Managing Communication Delays and Model Error in Demand Response for Frequency Regulation

Journal:	*IEEE Transactions on Power Systems*
Manuscript ID	Draft
Manuscript Type:	Transactions
Date Submitted by the Author:	n/a
Complete List of Authors:	Ledva, Gregory; University of Michigan, COE EECS - ECE Division ; Vrettos, Evangelos; Swiss Federal Institute of Technology (ETH Zurich), Information Techonolgy and Electrical Enginnering Mastellone, Silvia; ABB Corporate Research Center Switzerland Andersson, Goran; Eidgenossische Technische Hochschule Zurich, Information Technology and Electrical Engineering Mathieu, Johanna; University of Michigan, EECS
Technical Topic Area :	Power system control, Power system energy management
Key Words:	Air conditioning, Delay systems, Power demand, State estimation, Predictive control, On-off control, Thermal energy storage, Energy management

Managing Communication Delays and Model Error in Demand Response for Frequency Regulation

Gregory S. Ledva, *Student Member, IEEE,* Evangelos Vrettos, *Student Member, IEEE,* Silvia Mastellone, Göran Andersson, *Fellow, IEEE,* and Johanna L. Mathieu, *Member, IEEE*

***Abstract*—This work develops and compares several networked control and estimation algorithms to manipulate the power consumption of a population of residential thermostatically controlled loads to fulfill PJM frequency regulation requests given an imperfect communication network and modeling error. The algorithms rely on a model of the plant to reduce the effects of communication delays, and include a stochastic, predictive controller and two Kalman filter-based state estimation techniques. The first estimator uses a set of independent Kalman filters that run in parallel, and the second incorporates individual TCL models that rely on identified thermal parameters.**

We use simulations to examine 1) the algorithms' ability to adequately provide frequency regulation under a range of delay severities, and 2) the effect of increased modeling error. We find that both estimator-controller combinations provide acceptable frequency regulation with average delays of 20 seconds and minor modeling error. When we increase the modeling error by using a higher-order model to represent the TCLs within the plant, the first estimator provides acceptable frequency regulation while the second estimator provides poor frequency regulation.

***Index Terms*—Demand response, frequency regulation, delays, state estimation, optimal control, networked control**

I. Introduction

INCORPORATING more fluctuating, renewable power generation into the electricity network will usually lead to additional power production variability. To maintain the frequency within an acceptable range, generation resources must supply more reserves, which may require them to operate at inefficient operating points [1]. Alternatively, the manipulation of electric power demand using demand response is also capable of providing frequency regulation.

Common residential demand response methods include price-based demand manipulation and direct control of loads [2], e.g., residential thermostatically controlled loads (TCLs) such as air conditioners, heat pumps, and water heaters. Under normal operation, TCLs cycle on and off to maintain the temperature of an internal medium, e.g., a house's air temperature, around a user-defined set-point. Direct control strategies manipulate a TCL population's total power demand generally by adjusting either the user-defined temperature set-point, e.g., [3]–[5], or by requesting additional on/off switching, e.g., [6]–[8]. Aggregations of TCLs can be used to provide ancillary services such as frequency regulation to the power system [5]. Recently, researchers have developed non-disruptive load control strategies [9], ensuring TCLs operate within or very close to their normal temperature range [5].

TCLs are a spatially distributed resource, and coordinating the demand of thousands of them to provide frequency regulation requires sensing and communication infrastructure. This infrastructure enables TCLs to receive control inputs and to send information about their current operating state. However, the cost of this infrastructure can be prohibitive [10]. Using existing infrastructure, such as smart meters, is possible but the frequency of information retrieval is limited [11], and load control input delays can be significant [12]. Developing control algorithms for demand response that are robust to delays and respect the limitations of existing infrastructure may lower the cost of demand response implementations.

