

Simultaneous Identification of Linear Building Dynamic Model and Disturbance Using Sparsity-Promoting Optimization

Tingting Zeng*, Jonathan Brooks, and Prabir Barooah

University of Florida, Department of Mechanical and Aerospace Engineering,
Gainesville, Florida, USA *

Email: tingtingzeng@ufl.edu, jonathanbrooksuf@gmail.com, pbarooah@ufl.edu

*Corresponding Author

ABSTRACT

We propose a method that simultaneously identifies a dynamic model of a building's temperature in the presence of large, unmeasured disturbances, and a transformed version of the unmeasured disturbance. Our method uses ℓ_1 -regularization to encourage the identified disturbance to be approximately sparse, which is motivated by the piecewise constant nature of occupancy that determines the disturbance. We test our method using both open-loop and closed-loop simulation data. Results show that the identified model can accurately identify the transfer functions in both scenarios, even in the presence of large disturbances, and even when the disturbance does not satisfy the piecewise constant property.

1. INTRODUCTION

A dynamic model of a building's temperature is necessary for model-based control of building HVAC (Heating Ventilation and Air Conditioning) systems (Průvara et al., 2013). Due to the complexity of thermal dynamics, system identification from data is considered advantageous and there has been much work on it; see (Penman, 1990; Braun & Chaturvedi, 2002; Wang & Xu, 2006; Lin, Middelkoop, & Barooah, 2012; Li & Wen, 2014; Harb, Boyanov, Hernandez, Streblow, & Müller, 2016; Fux, Ashouri, Benz, & Guzzella, 2014; Kim, Cai, Ariyur, & Braun, 2016; Hu et al., 2016) and references therein. A particular challenge for model identification is that temperature is affected by large, unknown disturbances, especially the cooling load induced by the occupants. The occupant induced load refers to the heat gain directly due to the occupants' body heat and indirectly from lights and other equipment they use. Many system identification methods ignore these disturbances (Lin et al., 2012; Li & Wen, 2014; Harb et al., 2016), or use a specialized test building to measure the occupant induced load (Penman, 1990; Wang & Xu, 2006). Ignoring the disturbance can produce highly erroneous results (Kim et al., 2016).

In this paper we propose a method to estimate a dynamic model as well as a transformed version of the unknown disturbances from easily measurable input-output data. The proposed method, which we call SPDİR (Simultaneous Plant and Disturbance Identification through Regularization) is based on solving a ℓ_1 -regularized least-squares problem. The ℓ_1 penalty encourages the identified transformed disturbance to be sparse (Hastie, Tibshirani, & Friedman, 2009). The motivation for this is that the disturbance, which consists mostly of internal load due to occupants, is often piecewise-constant. For instance, large numbers of people enter and leave office buildings at approximately the same time. We show that this makes the transformed disturbances an approximately sparse signal, motivating the use of ℓ_1 regularization. We test our method via simulations, and results indicate that the method can estimate the thermal dynamic model and transformed disturbance with both open loop and closed loop data, even when the disturbance is not piecewise constant.

To the best of our knowledge, the only prior work on simultaneously identifying a dynamic model of a building's temperature dynamics and unmeasured disturbances from data are the recent references (Kim et al., 2016; Fux et

*This research is partially supported by NSF grants 1463316 and 1646229.

al., 2014; Hu et al., 2016; Coffman & Barooah, 2018). There are many differences between these references and our work. The method proposed in (Kim et al., 2016) estimates the plant parameters and an output disturbance (a disturbance that is added to the plant output) that encapsulates the effect of an unknown input disturbance. In contrast, the proposed method estimates an input disturbance. Both (Fux et al., 2014) and (Hu et al., 2016) take a similar approach: the model is estimated by using data from unoccupied periods (weekends in (Hu et al., 2016)) and assuming that the disturbance is zero during those periods. Once the model is identified this way, the disturbance is identified using data from occupied periods. Our method uses data collected during regular operation of a building and does not need data collected when the building is empty. Even when data from unoccupied periods is available, assuming the disturbance to be zero during that time is not desirable since doing so will prevent the disturbance from absorbing model mismatch. In contrast to all three methods, the method proposed here can enforce properties of the system that are known from the physics of the thermal processes, such as stability and signs of DC gains for certain input-output pairs. The proposed method consists of solving a convex optimization problem. The methods in (Kim et al., 2016; Fux et al., 2014) require solving non-convex optimization problems.

The rest of this paper is organized as follows. Section 2 formally describes the problem and establishes some properties that will be useful later. Section 3 describes the proposed algorithm. We provide evaluation results in Section 4. Finally, Section 5 concludes this work.

2. PROBLEM FORMULATION

The indoor zone temperature T_z is affected by three *known* inputs: (1) the heat gain added to the zone by the HVAC system, q_{hvac} (kW), (2) the outside air temperature T_{oa} ($^{\circ}\text{C}$), (3) the solar irradiance η^{sol} (kW/m²), and the *unknown* disturbance q_{int} (kW) which is the internal heat gain due to occupants, lights, and equipment used by the occupants. So $u(t) := [q_{\text{hvac}}(t), T_{oa}(t), \eta^{\text{sol}}(t)]^T \in \mathbb{R}^3$ and $w(t) = q_{\text{int}}(t) \in \mathbb{R}$. The only measurable output is the indoor temperature T_z , so $y(t) = T_z(t) \in \mathbb{R}$.

