Identification of aggregate building thermal dynamic model and unmeasured internal heat load from data

Zhong Guo*, Austin R. Coffman*,‡, Jeffrey Munk†, Piljae Im†, Prabir Barooah*
*University of Florida {zhong.guo, bubbaroney, pbarooah}@ufl.edu
†Oak Ridge National Labratory {munkjd, imp1}@ornl.gov

***Abstract*—This paper is on the problem of simultaneously identifying the parameters of an aggregate thermal dynamic model of a multi-zone building and unknown disturbances from input-output data. An aggregate model is a single-zone equivalent of a multi-zone building, and is useful for many purposes, including model based control of large heating, ventilation and air conditioning (HVAC) equipment that delivers thermal energy to the entire building. A key challenge in identification is the presence of unknown disturbance since it is not measurable but non-negligible.**

We first present a principled method to aggregate a multi-zone building model into a single zone model. We then provide a method to identify thermal parameters and the unknown disturbance for this aggregate (single-zone) model. Finally, we test our proposed identification algorithm to data generated from a virtual building. A key insight provided by the aggregation method allows us to recognize under what conditions the estimation of the disturbance signal will be necessarily poor and uncertain.

I. Introduction

A dynamic model of a building's zone temperature is useful in several applications involving heating, ventilation, and air conditioning (HVAC) systems, such as model-based control for improving indoor climate and reducing energy use [1], [2], limiting peak demand [3], [4], or providing ancillary services to the power grid [5], [6]. To be used in a control algorithm, the model should also be of low order. One way to achieve this is to estimate the model from input-output measurements.

There are many different model structures for modeling the thermal dynamics of a building zone. Among them the Resistance Capacitance (RC) network is a common subclass. There is a rich literature on identification of the thermal parameters of a RC network model for single zone commercial buildings [7], [8], [9], [10]. For single zone buildings, the inputs and output of the model are unambiguously defined. The output is the zone temperature, while the inputs are the heating (or cooling) rate injected by the HVAC system, ambient air temperature, solar irradiance and the internal heat load. The internal heat load is a sum of all sources of heat inside the building, such as from occupants' metabolism, and lights and appliances used by occupants. The internal heat load is an exogenous disturbance that is typically unknown and not measurable. Since this disturbance can be comparable in magnitude to the cooling provided by the HVAC system, it poses a challenge in model identification from measured input-output data [11]. Work on identifying the model in spite of the presence of this unknown disturbance is quite recent [11], [12], [13]. Earlier works have mostly ignored this disturbance; see [12] for a discussion of earlier approaches.

There are many challenges in extending modeling/identification approaches from a single to a multi-zone building. There are many more inputs and outputs, as well as more parameters and states, in case of a multi-zone model compared to a single-zone model. In particular, there can be as many internal heat loads (unknown disturbances) as there are zones. The identification of a multi-zone model is therefore more challenging than that of a single zone model [14]. Many works on model based control for multi-zone buildings instead use a "single-zone equivalent" model of the building, in which the average building temperature (averaged over the zones) and the sum of zone-level inputs are used [15], [16], [17]. Such a model is still useful in computing control commands for large equipment such as a chiller or an air handling unit that serves a multi-zone building.

In this paper we address the problem of identification of such an aggregate model. The first contribution of the paper is a principled method of aggregating a high-dimensional multi-zone model into a low-order single-zone equivalent model (that has the same structure as a single-zone model). We call it an *aggregate model* of the building. This step is necessary because of the ambiguity in the definition of inputs and outputs of the aggregate model. An outcome of this method is an insight into the appropriate definition of

‡ corresponding author.
ZG, AC, and PB are with the Department of Mechanical and Aerospace Engineering at the University of Florida. PI and JM are with the Building Envelope & Urban Systems Research group at Oak Ridge National Laboratory, Oak Ridge, Tennessee. The research reported here has been partially supported by the NSF through awards 1463316 (CMMI) and 1646229 (CPS), and DOE GMLC grant titled "virtual batteries". The authors would like to thank Naren Raman and Bo Chen for help with the Casadi software.

the internal heat load for the aggregate model. When the measurable outputs and inputs of the aggregate model are defined to be average of the zone-level signals, the aggregate internal heat load turns out to be a non-trivial function of many signals, not just the average of the internal heat loads of the zones. Our second contribution is a novel method to estimate both the model and the aggregate internal heat load from measurable input-output data. The third contribution is the evaluation of the proposed method with simulation generated data. Evaluation from the simulation dataset shows that the proposed method performs quite well. The insight on the difference between the average internal heat loads and the aggregate heat load is useful in evaluating the estimation results.

