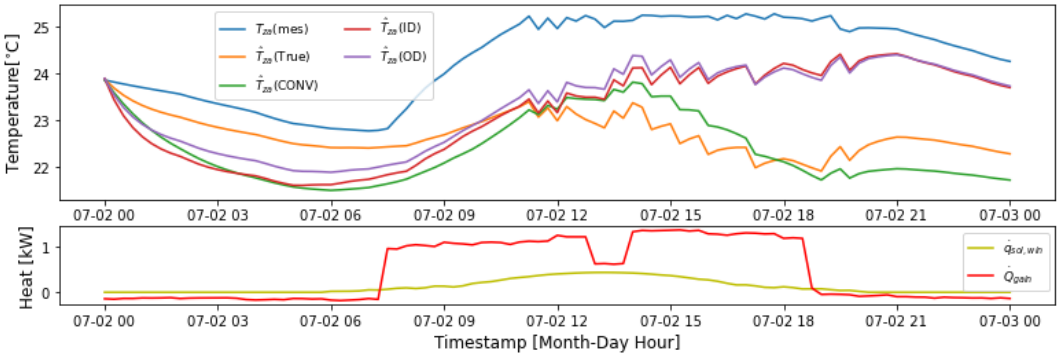


Figure 4: 1-day ahead temperature predictions of different approaches without unmeasured disturbances.

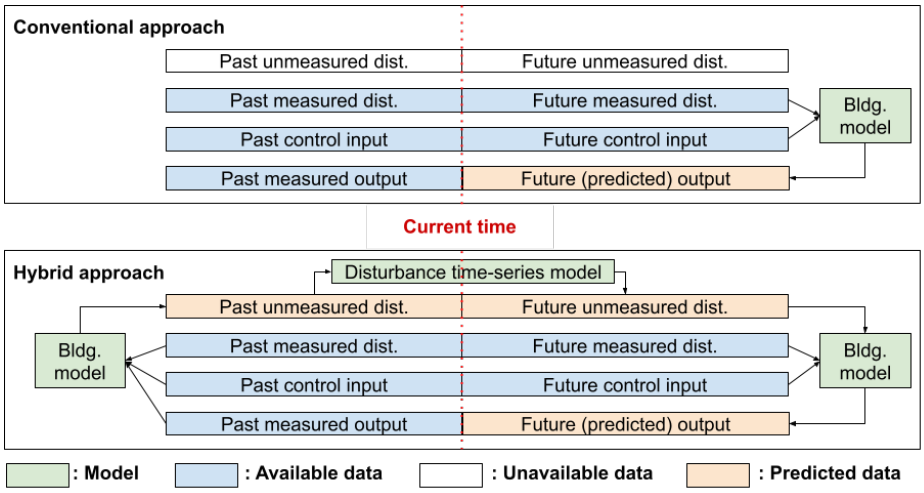


Figure 5: Comparison between Conventional and Hybrid approaches.

3. HYBRID MODELING APPROACH FOR PREDICTIVE APPLICATION

3.1 Overview of Hybrid modeling approach

In this section, we introduce a hybrid modeling approach that combines the OD system identification with a method of forecasting unmeasured disturbances for predictive applications. Fig. 5 shows the schematic diagram of the proposed Hybrid approach compared to the Conventional approach. As shown in Figures 5 and 6 (a), a building thermal model is a transfer function between the inputs (control inputs, measured disturbances, and unmeasured disturbances) and the output (temperature).

In Conventional approach, the building model is used to predict future output (i.e., temperature) by using the future measured disturbance (i.e., forecast of T_{oa} and $\dot{q}_{\mathrm{sol,win}}$) and future control inputs ($\dot{Q}_{\mathrm{hc}}$). However, as discussed in section 2.3 (see Fig. 3), this approach might not provide reliable prediction when the absence of unmeasured disturbances is not negligible.

On the other hand, the Hybrid approach attempts to include future unmeasured disturbance information through a disturbance model. The key steps are as follows. First, a gray-box model is trained through OD system identification. Then, we estimate the past unmeasured input disturbance ($\hat{\zeta}_{\mathrm{ID}}$) at each sampling time by utilizing the ID model (Eq. 5). Once the unmeasured input disturbance is estimated, then we perform the time series modeling explicitly, e.g., via machine learning techniques, to forecast the unmeasured disturbances. In this case, we use unmeasured input disturbance because the prediction errors on the input disturbance are filtered through the gray-box model on temperature prediction, while the prediction errors of the unmeasured output disturbance directly go to the temperature prediction.