Networked control theory addresses imperfect communication within control systems, see e.g., [13]. Ref. [14] investigates the impact of delays in frequency regulation including batteries and develops control algorithms to limit their effects. Within the demand response literature, [6]–[8], [15], [16] develop control strategies to address infrequent or unavailable state measurements, [17] investigates lost messages in optimal load scheduling, and [18] investigates the impact of, but does not compensate for, communication latencies.

In this paper, we develop non-disruptive control and estimation algorithms that enable aggregations of residential TCLs to provide ancillary services such as frequency regulation in the presence of significant communication system limitations, including delays, as well as substantial error within the model used by the algorithms. In practice, we would expect large model mismatch since it is difficult to develop a computationally-tractable and accurate model of the aggregate dynamics of large number of spatially-distributed, heterogeneous TCLs, especially given that many useful TCL parameters and states are not easy to measure.

This work builds upon and extends our preliminary work in [19]. The additional contributions of this paper are as follows: 1) we make small modifications to one of the control algorithms first proposed in [19]; 2) we track real PJM frequency regulation signals (rather than simple sinusoidal signals as in [19]) to evaluate the impact of delays on the adequacy of the frequency regulation from demand response; 3) we evaluate the control and estimation algorithms in tandem (rather than individually as in [19]); 4) we evaluate the impact of modeling error by testing the algorithms on two different simulated plants; 5) we find that both estimator-controller combinations can effectively mitigate communication delays; and 6) we find that one estimator is sensitive to the specific model used in the plant whereas the other is not.

The remainder of the paper is organized as follows: Sec-

G. S. Ledva and J. L. Mathieu are with the Department of Electrical Engineering & Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: gsledv@umich.edu; jlmath@umich.edu). E. Vrettos and G. Andersson are with the Power Systems Laboratory, ETH Zurich, 8092 Zurich, Switzerland (e-mail: vrettos@eeh.ee.ethz.ch; andersson@eeh.ee.ethz.ch). S. Mastellone is with the Corporate Research Center, ABB, 5405 Baden-Dättwil, Switzerland (e-mail: silvia.mastellone@ch.abb.com).

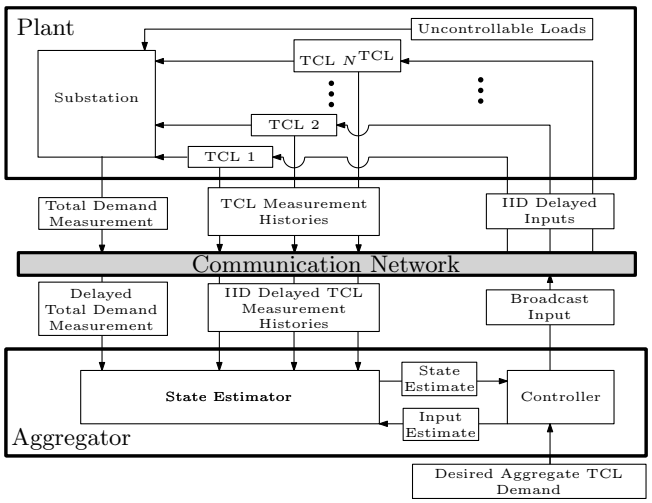


Fig. 1. An overview of the control framework.

tion II describes the problem framework. Section III presents individual and aggregate TCL models. Section IV develops two state estimation algorithms that we use in conjunction with a control algorithm that Section V develops. Section VI formulates a number of case studies and presents their results. Finally, Section VII discusses the conclusions.

II. Problem Framework

As shown in Fig. 1, we assume a problem framework that contains a plant, a communication network, and an aggregator. The plant consists of a set of N^{TCL} controllable TCLs, some uncontrollable loads, and a distribution substation that serves the total demand of the plant. We assume a smart meter acts as an interface between each TCL and the communication network, allowing two-way communication between the TCLs and the aggregator through an imperfect communication network as in [8].