A. *Contribution over prior art*

The concept of an aggregate model is not new in building modeling and control literature. Many Model Predictive Control (MPC) formulations reported in the literature require an aggregate model to determine building level control commands [17], [18]. There is in fact a plethora of works that reduces a model of a large commercial multi-zone building to that of a single "aggregate" zone. However, there are very few principled methods in the current literature on how to define the inputs and outputs of such an aggregate model, especially for system identification. It is a common practice to average the inputs and outputs over all the zones to form a single set of input/output signals. These signals are then assumed to be related by an arbitrary RC network model, which is then set up as a system identification problem to determine the parameters of the chosen RC network structure. The difference between our present work and what was just described, is that we *first* construct our aggregate model through analysis of a multi-zone model.

Since there is a rich literature on building model identification and control for a "single zone" building, we develop our aggregate model so that it mimics the form of a single zone model. We derive the aggregate model from the RC network models of the individual zones and constructed definitions of aggregate input/output signals. The derived aggregate model is shown itself to be an RC network model that is affected by additive time varying terms. These additive time varying terms arise from the potential asynchronous inputs affecting each different zone, which we term "aggregation errors." The thermal parameters of this aggregate model are shown to be time invariant and are also direct functions of the individual zone parameters. Additionally, The aggregate input and output signals are solely functions of the individual zone input and output signals, unlike other aggregation techniques [17] that require knowledge of the individual zone's thermal parameters to form aggregate inputs/outputs. Since the aggregate input/outputs only require knowledge of the individual inputs/outputs, and parameters of the aggregate model are time invariant, the aggregate model lends itself well to system identification approaches.

The identification method proposed here is an extension of the method proposed in our prior work [12], which identifies the parameters of a single-zone building model and the unknown disturbance signal affecting the building. The difference is that the method proposed here allows incorporation of constraints (such as non-negativity of the internal heat load) while that in [12] does not.

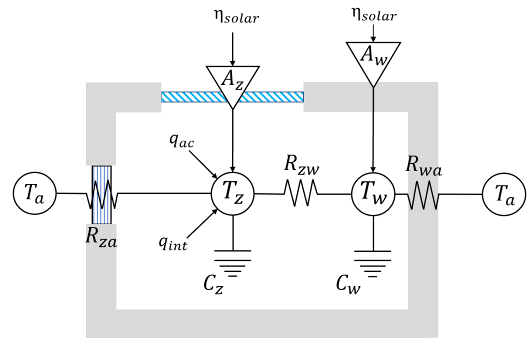

Fig. 1: 2R2C network on top of a single zone building.

II. Building Thermal Modeling Structure

A. *Single-zone Model*

A floor plan for a single zone commercial building is presented in Figure 1. A Resistance Capacitance (RC) network model, in particular, a 2R2C model, is overlayed on the floor plan. The 2R2C model refers to the following coupled differential equations that describe the evolution of the temperature of the building (zone), T_z, in response to various inputs:

$$\begin{aligned}\dot{T}_z(t) &= \frac{T_a(t)-T_z(t)}{R_{za}C_z} + \frac{1}{C_z}(q_{int}(t)-q_{ac}(t)) \\ &\quad + \frac{T_w(t)-T_z(t)}{R_{zw}C_z} + \frac{A_z}{C_z}\eta_{solar}(t), \\ \dot{T}_w(t) &= \frac{T_z(t)-T_w(t)}{R_{zw}C_w} + \frac{T_a(t)-T_w(t)}{R_{wa}C_w} + \frac{A_w}{C_w}\eta_{solar}(t),\end{aligned} \tag{1}$$

where T_w is a fictitious state (temperature of the wall, if you must), q_{ac} is the heat load due to the HVAC system, and η_{solar}, T_a, and q_{int} are non-controllable inputs, which are the solar irradiance, ambient (outside) temperature, and internal heat load, respectively. The parameters in this model are $\{R_{za}, C_z, R_{zw}, C_w, A_z, A_w\}$.

B. *Aggregate Multi-zone Model*

Here we describe how a model of a multi-zone building can be aggregated into a single-zone model of the type described in the previous section. In the sequel, we call the resulting single-zone model as the *aggregate model* of the multi-zone building. It is shown that the aggregate model is an RC network model either with time-varying thermal parameters, or with time invariant thermal parameters and additive time-varying terms.

To start the derivation, let N_z be the number of zones in the building, and each is modeled by a 2R2C network model using (1). *We now define the average input/output quantities as follows:*

$$\bar{T}_x(t) \triangleq \frac{1}{N_z}\sum_{j=1}^{N_z} T_x^j(t), \quad \bar{q}_{ac}(t) \triangleq \frac{1}{N_z}\sum_{j=1}^{N_z} q_{ac}^j(t). \tag{2}$$

$$\bar{q}_{int}(t) \triangleq \frac{1}{N_z}\sum_{j=1}^{N_z} q_{int}^j(t), \quad \bar{\eta}_{solar}(t) \triangleq \frac{1}{N_z}\sum_{j=1}^{N_z} \eta_{solar}^j(t).$$

where T_x^j is the temperature of x^{th} location for the j^{th} zone, e.g., T_z^2 represents the zone temperature of 2nd zone. In the sequel, *the bar is used to distinguish an average quantity*. Note that the average signals so defined can be computed from measurements of zone-level signals – though some, especially q_{int}^j – may be hard to measure.