The model we wish to identify is a black box model relating the known inputs and the unknown disturbance, to the measured output. We will later enforce constraints on the model's parameters by relating the model to a physics-based, second-order continuous-time model, making it an "grey-box" model.

2.1 Discrete-Time Model to be Identified

We start with the following second-order discrete-time transfer function model of the system, with a sampling time t_s :

$$y(z^{-1}) = \frac{1}{D(z^{-1})} \left[\sum_{j=1}^3 \left[\sum_{i=0}^2 \alpha_{ij} z^{-i} \right] u_j(z^{-1}) + \left[\sum_{i=0}^2 \beta_i z^{-i} \right] w(z^{-1}) \right] \quad (1)$$

where $D(z^{-1}) = 1 - \theta_1 z^{-1} - \theta_2 z^{-2}$, for some parameters θ_1, θ_2 and α_{ij}, β_i 's, and $u[k], w[k], y[k]$ are samples of the continuous-time signals $u(t), w(t), y(t)$. For future convenience, we rewrite it as

$$y(z^{-1}) = \frac{1}{D(z^{-1})} \left[K(z^{-1})^T u(z^{-1}) + \bar{w}(z^{-1}) \right] \quad (2)$$

where

$$K(z^{-1}) := \begin{bmatrix} \theta_3 z^{-2} + \theta_4 z^{-1} + \theta_5 \\ \theta_6 z^{-2} + \theta_7 z^{-1} + \theta_8 \\ \theta_9 z^{-2} + \theta_{10} z^{-1} + \theta_{11} \end{bmatrix} \quad (3)$$

and $\bar{w}(z^{-1})$ is the Z-transform of the *transformed disturbance* signal $\bar{w}[k]$ defined as

$$\bar{w}[k] := \beta_0 w[k] + \beta_1 w[k-1] + \beta_2 w[k-2]. \quad (4)$$

Performing an inverse Z-transform on (2)-(3), yields a difference equation, from which we obtain the linear regression form:

$$y[k] = \phi[k]^T \theta, \quad k = 3, \dots, k_{\max} \quad (5)$$

where $\theta^T := [\theta_p^T, \bar{w}^T]$, in which $\theta_p = [\theta_1, \dots, \theta_{11}]^T \in \mathbb{R}^{11}$, $\bar{w} = [\bar{w}_3, \dots, \bar{w}_{k_{\max}}]^T \in \mathbb{R}^{k_{\max}-2}$ and

$$\phi[k]^T := [y[k-1], y[k-2], u_1[k-2], u_1[k-1], u_1[k], u_2[k-2], \dots, u_2[k], u_3[k-2], \dots, u_3[k], e_{k-2}^T],$$

where e_k is the k -th canonical basis vector of $\mathbb{R}^{k_{\max}-2}$ in which the 1 appears in the k^{th} place, and $k_{\max}$ is the number of samples. Eq. (5) can be expressed as

$$y = \Phi \theta, \quad (6)$$

where $y := [y[3], \dots, y[k_{\max}]]^T \in \mathbb{R}^{k_{\max}-2}$ and

$$\Phi := \begin{bmatrix} \phi[3]^T \\ \vdots \\ \phi[k_{\max}]^T \end{bmatrix} \in \mathbb{R}^{k_{\max}-2 \times k_{\max}+9}.$$

The problem we seek to address is: *given time traces of inputs and outputs, $\{u[k], y[k]\}_1^{k_{\max}}$, determine the unknown parameter vector $p \in \mathbb{R}^{11}$ and the unknown transformed disturbance vector $\bar{w} := [\bar{w}_3, \dots, \bar{w}_{k_{\max}}]$, i.e., determine θ .*

It should be noted that the matrix Φ is not full column-rank, so there will be an infinite number of solutions to (6). We will use insights from a physics-based model to impose additional constraints on θ .

2.2 Insight from an RC Network ODE Model

Figure 1 shows a building (left) and a corresponding 2nd-order resistance-capacitance (RC) network model (right). RC-networks are a common modeling paradigm for building thermal dynamics (R.Kramer, Schijndel, & H.Schellen,

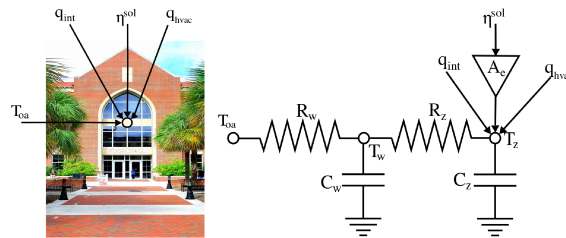


Figure 1: A photograph for Pugh Hall and a schematic of the “2R2C” model.