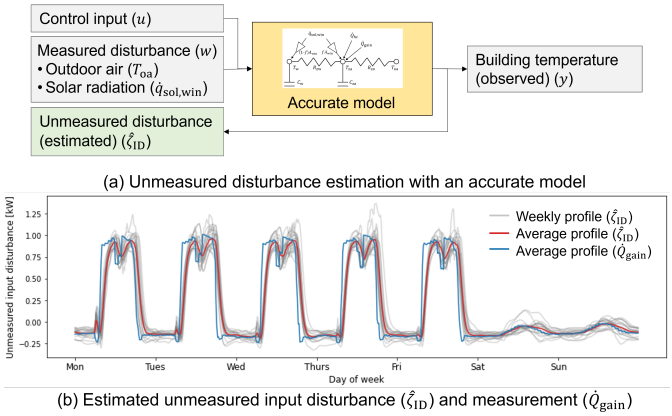


Figure 6: Unmeasured disturbance estimation with the accurate model.

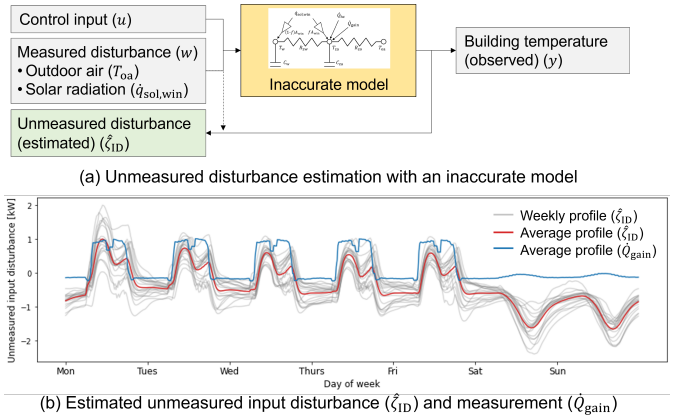


Figure 7: Unmeasured disturbance estimation with the inaccurate model.

3.2 Design of Hybrid model structure

The unmeasured disturbances are mainly composed of internal heat gains from occupants, appliances, and plug loads, and it is typically modeled as a function of daily/weekly schedules (Ellis, 2021; Hong et al., 2020). It is a valid approach when the gray-box model is accurate.

Fig. 6 shows the schematic diagram of unmeasured disturbance estimation ($\hat{\zeta}_{ID}$) with an accurate (True) model (a), and the average of estimated weekly profiles are compared with the average of measured values ($\dot{Q}_{gain}$) (b). Since the model is the True model, the estimation shows good agreement with the true profile.

On the other hand, Fig. 7 shows the schematic diagram of unmeasured disturbance estimation with an inaccurate model (a), and the average of estimated weekly profiles are compared with the average of the measured values ($\dot{Q}_{gain}$) (b). Although the unmeasured disturbance with an inaccurate model has a weekly pattern, it deviates from the measured values. In addition, each week's estimation has more variations compared to the accurate model case (Fig. 6 (a)), meaning it has some more dynamics in addition to the weekly pattern. As shown in Fig. 7 (a), the estimated unmeasured disturbance could be affected by control input and measured disturbance because the system model is not accurate.

From this investigation, a design matrix of input features and model types is established as shown in 8. Based on the design matrix, we develop several neural network models and evaluate the performance. Input features are composed of three components: time, pattern, and measured disturbance. Time (e.g., hour of week, etc.) and pattern features (past n days of $\hat{\zeta}_{ID}$) are typically used in the time-series modeling. It explains the time-specific and autoregressive pattern characteristics of time-series data. In addition, we also include measured disturbance (w) as shown in Fig. 7.

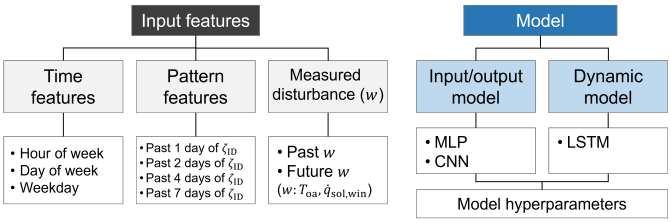


Figure 8: Design matrix of input features and model types for the unmeasured disturbance model.