By summing the N_z individual thermal models (1) over the j index and dividing by the number of zones, we have

$$\begin{aligned}\frac{1}{N_z}\sum_{j=1}^{N_z}\dot{T}_z^j(t) &= \frac{1}{N_z}\sum_{j=1}^{N_z}\left(\frac{T_a^j(t)-T_z^j(t)}{R_{za}^j C_z^j} + \frac{T_w^j(t)-T_z^j(t)}{R_{zw}^j C_z^j}\right) \\ &+ \frac{1}{N_z}\sum_{j=1}^{N_z}\left(\frac{q_{int}^j(t)-q_{ac}^j(t)}{C_z^j} + \frac{A_z^j}{C_z^j}\eta_{solar}^j(t)\right)\end{aligned} \tag{3}$$

$$\begin{aligned}\frac{1}{N_z}\sum_{j=1}^{N_z}\dot{T}_w^j(t) &= \frac{1}{N_z}\sum_{j=1}^{N_z}\left(\frac{T_z^j(t)-T_w^j(t)}{R_{zw}^j C_w^j} + \frac{A_w^j}{C_w^j}\eta_{solar}^j(t)\right) \\ &+ \frac{1}{N_z}\sum_{j=1}^{N_z}\frac{T_a^j(t)-T_w^j(t)}{R_{wa}^j C_w^j}.\end{aligned} \tag{4}$$

Notice that above we have assumed that the zones have no thermal interactions with each other. We have done so only in the interest of space; later we comment on the changes that occur if thermal interactions among zones are present.

Now we can construct the aggregate model by substituting definitions (2) into equations (3)-(4). Two approaches are available, leading to time varying model and time invariant model, respectively.

1) Time Varying Model: If we define a model from (3)-(4) that mimics the structure of the single-zone model (1) in terms of the average signals defined earlier, the thermal parameters will be time varying,

$$\begin{aligned}\dot{\bar{T}}_z(t) &= \frac{\bar{T}_a(t)-\bar{T}_z(t)}{\bar{\tau}_{za}(t)} - \frac{\bar{q}_{ac}(t)}{\bar{C}_z^{ac}(t)} + \frac{\bar{q}_{int}(t)}{\bar{C}_z^{int}(t)} \\ &+ \frac{\bar{T}_w(t)-\bar{T}_z(t)}{\bar{\tau}_{zw}(t)} + \bar{A}_z(t)\bar{\eta}_{solar}(t),\end{aligned} \tag{5}$$

$$\dot{\bar{T}}_w(t) = \frac{\bar{T}_a(t)-\bar{T}_w(t)}{\bar{\tau}_{wa}(t)} + \frac{T_z(t)-\bar{T}_w(t)}{\bar{\tau}_{wz}(t)} + \bar{A}_w(t)\bar{\eta}_{solar}(t), \tag{6}$$

where the expression for the thermal parameters are,

$$\bar{\tau}_{xy}(t) \triangleq \frac{N_z(\bar{T}_y(t)-\bar{T}_x(t))}{\sum_{j=1}^{N_z}\frac{T_y^j(t)-T_x^j(t)}{R_{xy}^j C_x^j}}, \tag{7}$$

$$\bar{C}_x^y(t) \triangleq \frac{N_z\bar{q}_y(t)}{\sum_{j=1}^{N_z}\frac{q_y^j(t)}{C_x^j}}, \tag{8}$$

$$\bar{A}_x(t) \triangleq \frac{\sum_{j=1}^{N_z}\frac{A_x^j}{C_x^j}\eta_{solar}^j(t)}{N_z\bar{\eta}_{solar}(t)}. \tag{9}$$

To save space, the sub/superscripts on the LHS values denote the relevant location (like the x subscript in (2)). The right hand side of these equations become time-invariant only if either thermal parameters or input variables are completely homogeneous over the zone index j. Otherwise, these defined "thermal parameters" will be time-varying depending on each zone's individual inputs and states, which is not an appealing situation.

2) Time Invariant Aggregate Model: Since the thermal parameters for the individual zones are out of our control and it is unlikely that the inputs will be completely homogeneous over each zone, we modify our interpretation of the average signals (2) so that the aggregate model becomes a RC network model with time invariant thermal parameters. Doing so requires defining the following deviation variables:

$$\begin{aligned}&\tilde{T}_x^j \triangleq T_x^j - \bar{T}_x, \quad \tilde{q}_{ac}^j \triangleq q_{ac}^j - \bar{q}_{ac}, \\ &\tilde{q}_{int}^j \triangleq q_{int}^j - \bar{q}_{int}, \quad \tilde{\eta}_{solar}^j \triangleq \eta_{solar}^j - \bar{\eta}_{solar}\end{aligned} \tag{10}$$