2012). The RC network model of the building shown in Figure 1 consists of two coupled ODE’s (ordinary differential equations):

$$\begin{aligned} C_z \dot{T}_z &= \frac{T_w - T_z}{R_z} + q_{\text{hvac}} + A_e \eta^{\text{sol}} + q_{\text{int}} \\ C_w \dot{T}_w &= \frac{T_{0a} - T_w}{R_w} + \frac{T_z - T_w}{R_z} \end{aligned} \quad (7)$$

where C_z, C_w, R_z, R_w are the thermal capacitances and resistances of the zone and wall, respectively, and A_e is the effective area of the building for incident solar radiation. All five parameters are positive. Defining the state vector as $x := [T_z, T_w]^T \in \mathbb{R}^2$, the ode can be expressed as:

$$\dot{x} = Fx + Gu + Hw, \quad y = Jx, \quad (8)$$

where u, w , and y are defined in Section 2, and $F \in \mathbb{R}^{2 \times 2}, G \in \mathbb{R}^{2 \times 3}, H \in \mathbb{R}^{2 \times 1}$ and $J \in \mathbb{R}^{1 \times 2}$ are appropriate matrices that are functions of the parameters C_z, C_w, R_z, R_w, A_e . In Laplace domain,

$$y(s) = \frac{1}{D(s)} \left[(s - f_{22}) (g_{11}u_1(s) + g_{13}u_3(s)) + f_{12}g_{22}u_2(s) + (s - f_{22})h_{11}w(s) \right] \quad (9)$$

where f_{ij}, g_{ij}, h_{ij} are the i, j -th entry of the matrices F, G, H (respectively) in (8), and

$$D(s) = s^2 + d_1 s + d_2,$$

$$\text{with } d_1 = \frac{1}{C_z R_z} + \frac{1}{C_w} \left(\frac{1}{R_z} + \frac{1}{R_w} \right), \quad d_2 = \frac{1}{C_z C_w R_z R_w}.$$

We now assume that the discrete-time system (1) was obtained by discretizing the continuous-time system (9) by Tustin transformation. It can be shown through straightforward calculations that the parameters of the discrete-time model – the θ_i 's – are related to those of the continuous-time model (9) as follows:

$$\theta_1 := \frac{8 - 2d_2 t_s^2}{D_0}, \quad \theta_2 := -\frac{d_2 t_s^2 - 2d_1 t_s + 4}{D_0},$$

$$\begin{bmatrix} \theta_3 & \theta_9 \\ \theta_4 & \theta_{10} \\ \theta_5 & \theta_{11} \end{bmatrix} := \frac{t_s}{D_0} \begin{bmatrix} -2 - f_{22} t_s \\ -2f_{22} t_s \\ 2 - f_{22} t_s \end{bmatrix} \begin{bmatrix} g_{11} & g_{13} \end{bmatrix}, \quad \begin{bmatrix} \theta_6 \\ \theta_7 \\ \theta_8 \end{bmatrix} := \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \frac{f_{12} g_{22} t_s^2}{D_0},$$

where $D_0 = d_2 t_s^2 + 2d_1 t_s + 4$. Similarly,

$$[\beta_0, \beta_1, \beta_2] = \frac{t_s [(2 + \varepsilon_0), 2\varepsilon_0, (-2 + \varepsilon_0)]}{C_z D_0}, \quad (10)$$

$$\text{where } \varepsilon_0 = -f_{22} t_s = \frac{t_s}{C_w} \left(\frac{1}{R_w} + \frac{1}{R_z} \right). \quad (11)$$

2.2.1 Insight I: Sparsity of transformed disturbance: We need a few definitions to talk about *approximately sparse* vectors, and *slowly varying* vectors.

Definition 1 1. A vector $x \in \mathbb{R}^n$ is (ε, f) -sparse if at most f fraction of entries of x are not in $[-\varepsilon, \varepsilon]$.

2. The change frequency $c_f(x)$ of a vector $x \in \mathbb{R}^n$ is the fraction of entries that are distinct from their previous neighbor: $c_f(x) = \frac{1}{n-1} |\{k > 1 | x_k \neq x_{k-1}\}|$, where $|A|$ denotes the cardinality of the set A . We say a vector x changes infrequently if $c_f(x) \ll 1$.

The following result shows that if the disturbance does not change frequently (which happens if it is piecewise constant), then the transformed disturbance is approximately sparse.

Proposition 1 Suppose the disturbance $w[k]$ is uniformly bounded in k , it changes infrequently, and $\varepsilon_0 \ll 1$ where ε_0 is defined in (11). Then, $\bar{w}[k]$ is $(\bar{\varepsilon}, 2c_f(w))$ -sparse, where $\bar{\varepsilon} = \frac{4}{C_z D_0} t_s w_u \varepsilon_0$ and w_u is an upper bound on $|w[k]|$. $\square$

Proof 1 It can be shown from (4) and (10) that

$$\bar{w}[k] = \frac{t_s}{C_z D_0} (2(w[k] - w[k-2]) - \varepsilon_0(w[k] + 2w[k-1] + w[k-2]))$$

Since w is bounded, $\exists w_l, w_u$ with $|w_l| \leq w_u$, $w_u \geq 0$ s.t. $w[k] \in [w_l, w_u]$. Since $c_f(w) \ll 1$ from the hypothesis, for at least $1 - 2c_f(w)$ fraction of k 's, $w[k] - w[k-2] = 0$, and for those k 's,

$$\bar{w}[k] = -\varepsilon_0 \frac{t_s}{C_z D_0} (w[k] + 2w[k-1] + w[k-2]) \in \left[\frac{-4\varepsilon_0 t_s w_u}{C_z D_0}, \frac{4\varepsilon_0 t_s w_l}{C_z D_0} \right] \subset \left[\frac{-4\varepsilon_0 t_s w_u}{C_z D_0}, \frac{4\varepsilon_0 t_s w_u}{C_z D_0} \right],$$

which proves the result.