We make several assumptions regarding the communication network and smart meters. Due to the capabilities of digital communication networks, we assume that multiple measurements and inputs can be transmitted within one communication packet, i.e., message. We also assume the messages are time-stamped [13] and the clocks are synchronized across the communication network nodes, allowing knowledge of previously realized delays and their resulting statistics. We assume that the communication network imposes independent and identically distributed (IID) delays on each message. We assume the smart meter can use logic to select an applicable input from a set of inputs, as explained in Section V. Finally, we assume each smart meter can collect histories of the TCL's internal air temperature and on/off mode measurements, but the smart meter can only transmit these state measurement histories infrequently, e.g., every fifteen minutes, due to communication limitations as in [8].

We assume the aggregator, which acts as a central controller, uses a state estimator and controller, both of which include a model of the plant. The aggregator induces TCL on/off switching by broadcasting inputs at each time-step (every two seconds). The inputs are designed to produce a desired aggregate TCL demand and are described in more detail in Section III-B. IID delays cause the inputs to arrive asynchronously, and so an estimated input is used by the aggregator. We assume the desired aggregate demand values are frequency regulation signals provided by the system operator.

The aggregator's state estimation algorithm produces an estimate of the TCL aggregation's state, which is described in Section III-B. We assume that the aggregator has access to measurements of the total substation demand and TCL state measurement histories as in [8]. As in [8], substation demand measurements are available at every time-step, and the aggregate TCL demand is estimated from these measurements by subtracting a prediction of the uncontrollable load. In this work, we add normally-distributed, zero-mean noise to the aggregate TCL demand to approximate the prediction error. As in [8] the TCL measurement histories are available infrequently (every 15 minutes) due to smart meter limitations.

The infrequent availability of possibly delayed TCL state measurements, means the aggregator relies on output feedback (i.e., the aggregate TCL demand estimates, referred to as "aggregate power measurements") at most time-steps to form the state estimate. The state estimate is then used by the aggregator to generate the control inputs.

III. Modeling

We use three previously developed models and describe them here for completeness. Two hybrid, heat transfer-based models represent individual TCLs, and a Markov chain model represents the TCL population.

The first individual TCL model, developed in [20], [21] and referred to as the three-state model, models household heating and cooling appliances using a mass temperature and an air temperature. The second, simpler TCL model, developed in [22], [23] and referred to as the two-state model, uses only an air temperature and can model all TCLs. We use these models to represent the TCLs within the plant, and also incorporate identified versions of the two-state model into the state estimator developed in Section III-A2.

The hybrid nature of these models, i.e., their usage of both discrete and continuous states, makes it computationally challenging to incorporate thousands of the models within optimization-based control algorithms. The Markov chain model, developed in [6] and referred to as the aggregate model, is a linear model of the TCL population's power demand dynamics, and it is easily incorporated into control algorithms. The algorithms in Sections IV and V use this aggregate model.

The following section describes the individual TCL models, and Section III-B describes the aggregate model. Note that we assume model parameters are time-invariant throughout, but the models can incorporate time-varying parameters.

A. Individual TCL Models

We first present a generic discrete-time state update equation for cooling TCLs below, then detail the difference between the two individual models in the following sections. Table I summarizes the model parameters where $[\alpha, \beta]$ corresponds to a uniform distribution and $\mathcal{N}(\alpha, \beta)$ corresponds to a normal distribution with mean α and variance β.