The interpretation of (10) is that each individual input and state can be represented as the average quantity (2) (zonal average) plus some deviation. With some tedious algebra, it can be shown that with the help of these definitions, the ODEs (5)-(6) can be transformed to:

$$\begin{aligned}\dot{\bar{T}}_z(t) =& \frac{\bar{T}_a(t)-\bar{T}_z(t)}{\bar{\tau}_{za}} + \frac{\bar{T}_w(t)-\bar{T}_z(t)}{\bar{\tau}_{zw}} \\ &+ \frac{1}{\bar{C}_z}\left(\bar{q}_{int}(t)-\bar{q}_{ac}(t)\right) + \bar{A}_z\bar{\eta}_{solar}(t) + \tilde{w}_z(t)\end{aligned} \tag{11}$$

$$\begin{aligned}\dot{\bar{T}}_w(t) =& \frac{\bar{T}_a(t)-\bar{T}_w(t)}{\bar{\tau}_{wa}} + \frac{\bar{T}_z(t)-\bar{T}_w(t)}{\bar{\tau}_{wz}} \\ &+ \bar{A}_w\bar{\eta}_{solar}(t) + \tilde{w}_w(t)\end{aligned} \tag{12}$$

where

$$\begin{aligned}\tilde{w}_z(t) \triangleq& \frac{1}{N_z}\sum_{j=1}^{N_z}\frac{\tilde{T}_a^j(t)-\tilde{T}_z^j(t)}{R_{za}^j C_z^j} + \frac{1}{N_z}\sum_{j=1}^{N_z}\frac{\tilde{T}_w^j(t)}{R_{zw}^j C_z^j} \\ &+ \frac{1}{N_z}\sum_{j=1}^{N_z}\frac{\tilde{q}_{int}^j(t)-\tilde{q}_{ac}^j(t)+A_z^j\tilde{\eta}_{solar}^j(t)}{C_z^j}\end{aligned} \tag{13}$$

$$\begin{aligned}\tilde{w}_w(t) \triangleq& \frac{1}{N_z}\sum_{j=1}^{N_z}\frac{\tilde{T}_a^j(t)-\tilde{T}_w^j(t)}{R_{wa}^j C_w^j} + \frac{1}{N_z}\sum_{j=1}^{N_z}\frac{\tilde{T}_z^j(t)}{R_{zw}^j C_w^j} \\ &+ \frac{1}{N_z}\sum_{j=1}^{N_z}\frac{A_w^j}{C_w^j}\tilde{\eta}_{solar}^j(t)\end{aligned} \tag{14}$$

The signals $\tilde{w}_z, \tilde{w}_w$ are additive time-varying terms that represent model mismatch due to the non-synchronous inputs and states for each zone and wall, respectively. We term these additive terms "aggregation errors" and they would be zero only if all deviation variables are synchronous. *We define the aggregate internal heat load as:*

$$\bar{q}_{agg}(t) \triangleq \bar{q}_{int}(t) + \tilde{w}_z(t)\bar{C}_z, \tag{15}$$

which leads to the final construction of the aggregate model,

$$\begin{aligned}\dot{\bar{T}}_z(t) =& \frac{\bar{T}_a(t) - \bar{T}_z(t)}{\bar{\tau}_{za}} + \frac{\bar{T}_w(t) - \bar{T}_z(t)}{\bar{\tau}_{zw}} + \bar{A}_z\bar{\eta}_{solar}(t) \\ &+ \frac{1}{\bar{C}_z}(\bar{q}_{agg}(t) - \bar{q}_{ac}(t)) \end{aligned}\tag{16}$$

$$\begin{aligned}\dot{\bar{T}}_w(t) =& \frac{\bar{T}_a(t) - \bar{T}_w(t)}{\bar{\tau}_{wa}} + \frac{\bar{T}_z(t) - \bar{T}_w(t)}{\bar{\tau}_{wz}} \\ &+ \bar{A}_w\bar{\eta}_{solar}(t) + \tilde{w}_w(t). \end{aligned}\tag{17}$$

The definition of $\bar{\tau}_{xy}$, $\bar{C}_x$, and $\bar{A}_x$ are now as follows:

$$\bar{\tau}_{xy} \triangleq \frac{N_z}{\sum_{j=1}^{Nz} \frac{1}{R^j_{xy}C^j_x}}, \quad \bar{C}_x \triangleq \frac{N_z}{\sum_{j=1}^{Nz} \frac{1}{C^j_x}}, \quad \bar{A}_x \triangleq \frac{\sum_{j=1}^{Nz} \frac{A^j_x}{C^j_x}}{N_z}. \tag{18}$$

Comment 1: The derived model (16)-(17) represents the time evolution for "average" quantities of building zones. That is, if data is collected from individual zones in a building and is aggregated (according to (2)) then equation (16)-(17) describes how these states should change over time. Furthermore, if this average data is utilized in system identification of an equivalent single-zone RC-network model, then (18) informs us how the thermal parameters of such a model are related to thermal parameters of each individual model. More importantly, (15) shows that the aggregate disturbance $\bar{q}_{agg}$ is not the average internal heat load, $\bar{q}_{int}$, but that there is a additive term ($\tilde{w}_z$) that is related to all the other states and inputs in a complex manner.