Since the product RC is large for large buildings, of the order of few hours (Kim et al., 2016), it follows from (11) that ε_0 is small for such buildings. In addition, both ε_0 and $\bar{\varepsilon}$ can be made as small as possible by choosing t_s sufficiently small. The assumption in the proposition, that ε_0 is small, is therefore not a strong one.

2.2.2 Insight II: Constraints on parameters: The constraints described below are straightforward to derive, but involve - in a few cases - extensive algebra. We therefore omit the details here; they can be found in the expanded version (Zeng, Brooks, & Barooah, 2017).

Stability It can be shown that due to the resistances and capacitances in (7) being positive, the continuous time model (9) is BIBO stable. Since Tustin transformation preserves stability, all poles of the transfer function (1) should be inside the unit circle (Ogata, 1995). It can be shown that this is equivalent to

$$-\theta_2 < 1, \quad \theta_2 + \theta_1 < 1, \quad (12)$$

$$\theta_2 - \theta_1 < 1. \quad (13)$$

Sign of parameters By using the positivity of the parameters R_w, R_z, C_w, C_z , it can be shown after some tedious algebra that if $t_s < 2\min\{\frac{C_w R_w R_z}{R_z + R_w}, \sqrt{R_z C_z R_w C_w}, \frac{\min(R_z C_z, R_z C_w, R_w C_w)}{3}\}$, the following holds:

$$\theta_i > 0, i \in \{1, 4, 5, 6, 7, 8, 10, 11\},$$

$$\theta_2 < 0, \theta_3 < 0, \theta_9 < 0$$

Positive DC-gain An increase in any of the inputs $q_{\text{hvac}}, T_{\text{oa}}, \eta^{\text{sol}}$ represents an increase in the cooling load for the building. A steady state increase in any of these inputs must therefore lead to a steady state increase in the indoor temperature T_z . In other words, the corresponding DC gains must be positive. Using the previously established fact that the denominator coefficients are positive (see (12)) it can be shown that positive DC gains are equivalent to

$$\theta_3 + \theta_4 + \theta_5 > 0,$$

$$\theta_6 + \theta_7 + \theta_8 > 0, \quad (14)$$

$$\theta_9 + \theta_{10} + \theta_{11} > 0.$$

In order to ensure existence of a solution (Luenberger, Ye, et al., 1984), the above constraints are relaxed from a strict inequality to a non-strict one. Additionally, the redundant inequalities (13) and (14) are removed since they do not change the feasible region, where the proof is provided in (Zeng et al., 2017).

The remaining, linearly independent constraints can be compactly written as $g(\theta) \leq 0$ with $g : \mathbb{R}^{11} \rightarrow \mathbb{R}^{15}$, in which the inequality is entry-wise.

3. PROPOSED METHOD

Let $S := [0_{k_{\max}-2 \times 11} \mid I_{k_{\max}-2}]$ so that $S\theta = \bar{w}$. Since we expect w to be piecewise constant and slowly varying, $\bar{w}$ should be approximately sparse (Proposition 1). We thus seek a solution to $\mathbf{y} = \Phi\theta$ so that $S\theta$ is sparse, by posing the following optimization problem

$$\begin{aligned} \hat{\theta} = \arg \min_{\theta} \quad & \frac{1}{2} \|\mathbf{y} - \Phi\theta\|_2^2 + \lambda \|S\theta\|_1 \\ \text{s. t. } \quad & g(\theta) \leq 0 \end{aligned} \quad (15)$$

where $\lambda > 0$ is a user-defined weighting factor. The ℓ_1 norm penalty is to encourage sparsity. The estimated plant parameters $\hat{\theta}_p$ and estimated transformed disturbance $\hat{w}$ can be recovered from $\hat{\theta}$ since $\theta^T = [\theta_p^T, \bar{w}^T]$.

Regularity of constraints is useful for optimization algorithms to perform well (Andreani, Martinez, Santos, & Svaiter, 2014), and the next result establishes regularity. The proof of the result is straightforward and available in the expanded version (Zeng et al., 2017), but omitted here due to space constraints.

We call a point θ *physically meaningful* if none of the three SISO transfer functions in (2) is identically zero.

Proposition 2 *The optimization problem (15) is feasible, convex, and every physically meaningful feasible θ is a regular point of the constraints.*

The optimization problem (15) is convex. All numerical results presented in this paper were obtained by using the `cvx` package for solving convex problems in MATLAB[®] (Grant & Boyd, 2011).