Denote the set of TCLs $\mathcal{J}^{\text{TCL}} = \{1, 2, \ldots, N^{\text{TCL}}\}$. Each TCL $j \in \mathcal{J}^{\text{TCL}}$ has continuous-time matrix parameters $A^{\text{c},j}$,

TABLE I
TCL MODEL PARAMETERS

Parameter	Description	Three-State Value	Two-State Value
θ^{set}	Temperature Set-Point [°C]	[23, 25]	[23, 25]
θ^{db}	Temperature Dead-band [°C]	[0.85, 1.15]	[0.85, 1.15]
θ^{o}	Outdoor Temperature [°C]	32	32
U^{m}	Envelope Conductance [$\frac{\text{kW}}{^\circ\text{C}}$]	[4.35, 5.87]	-
U^{a}	Internal Conductance [$\frac{\text{kW}}{^\circ\text{C}}$]	[0.26, 0.35]	[0.41, 0.56]
Λ^{m}	Mass Heat Capacitance [$\frac{\text{kWh}}{^\circ\text{C}}$]	[1.93, 2.60]	-
Λ^{a}	Air Heat Capacitance [$\frac{\text{kWh}}{^\circ\text{C}}$]	[0.48, 0.64]	[0.51, 0.70]
Q^{m}	TCL Mass Heat Gain [kW]	0	-
Q^{a}	TCL Air Heat Gain [kW]	$\mathcal{N}(0, \epsilon)$	$\mathcal{N}(0, \epsilon)$
ϵ	Variance for Q^{a} [kW^2]	2.5E-7	2.5E-7
Q^{h}	TCL Heat Transfer [kW]	[-16, -12]	[-16, -12]
η	Coefficient of Performance [-]	3	3
Δt	Time-Step Duration [s]	2	2

$B^{\text{c},j}$, and $E^{\text{c},j}$, and the discrete-time matrix parameters A^j, B^j, and E^j are formed using [24, p. 315]. The vector θ_t^j denotes the continuous states, which are the TCL's internal temperature(s) at time-step t. The discrete state corresponds to the scalar on/off mode m_t^j, and d_t^j is a disturbance vector. The discrete-time state-update equations are

$$\theta_{t+1}^j = A^j \theta_t^j + B^j m_t^j + E^j d_t^j \qquad j \in \mathcal{J}^{\text{TCL}} \quad \text{(1a)}$$

$$m_{t+1}^j = \begin{cases} 0 & \text{if } \theta_{t+1}^{a,j} < \theta^{\text{set},j} - \theta^{\text{db},j}/2 \\ 1 & \text{if } \theta_{t+1}^{a,j} > \theta^{\text{set},j} + \theta^{\text{db},j}/2 \\ m_t^j & \text{otherwise,} \end{cases} \qquad j \in \mathcal{J}^{\text{TCL}} \quad \text{(1b)}$$

where (1a) updates the internal temperatures, (1b) updates the on/off mode, and $\theta_t^{a,j}$ is the element of θ_t^j that corresponds to the TCL's air temperature, which is being regulated. The power demand of TCL j is $P_t^j = (|Q^{\text{h},j}|\ m_t^j)/\eta^j$ with $Q^{\text{h},j} < 0$ for cooling TCLs. The output of the TCL model, or the values that can be measured, is $y_t^j = \begin{bmatrix}\theta_t^{a,j} & m_t^j\end{bmatrix}^\top$.

1) Three-State Individual TCL Model: In the three-state model $\theta_t^j = \begin{bmatrix}\theta_t^{a,j} & \theta_t^{m,j}\end{bmatrix}^\top$ where $\theta_t^{m,j}$ denotes the TCL's mass temperature. The disturbance vector is $d_t^j = \begin{bmatrix}\theta^{\text{o}} & Q_t^{\text{a},j} & Q^{\text{m},j}\end{bmatrix}^\top$, where the heat injections $Q_t^{\text{a},j}$ and $Q^{\text{m},j}$ arise due to solar irradiance and heat gain within the household due to occupants and additional appliances. The model's continuous-time matrices are

$$A^{\text{c},j} = \begin{bmatrix} -\left(U^{\text{a},j} + U^{\text{m},j}\right)/\Lambda^{\text{a},j} & U^{\text{m},j}/\Lambda^{\text{a},j} \\ U^{\text{m},j}/\Lambda^{\text{m},j} & -U^{\text{m},j}/\Lambda^{\text{m},j} \end{bmatrix}$$

$$B^{\text{c},j} = \begin{bmatrix} Q^{\text{h},j}/\Lambda^{\text{a},j} & 0 \end{bmatrix}^\top$$

$$E^{\text{c},j} = \begin{bmatrix} U^{\text{a},j}/\Lambda^{\text{a},j} & 1/\Lambda^{\text{a},j} & 0 \\ 0 & 0 & 1/\Lambda^{\text{m},j} \end{bmatrix}.$$