Comment 2: Although we have neglected thermal interactions in the derivation of the aggregate model, those interactions only change the interpretation of $\tilde{w}_z$. The RC network structure remains and the thermal parameters of the aggregate model (18) remain the same.

Thermal interactions between the zones is modeled as an additional term to T^j_z in (1), $\sum_{i\neq j} \frac{T^i_z - T^j_z}{R^{ij}_z C^j_z}$. If each $C^j_z = C_0$ for all j, the thermal interactions will cancel completely in the aggregate model, since $R^{ij}_z = R^{ji}_z$ by definition. If each C^j_z is distinct then the aggregation error $\tilde{w}_z$ changes to $\tilde{w}^x_z$, where

$$\tilde{w}^x_z := \tilde{w}_z + \frac{1}{N_z}\sum_{i=1}^{N_z}\sum_{j=1}^{N_z} \frac{\tilde{T}^i_z - \tilde{T}^j_z}{R^{ij}_z C^j_z}. \tag{19}$$

III. System Identification

A. Problem Statement

We now turn to the problem of identifying the aggregate model (16)-(17) from input-output data collected from multiple zones. That is, given time traces (of length N_t) of the average output and inputs, $\{\bar{T}_{z,k}\}_{k=1}^{N_t}$, $\{\bar{q}_{ac,k}\}_{k=1}^{N_t}$, $\{\bar{T}_{a,k}\}_{k=1}^{N_t}$, and $\{\bar{\eta}_{solar,k}\}_{k=1}^{N_t}$ we wish to identify the aggregate thermal parameters $\{\bar{\tau}_{za}, \bar{\tau}_{zw}, \bar{\tau}_{wa}, \bar{\tau}_{wz}, \bar{C}_z, \bar{A}_z\, \bar{A}_w\}$. Recall that (i) these time traces can be computed from measurements collected from each zone of the building, (ii) the presence of the unknown disturbances in the model, $\bar{q}_{agg}(t), \tilde{w}_w(t)$ presents a serious hurdle, especially since $\bar{q}_{agg}$ is large; sometimes comparable to the cooling provided by the HVAC system [11].

Our proposed approach is inspired by that in [12], in which a simple dynamic model of aggregate internal heat load is proposed and then estimation of this unknown signal is cast as a state estimation problem. Recall that $\bar{q}_{agg} = \bar{q}_{int} + \tilde{w}_z\bar{C}_z$. In this work we assume that $\bar{q}_{int} >> \tilde{w}_z\bar{C}_z$, that is the average internal heat load is much larger than the scaled aggregation errors. Due to this, we adapt the following dynamic model for $\bar{q}_{agg}$ [12]:

$$\frac{d}{dt}\bar{q}_{agg}(t) = 0 \tag{20}$$

The justification for this model comes from the usual pattern of occupancy in commercial buildings, in which occupants typically enter and exit the building in bulk. This bulk transfer of people would correspond to a piece-wise constant occupant-induced heat load signal, which would consequently have a time-derivative whose value is 0 for most of the time. Note that the dynamic model for the aggregate internal heat load can be coupled with the aggregate RC network model (16)-(17) to form the aggregate model:

$$\begin{aligned}\begin{bmatrix}\dot{\bar{T}}_z \\ \dot{\bar{T}}_w \\ \dot{\bar{q}}_{agg}\end{bmatrix} =& \begin{bmatrix}\frac{-1}{\bar{\tau}_{za}} + \frac{-1}{\bar{\tau}_{zw}} & \frac{1}{\bar{\tau}_{zw}} & 1 \\ \frac{1}{\bar{\tau}_{wz}} & \frac{-1}{\bar{\tau}_{wa}} + \frac{-1}{\bar{\tau}_{wz}} & 0 \\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\bar{T}_z \\ \bar{T}_w \\ \bar{q}_{agg}\end{bmatrix} \\ &+ \begin{bmatrix}\frac{1}{\bar{\tau}_{za}} & \bar{A}_z & \frac{1}{\bar{C}_z} \\ \frac{1}{\bar{\tau}_{wa}} & \bar{A}_w & 0 \\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\bar{T}_a \\ \bar{\eta}_{solar} \\ \bar{q}_{ac}\end{bmatrix},\end{aligned} \tag{21}$$

where we have dropped $\tilde{w}_w$ for it is not accessible in reality. This can then be expressed in compact continuous time state space notation as,

$$\dot{x}(t) = A(\theta)x(t) + B(\theta)u(t) \tag{22}$$

$$z(t) = Cx(t) \tag{23}$$

where the definitions of $x(t), u(t)$, and θ follow from comparison to (21).