Three types of models are used in this study: Multi-Layer Perceptron (MLP), Convolution Neural Network (CNN), and Long short-term memory (LSTM). MLP and CNN have input-output structure that maps input time-series vector ($\mathbf{x} \in \mathbb{R}^{(n_x n_{k,x})}$) to output time-series vector ($\mathbf{y} \in \mathbb{R}^{n_{k,y}}$) (i.e., feed-forward network). Input time-series vector is 1-dimensional vector concatenated by input features for a given input time-period ($n_{k,x}$), and output time-series vector is 1-dimensional vector of estimated unmeasured disturbance for the prediction time-horizon ($n_{k,y}$). On the other hand, LSTM is a dynamical model that maps input features ($\mathbf{x}_k \in \mathbb{R}^{n_x}$) to output ($\mathbf{y}_k \in \mathbb{R}^1$) in each time-step ($k$) (i.e., feed-back network). All the three models are commonly used in deep-learning, so we simply summarize the structures below due to the lack of space. MLP is a fully connected feed-forward neural network (Murphy, 2022). The output time-series vector ($\mathbf{y}$) is modeled from input time-series vector ($\mathbf{x}$) through n_{layer} numbers of hidden linear layers with activation function (φ) and dropout layer (Eq. 6).

$$\begin{aligned} &\mathbf{z}_0 = \text{dropout}(\varphi(\mathbf{W}_0\mathbf{x} + \mathbf{b}_0)) \\ &\text{for } i \text{ in } 1 : n_{\text{layer}} : \\ &\quad \mathbf{z}_i = \text{dropout}(\varphi(\mathbf{W}_i\mathbf{z}_{i-1} + \mathbf{b}_i)) \\ &\mathbf{y} = \mathbf{W}_y\mathbf{z}_{n_{\text{layer}}} + \mathbf{b}_y \end{aligned} \tag{6}$$

where $\mathbf{W}_i$ and $\mathbf{b}_i$ are weights and bias parameters that maps hidden layers from $\mathbf{z}_{i-1}$ to $\mathbf{z}_i$ is hidden. φ is activation function, dropout is a dropout layer, which is commonly used to prevent from over-fitting (Murphy, 2022).

CNN is one type of neural network that maps input-output data and is widely used for image classification problems because it extracts meaningful features from raw data through the convolution kernel (Murphy, 2022). The structure is similar to MLP, but the convolution filter and max-pooling layer are used instead of the linear layer, and one more linear layer is for the last convolution layer (Eq. 7).

$$\begin{aligned} &\mathbf{z}_0 = \varphi(\text{maxpool}(\text{conv}(\mathbf{x}))) \\ &\text{for } i \text{ in } 1 : n_{\text{layer}} : \\ &\quad \mathbf{z}_i = \varphi(\text{maxpool}(\text{conv}(\mathbf{z}_{i-1}))) \\ &\mathbf{y} = \mathbf{W}_y(\text{dropout}(\varphi(\mathbf{W}_{n_{\text{layer}}+1}\mathbf{z}_{n_{\text{layer}}} + \mathbf{b}_{n_{\text{layer}}+1}))) + \mathbf{b}_y \end{aligned} \tag{7}$$

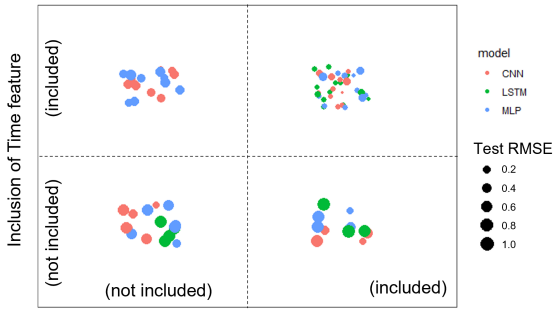
where conv is convolution filter and maxpool is max-pooling layer.

LSTM is one type of recurrent neural network (RNN) using LSTM cell that maps input ($\mathbf{x}_k \in \mathbb{R}^{n_x}$) to output ($\mathbf{y}_k \in \mathbb{R}^1$) in each time ($k$) (Eq. 8) (Murphy, 2022). There are several types of cells for RNN, but LSTM, which includes both short- and long-term memory states, shows a good performance for several time-series tasks.