Table I's "Three-State Value" column parameterizes a population of residential air conditioners using nominal parameters from [25]. However, we set the outdoor temperature θ^{o} to simulate a reasonably hot day, we assume $Q_t^{\text{a},j}$ is zero-mean and normally-distributed to include random air temperature disturbances as in [19], and we set $Q^{\text{m},j} = 0$. The results in Section VI-B include a discussion of the algorithms' ability to accommodate positively biased heat injections.

2) Two-State Individual TCL Model: In the two-state model $\theta_t^j = \theta_t^{\text{a},j}$ and $d_t^j = \begin{bmatrix}\theta^{\text{o}} & Q_t^{\text{a}}\end{bmatrix}^\top$. The resulting continuous-time matrices are $A^{\text{c},j} = -U^{\text{a},j}/\Lambda^{\text{a},j}$, $B^{\text{c},j} = Q^{\text{h},j}/\Lambda^{\text{a},j}$ and $E^{\text{c},j} = \begin{bmatrix}U^{\text{a},j}/\Lambda^{\text{a},j} & 1/\Lambda^{\text{a},j}\end{bmatrix}$. We set the parameters in Table I's "Two-State Value" column equal to the three-state model values where applicable, but we set $U^{\text{a},j}$ and $\Lambda^{\text{a},j}$ such that the nominal cycle time is comparable to that of the three-state model.

B. Aggregate TCL Population Model

The aggregate model assumes that the two-state model of Section III-A2 is the underlying individual TCL model for two reasons. First, while an aggregate model exists for the three-state TCL model [5], measurements of $\theta_t^{m,j}$ are not easy to obtain, and practical construction of the state-transition matrix from available measurements is an open question. Second, incorporating the three-state aggregate model into the controller developed in Section V would be computationally prohibitive. In Section VI, we evaluate the impact of this assumption on the control and estimation algorithms' performance by simulating TCLs within the plant using the three-state model.

The aggregate model uses an aggregate state $x_t \in \mathbb{R}^{N^{\text{x}}}$ where N^{x} is the number of discrete states, and each element of the state vector corresponds to the portion of TCLs within the discrete state. The discrete states are formed by first defining a normalized temperature deadband and then dividing it into $N^{\text{x}}/2$ temperature intervals. Each interval contains two states – one for TCLs that are drawing power and one for TCLs that are not. At each time-step, a TCL maps to the aggregate state that contains its normalized internal air temperature and corresponds to its current on/off mode. The state transition matrix $A \in \mathbb{R}^{N^{\text{x}} \times N^{\text{x}}}$ is a transposed Markov Transition Matrix, formed from historical TCL measurements, that describes the probability of TCLs transitioning between states in a time-step.

The input $u_t \in \mathbb{R}^{N^{\text{x}}/2}$ is the portion of TCLs that we want to force from the "on" bin of a temperature interval into the corresponding "off" bin or vice versa. The matrix $B \in \mathbb{R}^{N^{\text{x}}/2 \times N^{\text{x}}}$ ensures the TCLs are forced into the opposite on/off bin within the same temperature interval. Before transmitting input vectors to the TCLs, this input is converted into a switching probability by normalizing each input element with the corresponding state element. To implement these switching probabilities, each TCL first selects the probability corresponding to its current internal temperature and on/off mode. Then, each TCL determines whether it switches by drawing a random number. We assume the local TCL controller disregards switching requests when necessary to maintain the temperature within the normal operating range.