The aggregate level identification problem can now be posed as follows (with sampled data): given N_t time-samples of the average measured inputs $u[k] = [\bar{T}_a[k], \bar{q}_{ac}[k], \bar{\eta}_{sol}[k]]^T \in \mathbb{R}^3$ for $k = 1, \ldots, N_t$ and the measured average output $z[k] = \bar{T}_z[k]$, identify the unknown thermal parameters, $\theta = [\bar{\tau}_{za}, \bar{\tau}_{zw}, \bar{\tau}_{wa}, \bar{\tau}_{wz}, \bar{C}_z, \bar{A}_z\, \bar{A}_w] \in \mathbb{R}^7$, and aggregate internal heat load signal samples $\bar{q}_{agg}[k] \in \mathbb{R}$ for $k = 0, \ldots, N_t - 1$.

B. Proposed Method

Our proposed method involves solving a constrained nonlinear optimization problem in order to obtain estimates for the aggregate thermal parameters and aggregate indoor heat load. This optimization problem is inspired from the "batch estimation" approach common in state estimation [19]. Since $\bar{q}_{\text{agg}}$ has been recast as a state variable, estimation of the state produces an estimate for this quantity.

If the thermal parameters were already known, a Kalman filter would be adequate to estimate the state since the model is linear. However, since the thermal parameters are unknown the problem turns into a nonlinear state estimation problem. There are many available choices for non-linear state estimation, such as particle filtering, the extended Kalman filter, or moving horizon/batch estimation approaches [19], [20]. In this work we elect the latter approach because it allows for an easy incorporation of inequality constraints on the state estimates. Particularly, the aggregate internal heat load is enforced to be positive since occupant-induced load usually provides heating.

To setup the batch optimization problem, the continuous time aggregate model (21) is discretized using a first order backward Euler method with a sampling time t_s. Additionally, process and measurement noise are included to account for modeling error. The corresponding discrete time model is then,

$$x_{k+1} = x_k + t_s\left(A(\theta)x_k + B(\theta)u_k\right) + G_d\xi_k, \tag{24}$$

$$z_k = Cx_k + \nu_k, \tag{25}$$

where ξ_k and ν_k are white noise sequences that capture the modeling error and sensor noise, respectively. Here we only model $\bar{q}_{\text{agg}}$ as having process noise (i.e., $G_d = [0, 0, 1]^T$). The reason is that since we are also identifying system parameters, adding noise to the other states would greatly increase the number of degrees of freedom and likely produce spurious results.

The batch estimation problem is now posed as,

$$\begin{aligned}
\min_{\theta, x_0^e, \{\xi_k\}_{k=0}^{N_t-1}} \Bigg(& (x_0^e)^T P_{x_0}^{-1}(x_0^e) \;+\; (\theta-\bar{\theta})^T P_\Theta^{-1}(\theta-\bar{\theta}) \\
& + \;\lambda \sum_{k=0}^{N_t-1} |\xi_k| \;+\; R^{-1}\sum_{k=1}^{N_t}\nu_k^2 + \;\alpha\sum_{k=1}^{N_t}\bar{q}_{agg,k}^2 \Bigg) \\
\text{s.t.} \quad & \forall\; k \in \{0, ..., N_t - 1\} \\
& x_{k+1} = x_k + t_s\Big(A(\theta)x_k + B(\theta)u_k\Big) + G_d\xi_k \\
& x_k \in X_k \quad \theta \in \Theta
\end{aligned}$$

where X_k represents the state inequality constraint at time step k and Θ represents the thermal parameter constraint set. The scalar values λ, R^{-1} represent the cost for the process and measurement noise, respectively. The matrix $P_{x_0}^{-1}$ is a weighting matrix for x_0^e which is the difference from the initial guess of the initial state x_0 of (21). The matrix and vector $P_\Theta^{-1}, \bar{\theta}$ are design choices and represent a quadratic cost and center of the parameter vector, respectively. Incorporating this term in the objective function allows for us to add prior knowledge of the parameter values to the estimation procedure.

We have incorporated an absolute value penalty on the process noise for the aggregate internal heat load estimate.

TABLE I: Parameter estimates from open loop simulation data of the virtual building.