$$\begin{aligned} &\text{for } i \text{ in } 1 : (n_{k,x} + n_{k,y}) : \\ &\quad \mathbf{h}_{i+1}, \mathbf{c}_{i+1}, = \text{LSTM}(\mathbf{h}_i, \mathbf{c}_i, \mathbf{x}_i) \\ &\text{for } i \text{ in } 1 : (n_{k,y}) : \\ &\quad \mathbf{y}_i = \mathbf{W}_y(\text{dropout}(\varphi(\mathbf{W}_h\mathbf{h}_i + \mathbf{b}_h))) + \mathbf{b}_y \end{aligned} \tag{8}$$

where $\mathbf{h}$ and $\mathbf{c}$ are hidden and cell state vector that holds short- and long-term memory, respectively. $\mathbf{x}_i$ is subset of $\left(\text{time}_i, w_i, \hat{\zeta}_{\text{ID},i}\right)$, and $\hat{y}_{(n_{k,x}+1):(n_{k,x}+n_{k,y})}$ are used to for $\hat{\zeta}_{(n_{k,x}+1):(n_{k,x}+n_{k,y})}$.

For all three models, there are several hyperparameters such as types of activation function, sizes and number of hidden layers, sizes of convolution filter and pooling layers, etc. Model training is conducted on various combinations



(a) Test RMSE vs. model and input features

Test RMSE=β_0+$\beta_1$time+$\beta_2$future w+β_3CNN+β_4LSTM

Variable	β	p-value
Intercept	0.532	2e-16***
time	-0.243	2e-16***
future w	-0.091	1e-4***
CNN	-0.037	0.128
LSTM	0.0067	0.807

(b) Regression of test RMSE vs. model and input features

Figure 9: Impact of time feature and future unmeasured disturbance on test RMSE (test RMSE data is scattered in each box for readability without any meaning).

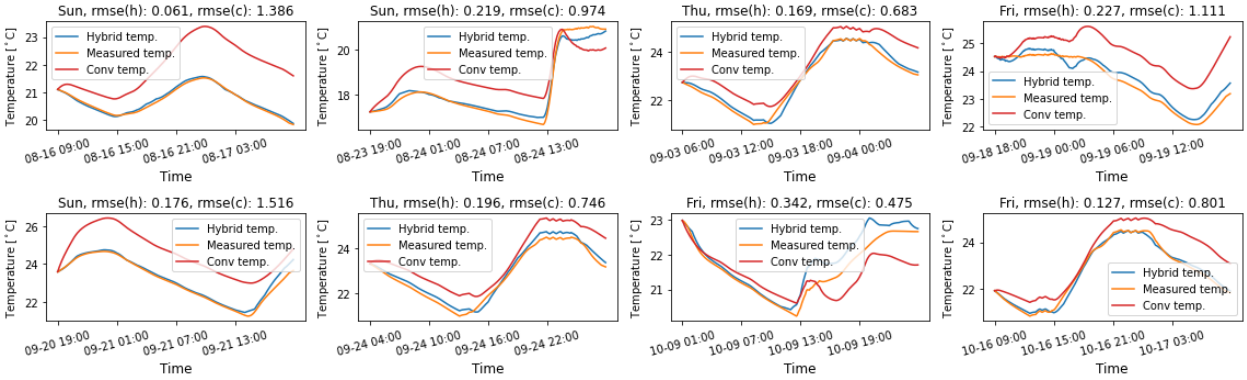


Figure 10: One-day ahead temperature prediction of Hybrid and Conventional approaches on test data (h: Hybrid, c: Conventional).

of hyperparameters, and the best hyper-parameter sets that shows the minimum RMSE on test data is chosen. The details are omitted due to the lack of the space. All the models are developed using PyTorch (Paszke et al., 2019). One month of data (June, 2020) is used for training, and the following four months of data (July-Oct, 2020) is used for the test. All the data is resampled in hourly interval and then scaled by standardization (i.e., z-score normalization).

4. RESULT AND CONCLUSION

One of the main contributions of this study is to propose a Hybrid modeling approach in consideration of the limitation of the gray-box system identification. To validate the design approach proposed in section 3.2, we compare the test RMSE values of various models of input features and model structures. Specifically, the impact of time feature, future measured disturbance (w), and model type on test RMSE is tested using linear regression (Fig. 9 (b)) and visualized in Fig. 9 (a). The inclusion of any time features and future w clearly shows the better performance, and it is statistically significant (p-value<0.05). However, the model types do not show the statistical differences.