3.1 Regularization Parameter Selection

Theorem 1 (Duan et al., 2016) *For $\Phi \in \mathbb{R}^{n \times p}$, $S \in \mathbb{R}^{m \times p}$, $\tilde{G}_1 \in \mathbb{R}^{k \times p}$, if matrix $H = [\Phi^T, S^T, \tilde{G}_1^T]^T$ has full column rank and $[S^T, \tilde{G}_1^T]^T$ has full row rank, then the generalized Lasso (15) with active constraints written as $-\tilde{G}_1 \theta + \tilde{G}_0 \leq 0$ can be transformed into the following classic Lasso form:*

$$\begin{aligned} \hat{\chi} = \arg \min_{\chi} \quad & \frac{1}{2} \|z - \Psi \chi\|_2^2 + \lambda \|\chi\|_1 \\ \text{s. t.} \quad & -\bar{F}_1 \chi + \bar{F}_0 \leq 0 \end{aligned} \quad (16)$$

where

$$\begin{aligned} \chi &= S\theta & \bar{F}_1 &= Q_3 Q_2^\dagger \\ \Psi &= Q_1 Q_2^\dagger & \bar{F}_0 &= -Q_3 Q_1^\dagger y \\ z &= [I_n - Q_1(I_p - Q_0^\dagger Q_0)Q_1^\dagger]y - Q_1 Q_3^\dagger \tilde{G}_0, \end{aligned}$$

and

$$H = \begin{bmatrix} \Phi \\ S \\ \tilde{G}_1 \end{bmatrix} = QR = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} R = \begin{bmatrix} Q_1 \\ Q_0 \end{bmatrix} R, \quad Q_1 \in \mathbb{R}^{n \times p}, Q_2 \in \mathbb{R}^{m \times p}, Q_3 \in \mathbb{R}^{k \times p}, Q_0 \in \mathbb{R}^{(m+k) \times p}.$$

Our problem falls into the category of “generalized Lasso” and can be transformed into the classic form as it satisfy conditions in theorem 1.

Two common heuristics for choosing λ for the classic Lasso problem are cross-validation (Pendse, 2011) and L-curve-based curvature methods (Hansen, 1992). However, neither of them is applicable to our problem as they have implementation requirements that our problem does not satisfy.

Cross validation divides datasets into K folders and requires that parameters to be retrieved are the same among such folders, whereas parameters in our problem contain transformed disturbance, which may differ from one day to the next. The L-curve, which is “a log-log plot of the norm of a regularized solution $\|\chi\|_1$ versus the norm of the corresponding residual norm $\|z - \Psi \chi\|_2$ ”, can graphically display the trade-off between the size of a regularized solution and its fit to the given data, where optimal regularization parameter that minimize the trade-off lies at the corner of such L-curve. For the L-curve method, a solution path that changes monotonically with respect to λ is essential, i.e., $(\Psi^T \Psi)^{-1}$ in (16) needs to be diagonally dominant (Duan et al., 2016). That too is also not satisfied in our case.

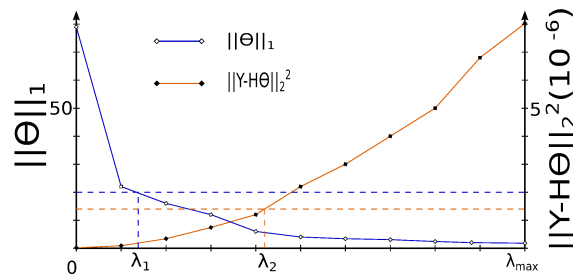


Figure 2: Illustration of regularization parameter selection

Theorem 2 For classic Lasso, $\exists \lambda_{\max} = \|\Psi^T z\|_{\infty}$, such that $\forall \lambda \geq \lambda_{\max}$, $\chi = 0$ is a solution to (16).

Proof 2 From optimality condition for (16), we know

$$-\Psi^T(z - \Psi\chi) + \lambda v = 0, \quad (17)$$

where v is the sub-gradient of χ as defined as

$$v_i \in \begin{cases} \{+1\}, & \chi_i > 0 \\ \{-1\}, & \chi_i < 0, \\ [-1, 1], & \chi_i = 0. \end{cases} \quad i = 1, \dots, p \quad (18)$$

Substituting $\lambda \geq \|\Psi^T z\|_{\infty}$ and $\chi = 0$ into (17) and solving for v , we get $v = \frac{\Psi^T z}{\|\Psi^T z\|_{\infty}}$, which implies (18) also holds as $\|v_i\| \leq 1, \forall i$. Therefore we prove $\chi = 0$ is one solution to (16).

According to Theorem 2, $\chi = 0$ is a solution to (16) $\forall \lambda \geq \lambda_{\max}$. Notice that λ in (15) and (16) are identical, that is, $\lambda_{\max}$ determined from (16) can be directly used in (15). Combined with Theorem 1, we have $\chi = S\theta = [\theta_p^T, \bar{w}^T]^T = 0$ is a solution to (15) $\forall \lambda \geq \lambda_{\max}$.

We use the following heuristic to choose λ , which is inspired by the L-curve method. First, plot both the solution norm and residual norm individually against λ by repeatedly solving Problem (15) for $\lambda \in [0, \lambda_{\max}]$. An illustration of these two plots is shown in Figure 2. Second, identify a value λ_1 so that the solution norm is smaller than a user-defined threshold for any $\lambda > \lambda_1$, and then identify λ_2 so that the residual norm is smaller than a user-defined threshold for any $\lambda < \lambda_2$. If $\lambda_2 > \lambda_1$, choose λ to be λ_1 . If not, pick another threshold, and continue until this condition is met.