Parameter	Estimate	True Value	units
$\bar{\tau}_{za}$	0.7652	0.7899	hour(s)
$\bar{\tau}_{zw}$	0.5919	0.5869	hour(s)
$\bar{C}_z$	0.7050	0.7147	kWh/K
$\bar{A}_z$	0.7884	0.5700	Km^2/kWh
$\bar{\tau}_{wa}$	24.5098	19.3798	hour(s)
$\bar{\tau}_{wz}$	2.6795	2.8441	hour(s)
$\bar{A}_w$	4.2955	4.5537	Km^2/kWh

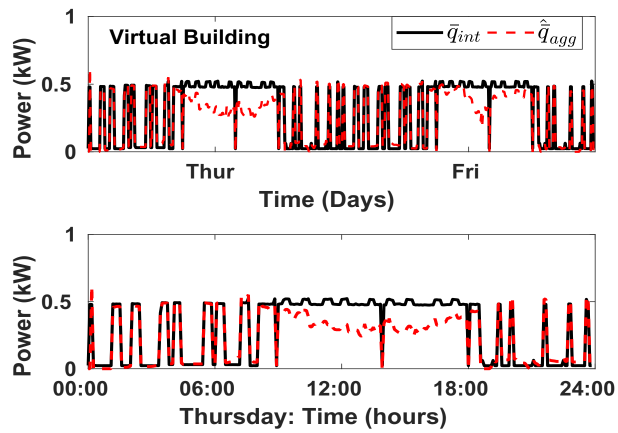

Fig. 2: Aggregate Disturbance estimate $\hat{\bar{q}}_{agg}$ and internal disturbance $\bar{q}_{int}$ for the simulated virtual building.

This is because we expect the derivative of $\bar{q}_{int}$ (which is a main component of $\bar{q}_{agg}$) to be sparse. Additionally, we offer a penalty term on the aggregate internal heat load in the objective function for system identification. Since the aggregate internal heat load acts as a "slack variable" for the aggregate zone temperature, we do not want it to make up for all of the model error. Adding a penalty in the cost function enforces this.

Mathematically, this problem represents the minimization of a convex function over a non-convex constraint set. This makes the overall problem a non-convex optimization problem, which we solve with casadi [21] and the NLP solver IPOPT [22].

IV. Evaluation with simulation data

We present the result for a 2 day open loop simulation of a building model ("virtual building") with 5 independent zones, each modeled by a 2R2C network model (1). The simulated data are aggregated according to the aggregation method described in Section II-B. This data is used by the proposed identification method to estimate the parameters of an aggregate 2R2C network model (16)-(17) along with the aggregate internal heat load. The parameter estimation results are shown in Table I. As one can see from the table, the parameters are estimated quite accurately. The estimated aggregate internal heat load $\hat{\bar{q}}_{agg}$ and the true average internal heat load $\bar{q}_{int}$ are shown in Figure 2.

Figure 3 shows the estimated aggregate internal heat load $\hat{\bar{q}}_{\text{agg}}$, the true aggregate internal heat load $\bar{q}_{\text{agg}}$, and the

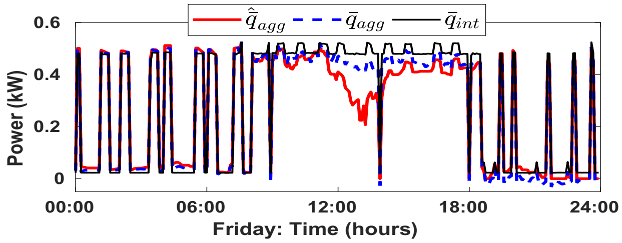

Fig. 3: Comparison of $\bar{q}_{\mathrm{agg}}$, its estimate $\hat{\bar{q}}_{agg}$, and $\bar{q}_{int}$.

average internal heat load $\bar{q}_{int}$. The estimate is quite accurate at night, but there are larger errors during the daytime. The reason for this feature is not completely clear yet, but part of the error can be easily explained from the analysis performed earlier. In particular, recall from (15) that the aggregate internal heat loads differs from the average heat load by $\tilde{w}_z(t)C_z$. In fact this difference is clearly seen in Figure 3. Since the estimation method is designed to estimate the disturbance affecting the aggregate model ($\bar{q}_{\mathrm{agg}}$), which is not the average internal heat load ($\bar{q}_{int}$), even if the estimation method performs perfectly there will be some non-vanishing difference between the estimate and the true average internal heat load. This observation should be kept in mind in interpreting the estimation results. Otherwise the estimation method may appear to perform more poorly than it actually is.

V. Conclusion

We derive an aggregate (single-zone equivalent) model of a multi-zone building. Our method shows that unless each of the individual zones are identical (i.e synchronous inputs and identical thermal parameters) the aggregate model will be effected by time varying additive terms. These additive terms act as an additional heating/cooling source to the model. In fact, in [23] the authors apply similar definitions of the aggregate input/output signals to collected data from a multi-zone building. When the cooling load is predicted, they observe that model errors were most sensitive to the magnitude of the internal heat loads. The authors in [23] provide one explanation for this. However, another possible explanation is that the additive terms due to aggregation were not accounted for.

Additionally, we propose an identification technique to identify the parameters of the derived aggregate model and the unmeasured internal heat load. The identification method is then tested on data from an open loop simulation with known internal heat load. The results corroborate the effectiveness of our methods.

There are two main avenues for future work: examination of the effect of latent heat, and identification of a multi-zone building model.