Since there are no significant differences between models that include time feature and future w, we choose one specific model (LSTM with weekday time feature, 2 days historic ζ_{ID}, and both past and future w) that shows consistent performances over different data and practical issues (e.g., the required amount of historic data, number of parameters, etc.). The one-day ahead predictions of the Hybrid approach on test data are randomly selected and visualized in Fig. 10. The Hybrid (h) approach shows better prediction than Conventional (c) approach, and the average RMSE of Hybrid approach is 0.24°C while Conventional approach shows 1.1°C.

In this study, we propose the Hybrid approach that overcomes the gray-box model in a predictive application for

real buildings. Throughout the investigation of the various system identification approaches and their limitations, we design a neural network model that predicts unmeasured disturbance. The Hybrid approach shows better performance in temperature prediction on the test data. In the future study, we will investigate the performance of the proposed model in various real data with hyperparameter sensitivity analysis. Also, it is necessary to be discussed how to make this approach in scale over spaces (i.e., a limited number of data) and time (i.e., different seasons). One practical solution is to provide a guideline to deploy the model, but it is worth testing the applicability of Bayesian model update or Transfer learning approach.

REFERENCES

ASHRAE. (2017). 2017 ASHRAE handbook: Fundamentals, chapter 18 SI: Nonresidential cooling and heating load calculations.

Braun, J., & Chaturvedi, N. (2002). An inverse gray-box model for transient building load prediction. *HVAC&R Research*, *8*(1), 73–99. https://doi.org/10.1080/10789669.2002.10391290

Coffman, A. R., & Barooah, P. (2018). Simultaneous identification of dynamic model and occupant-induced disturbance for commercial buildings. *Building and Environment*, *128*, 153–160. https://doi.org/10.1016/j.buildenv.2017.10.020

Dong, B., Li, Z., Rahman, S. M., & Vega, R. (2016). A hybrid model approach for forecasting future residential electricity consumption. *Energy and Buildings*, *117*, 341–351. https://doi.org/10.1016/j.enbuild.2015.09.033

Ellis, M. J. (2021). Machine learning enhanced grey-box modeling for building thermal modeling. *2021 American Control Conference (ACC)*, 3927–3932. https://doi.org/10.23919/ACC50511.2021.9482715

Ham, S. W., Karava, P., Bilionis, I., & Braun, J. (2021). Real-time model for unit-level heating and cooling energy prediction in multi-family residential housing. *Journal of Building Performance Simulation*, *14*(4), 420–445. https://doi.org/10.1080/19401493.2021.1968495

Hong, T., Macumber, D., Li, H., Fleming, K., & Wang, Z. (2020). Generation and representation of synthetic smart meter data. *Building Simulation*, *13*(6), 1205–1220. https://doi.org/10.1007/s12273-020-0661-y

Kim, D., Cai, J., Ariyur, K. B., & Braun, J. E. (2016). System identification for building thermal systems under the presence of unmeasured disturbances in closed loop operation: Lumped disturbance modeling approach. *Building and Environment*, *107*, 169–180. https://doi.org/10.1016/j.buildenv.2016.07.007

Kim, D., Cai, J., Braun, J. E., & Ariyur, K. B. (2018). System identification for building thermal systems under the presence of unmeasured disturbances in closed loop operation: Theoretical analysis and application. *Energy and Buildings*, *167*, 359–369. https://doi.org/10.1016/j.enbuild.2017.12.007

Lawrence Berkeley National Laboratory. (2021). *FLEXLAB the world's most advanced integrated building and grid technologies testbed*. Retrieved December 24, 2021, from https://flexlab.lbl.gov/

Lee, Z. E., & Zhang, K. M. (2021). Scalable identification and control of residential heat pumps: A minimal hardware approach. *Applied Energy*, *286*, 116544. https://doi.org/10.1016/j.apenergy.2021.116544

Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press. probml.ai

O'Neill, Z., Narayanan, S., & Brahme, R. (2010). Model-based thermal load estimation in buildings. *Proceedings of SimBuild*, *4*(1), 474–481.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d. Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

Rouchier, S., Jiménez, M. J., & Castaño, S. (2019). Sequential monte carlo for on-line parameter estimation of a lumped building energy model. *Energy and Buildings*, *187*, 86–94. https://doi.org/10.1016/j.enbuild.2019.01.045

ACKNOWLEDGMENT

The work was supported by the California Energy Commission-sponsored project entitled "HP-Flex: Next Generation Heat Pump Load Flexibility" and DOE BTO sponsored CRADA project entitled "A Low-cost, Scalable Control Solution for Grid-Interactive Small and Medium Sized Commercial Buildings".