4. EVALUATION OF PROPOSED ALGORITHM

The continuous-time RC model (8) is used to generate training and validation data. The parameters of the model were chosen by manual calibration of the model to data collected from a large auditorium in a campus building at the University of Florida (Pugh Hall; shown in Figure 1). Four scenarios are tested:

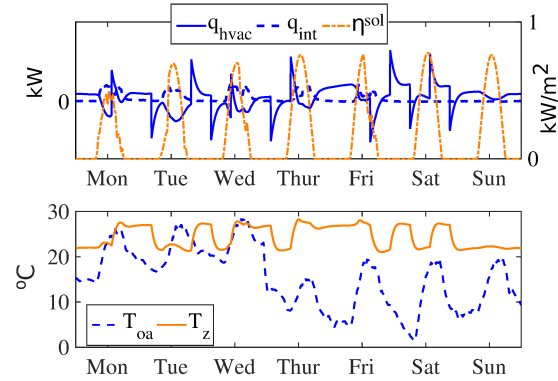
1. *OL-PW*: Open-loop with piecewise-constant disturbance;
2. *OL-NPW*: Open-loop with not piecewise-constant disturbance;
3. *CL-PW*: Closed-loop with piecewise-constant disturbance;
4. *CL-NPW*: Closed-loop with not piecewise-constant disturbance;

If the disturbance $w[k]$ is piecewise constant, since that is slowly varying, the transformed disturbance $\bar{w}[k]$ will be approximately sparse. The algorithm is expected to perform well in the OL-PW scenario since it satisfies the piecewise constant assumption the method is based on, and identification with open-loop data is generally easier than with closed loop (Ljung, 1999). The CL-NPW scenario is the most relevant in practice, but it is likely to be the most challenging for the method. In all four scenarios, the same input data sequence for ambient temperature (from weatherunderground.com) and solar irradiance data (from NSRDB: <https://nsrdb.nrel.gov/>), both for Gainesville, FL, are used. In the two open-loop scenarios, the input component q_{hvac} is somewhat arbitrarily chosen, while in the two closed-loop scenarios, it is decided by a PI-controller that tries to maintain the zone temperature near a setpoint T^{ref} . To have exciting input to aid in identification, the setpoint T^{ref} is chosen to be a PRBS sequence (Ljung, 1999). To ensure that occupant comfort is not compromised, the setpoint is constrained to lie within 22°C and 27°C. The disturbance signal q_{int} is picked somewhat arbitrarily during manual calibration of the RC network model to Pugh Hall data. The training data are shown in Figure 3.

Notice from the figure that the disturbance q_{int} is large; sometimes as large as the cooling power provided by the HVAC system.

Table 1: Plant parameters and errors in their estimates.

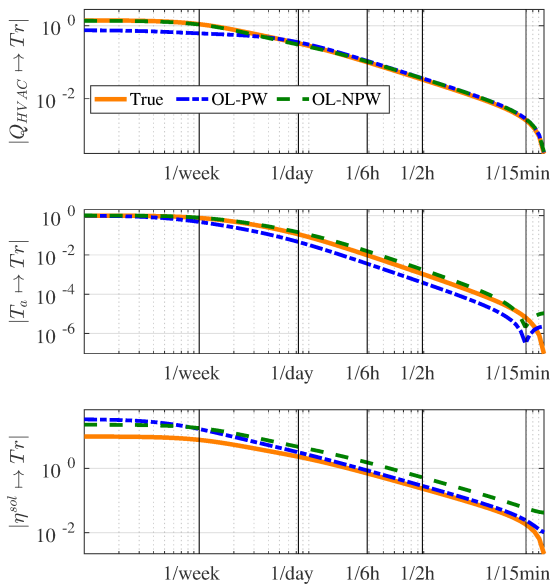
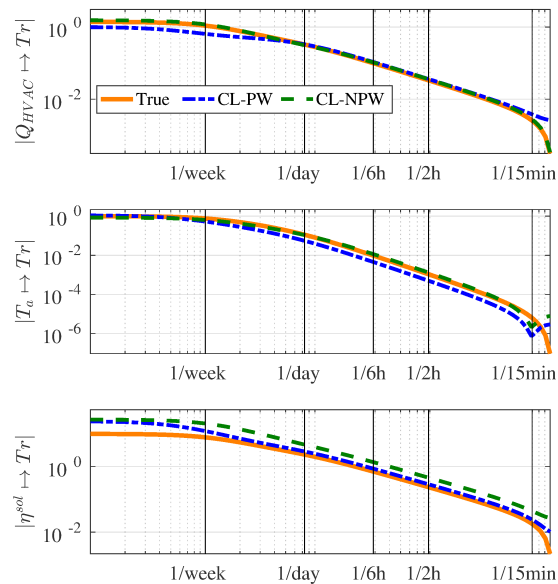
	θ	$\frac{\theta - \hat{\theta}}{\hat{\theta}} \%$		input
		(OL-PW)	(CL-NPW)	
θ_1	1.98×10^{-0}	-0.075	0.042	q_{hvac}
θ_2	-9.76×10^{-1}	-0.151	0.085	
θ_3	-4.35×10^{-3}	-9.214	-8.024	
θ_4	5.21×10^{-5}	-59.48	-108.2	
θ_5	4.40×10^{-3}	-7.493	-6.36	
θ_6	1.86×10^{-5}	-18.64	-48.90	T_{oa}
θ_7	3.72×10^{-5}	38.15	22.35	
θ_8	1.86×10^{-5}	-39.89	-68.32	
θ_9	-3.05×10^{-2}	-112.6	-232.1	η^{sol}
θ_{10}	3.65×10^{-4}	-12300	-19320	
θ_{11}	3.08×10^{-2}	33.18	-2.881	