References

[1] A. Afram and F. Janabi-Sharifi, "Theory and applications of HVAC control systems–a review of model predictive control (MPC)," *Building and Environment*, vol. 72, pp. 343–355, 2014.

[2] Y. Ma, A. Kelman, A. Daly, and F. Borrelli, "Predictive control for energy efficient buildings with thermal storage: Modeling, stimulation, and experiments," *IEEE Control Systems Magazine*, vol. 32, no. 1, pp. 44–64, Feb 2012.

[3] F. B. Morris, J. E. Braun, and S. J. Treado, "Experimental and simulated performance of optimal control of building thermal storage," *ASHRAE transactions*, vol. 100, no. 1, pp. 402–414, 1994.

[4] Y. Sun, S. Wang, F. Xiao, and D. Gao, "Peak load shifting control using different cold thermal energy storage facilities in commercial buildings: A review," *Energy conversion and management*, vol. 71, pp. 101–114, 2013.

[5] H. Hao, T. Middelkoop, P. Barooah, and S. Meyn, "How demand response from commercial buildings will provide the regulation needs of the grid," in *50th Annual Allerton Conference on Communication, Control and Computing*, October 2012, invited paper.

[6] Y. Lin, P. Barooah, S. Meyn, and T. Middelkoop, "Experimental evaluation of frequency regulation from commercial building hvac systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 776–783, March 2015.

[7] S. F. Fux, A. Ashouri, M. J. Benz, and L. Guzzella, "EKF based self-adaptive thermal model for a passive house," *Energy and Buildings*, vol. 68, Part C, pp. 811 – 817, 2014.

[8] P. Bacher and H. Madsen, "Identifying suitable models for the heat dynamics of buildings," *Energy and Buildings*, vol. 43, no. 7, pp. 1511 – 1522, 2011.

[9] S. Wang and X. Xu, "Parameter estimation of internal thermal mass of building dynamic models using genetic algorithm," *Energy Conversion and Management*, vol. 47, no. 1314, pp. 1927 – 1941, 2006.

[10] K. K. Andersen, H. Madsen, and L. H. Hansen, "Modelling the heat dynamics of a building using stochastic differential equations," *Energy and Buildings*, vol. 31, no. 1, pp. 13 – 24, 2000.

[11] D. Kim, J. Cai, K. B. Ariyur, and J. E. Braun, "System identification for building thermal systems under the presence of unmeasured disturbances in closed loop operation: Lumped disturbance modeling approach," *Building and Environment*, vol. 107, pp. 169 – 180, 2016.

[12] A. Coffman and P. Barooah, "Simultaneous identification of dynamic model and occupant-induced disturbance for commercial buildings," *Building and Environment*, vol. 128, no. 153-160, 2018.

[13] T. Zeng, J. Brooks, and P. Barooah, "Simultaneous identification of building dynamic model and disturbance using sparsity-promoting optimization," in *5th International Conference on High Performance Buildings*, July 2018, pp. 1–10.

[14] T. Zeng and P. Barooah, "Identification of network dynamics and disturbance for a multi-zone building," in *2nd IFAC Conference on Cyber-Physical and Human Systems*, December 2018.

[15] T. Pan, J. Salom, and R. Costa-Castell, "Configurations of model predictive control to exploit energy flexibility in building thermal loads," in *2018 IEEE Conference on Decision and Control (CDC)*, Dec 2018, pp. 3177–3182.

[16] J. Ma, S. J. Qin, and T. Salsbury, "Application of economic mpc to the energy and demand minimization of a commercial building," *Journal of Process Control*, vol. 24, no. 8, pp. 1282 – 1291, 2014, economic nonlinear model predictive control.

[17] N. R. Patel, J. B. Rawlings, M. J. Ellis, M. J. Wenzel, and R. D. Turney, "An economic model predictive control framework for distributed embedded battery applications," 2018.

[18] G. Serale, M. Fiorentini, A. Capozzoli, D. Bernardini, and A. Bemporad, "Model predictive control (mpc) for enhancing building and hvac system energy efficiency: Problem formulation, applications and opportunities," *Energies*, vol. 11, no. 3, 2018.

[19] D. G. Robertson, J. H. Lee, and J. B. Rawlings, "A moving horizon-based approach for least-squares estimation," *AIChE Journal*, vol. 42, no. 8, pp. 2209–2224, 1996.

[20] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[21] J. Andersson, J. Åkesson, and M. Diehl, "Casadi: A symbolic package for automatic differentiation andoptimal control," in *Recent Advances in Algorithmic Differentiation*, S. Forth, P. Hovland, E. Phipps, J. Utke, and A. Walther, Eds. Springer Berlin Heidelberg, 2012, pp. 297–307.

[22] A. Wächter and L. B, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, March 2006.

[23] J. E. Braun and N. Chaturvedi, "An inverse gray-box model for transient building load prediction," *HVAC&R Research*, vol. 8, no. 1, pp. 73–99, 2002.