**Figure 3:** Training data for algorithm evaluation. The data η^{sol} , T_{oa} , q_{int} shown here are used in all four scenarios; q_{hvac} , T_z shown here are for the CL-NPW scenario.

4.1 Plant Identification Results

Parameters Table 1 shows the true values of the plant parameters, $\theta_1, \dots, \theta_{11}$. It also shows the corresponding estimation errors (in percentage) for the OL-PW and CL-NPW scenarios. We can see from the table that performance of the method is similar with both open-loop and closed-loop data. Second, the two parameters that determine the characteristic equation are estimated highly accurately. Third, there is more error in the estimate of numerators. While some are more accurate than others, the numerator coefficients corresponding to the input η^{sol} has the most error. A possible reason for this high error is the lack of richness in the η^{sol} data. See Figure 3: η^{sol} is the least rich among all the input signals.

Results for the remaining two scenarios are similar, but are not shown due to space constraints.

**Figure 4:** Bode magnitude plots of true and identified systems for OL scenarios.**Figure 5:** Bode magnitude plots of true and identified systems for CL scenarios.

Frequency response For prediction accuracy, frequency response is more important than individual parameters. Figure 4 and 5 compare the frequency response of the identified plants with their true values for the two open-loop and closed-loop scenarios, respectively. Notice that just as in case of parameters, the estimates corresponding to the input η^{sol} is the poorest. We believe this is due to the lack of sufficient excitation in the data; cf. Figure 3. The input T_{oa} also has low excitation at higher frequencies, and therefore has poor estimates in higher frequencies.

For the transfer functions from inputs q_{hvac} to output T_z , the maximum absolute error in the estimated frequency response is:

$$\max_{\omega} \frac{|\hat{G}_{q_{\text{hvac}} T_z}(j\omega) - G_{q_{\text{hvac}} T_z}(j\omega)|}{|G_{q_{\text{hvac}} T_z}(j\omega)|} = 0.529$$

and occurs at Nyquist frequency for CL-PW scenario. The maximum errors for the transfers functions from T_{oa} and η_{sol} to T_z occur at the Nyquist frequency.

4.2 Disturbance

The estimated transformed disturbance, $\hat{w}$, for all four scenarios are shown in Fig. 6. The estimates are quite accurate when the true values are large, but less accurate otherwise. However, the estimates capture the trend of the true values quite accurately, even when the true disturbance is not piecewise constant, in which case the transformed disturbance may be neither approximately sparse nor slowly varying.

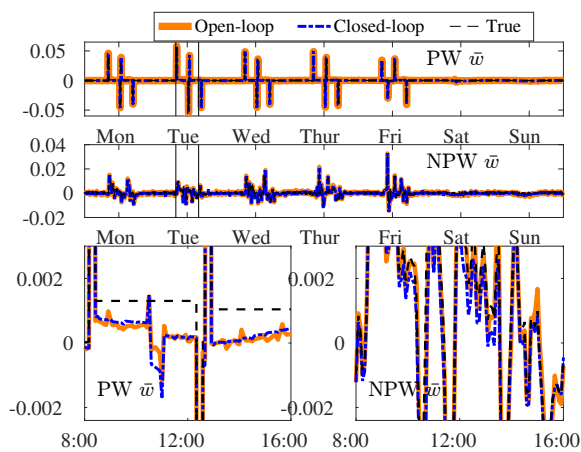


Figure 6: Comparison of identified and actual transformed disturbance. Bottom two plots are zoomed version on Tuesday of the top two plots.

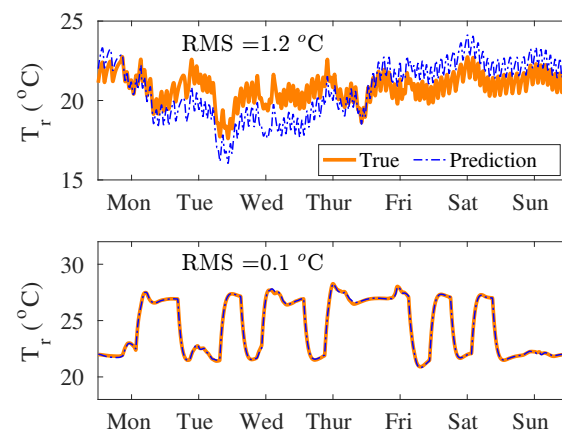


Figure 7: Predicted and actual zone temperature (validation dataset). Upper: OL-PW; Lower: CL-NPW.

4.3 Validation through Temperature Prediction

The plant identified with data from one week is used to predict temperatures in another week. The disturbance data is the same between the calibration and validation data sets but the input u and output y data sets are distinct. The RMS value of the prediction error of zone temperature is 1.2 °C for OL-PW case and 0.1 °C for CL-NPW case; see Figure 7. For validation, we use the solar irradiance and ambient temperature from the validation data set; the temperature setpoint is arbitrarily picked as another PRBS sequence lies within [22,27] °C, and the disturbance is piecewise-constant (see Figure 3). Compared to the large inaccuracies in the estimated plant parameters, predictions of the zone temperature, shown in Figure 7, are much smaller. As we can see from the figure, the error is more pronounced in some days of the week, while extremely small in other days.

5. CONCLUSION

The main advantages of the method is posing the estimation problem as a convex optimization problem with constraints from physical insights about the system and the disturbances, without requiring specially collected data. Previous

methods lacked both convexity and/or physically meaningful constraints. The main limitation is that the identified disturbance is a linear transformation of the true disturbance with unknown coefficients. This presents a challenge in verifying the disturbance estimates when the method is applied to data from a real building. Extracting w from $\bar{w}$, so that the estimate can be verified in a test setting, is a topic of future work.

REFERENCES

- Andreani, R., Martinez, J., Santos, L., & Svaiter, B. (2014). On the behaviour of constrained optimization methods when Lagrange multipliers do not exist. *Optimization Methods and Software*, 29(3), 646–657. doi: 10.1080/10556788.2013.841692
- Braun, J. E., & Chaturvedia, N. (2002). An inverse gray-box model for transient building load prediction. *HVAC&R Research*, 8, 73–99. doi: 10.1080/10789669.2002.10391290
- Coffman, A., & Barooah, P. (2018). Simultaneous identification of dynamic model and occupant-induced disturbance for commercial buildings. *Building and Environment*, 128(153–160). Retrieved from <https://doi.org/10.1016/j.buildenv.2017.10.020>
- Duan, J., Soussen, C., Brie, D., Idier, J., Wan, M., & Wang, Y.-P. (2016). Generalized lasso with under-determined regularization matrices. *Signal processing*, 127, 239–246.
- Fux, S. F., Ashouri, A., Benz, M. J., & Guzzella, L. (2014). EKF based self-adaptive thermal model for a passive house. *Energy and Buildings*, 68, 811–817.
- Grant, M., & Boyd, S. (2011, February). CVX: Matlab software for disciplined convex programming, version 1.21. <http://cvxr.com/cvx>.
- Hansen, P. C. (1992). Analysis of discrete ill-posed problems by means of the l-curve. *SIAM review*, 34(4), 561–580.
- Harb, H., Boyanov, N., Hernandez, L., Streblow, R., & Müller, D. (2016). Development and validation of grey-box models for forecasting the thermal response of occupied buildings. *Energy and Buildings*, 117, 199–207.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Second ed.). Springer.
- Hu, Q., Oldewurtel, F., Balandat, M., Vrettos, E., Zhou, D., & Tomlin, C. J. (2016). Building model identification during regular operation-empirical results and challenges. In *Proceedings of the american control conference* (pp. 605–610).
- 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. doi: <http://dx.doi.org/10.1016/j.buildenv.2016.07.007>
- Li, X., & Wen, J. (2014). Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Review*, 37, 517–537.
- Lin, Y., Middelkoop, T., & Barooah, P. (2012, December). Issues in identification of control-oriented thermal models of zones in multi-zone buildings. In *IEEE conference on decision and control* (p. 6932 – 6937). doi: 10.1109/CDC.2012.6425958
- Ljung, L. (1999). *System identification: Theory for the user* (2nd ed.). Prentice Hall.
- Luenberger, D. G., Ye, Y., et al. (1984). *Linear and nonlinear programming* (Vol. 2). Springer.
- Ogata, K. (1995). *Discrete-time control systems* (Vol. 8). Prentice-Hall Englewood Cliffs, NJ.
- Pendse, G. V. (2011). A tutorial on the lasso and the shooting algorithm. *Technical report, PAIN Group, Imaging and Analysis Group-McLean Hospital, Harvard Medical School*, 8.
- Penman, J. (1990). Second order system identification in the thermal response of a working school. *Building and Environment*, 25(2), 105–110.
- Prívara, S., Cigler, J., Váňa, Z., Oldewurtel, F., Sagerschnig, C., & Žáčková, E. (2013). Building modeling as a crucial part for building predictive control. *Energy and Buildings*, 56(0), 8 – 22. doi: <http://dx.doi.org/10.1016/j.enbuild.2012.10.024>
- R.Kramer, Schijndel, J., & H.Schellen. (2012). Simplified thermal and hygric building models: a literature review. *Frontiers of Architectural Research*, 1, 318–325.
- Wang, S., & Xu, X. (2006). Parameter estimation of internal thermal mass of building dynamic models using genetic algorithm. *Energy conversion and management*, 47(13), 1927–1941.
- Zeng, T., Brooks, J., & Barooah, P. (2017). Simultaneous identification of building dynamic model and disturbance using sparsity-promoting optimization. *ArXiv.org*. (arXiv:1711.06386)