The output of the aggregate model y_t depends on whether both aggregate state and aggregate power measurements are available at time-step t. The set $\mathcal{T}^{\text{S}}$ denotes time-steps where both aggregate state and aggregate power measurements are available, and $y_t \in \mathbb{R}^{N^{\text{x}}+1}$ at these time-steps. Otherwise only aggregate power measurements are available and $y_t \in \mathbb{R}$. The resulting linear system is

$$x_{t+1} = Ax_t + Bu_t + w_t \quad \text{(3a)}$$

$$y_t = \begin{cases} C^{\text{P}} x_t + v_t^{\text{P}} & t \notin \mathcal{T}^{\text{S}} \\ \begin{bmatrix} C^{\text{S}} \\ C^{\text{P}} \end{bmatrix} x_t + \begin{bmatrix} v_t^{\text{P}} \\ v_t^{\text{S}} \end{bmatrix} & t \in \mathcal{T}^{\text{S}}, \end{cases} \quad \text{(3b)}$$

where $w_t \in \mathbb{R}^{N^{\mathrm{x}}}$ is process noise including modeling error, $v_t^{\mathrm{S}} \in \mathbb{R}^{N^{\mathrm{x}}}$ is the aggregate state's measurement noise, $v_t^{\mathrm{P}} \in \mathbb{R}$ is the aggregate power's measurement noise, C^{S} is an $N^{\mathrm{x}} \times N^{\mathrm{x}}$ identity matrix, $C^{\mathrm{P}} = \overline{P}^{\mathrm{on}} N^{\mathrm{TCL}}[0 \cdots 0 \mid 1 \cdots 1]$, and $\overline{P}^{\mathrm{on}}$ is the average power demand of TCLs that are drawing power, which is time-varying but we approximate it as a constant.

IV. State Estimation Algorithms

Our state estimation algorithms use a networked, time-varying Kalman filter from [26] that incorporates aggregate state and power measurements. The networked Kalman filter excludes measurements that have not arrived from the calculations at each time-step using binary indicator variables. As delayed measurements arrive, the networked Kalman filter fully incorporates the new information by updating a history of estimates. Measurement delays are treated deterministically within the estimator since delays associated with measurements that have arrived are known. Whereas [26] does not include inputs within the estimator's dynamic model, we use estimated inputs that are described in Section V-B.

To incorporate delayed measurements, past estimator values must be stored so that they can be updated. The memory requirement can be reduced by excluding measurements with delays longer than a preset threshold; for generality, we do not set a delay threshold in this work. Section IV-A presents the networked Kalman filter, and Section IV-B presents two variations of it, which we apply to our problem.

A. The Networked Kalman Filter

Within this section and Section V, we use the time indexing notation $\psi_{k|t}$ where ψ is an arbitrary value, and t denotes the time of the calculation. In this section, $k \leq t$ indexes a historical horizon of time-steps. The horizon length N_t^{kf} is set at each calculation time t as the number of time-steps of the newly-arrived measurements' largest delay, and the set of time-steps within the historical horizon is $\mathcal{K}_t^{\mathrm{kf}} = \{t - N_t^{\mathrm{kf}}, \ldots, t\}$. The set $\mathcal{K}_t^{\mathrm{kf}}$ includes past time-steps requiring an update to incorporate the newly arrived measurements into the state estimate, and the present time-step for which a new state estimate must be generated.

The binary, scalar variables $\gamma_{k|t}^{\mathrm{S}}$ and $\gamma_{k|t}^{\mathrm{P}}$ indicate whether the aggregate state and aggregate power measurements sampled at time-step k have arrived by time-step t. The indicators are 0 if the corresponding measurement has not arrived by time-step t, and 1 otherwise. The indicator $\gamma_{k|t}^{\mathrm{S}}$ is also set to 0 when the aggregate state is not measured. The Kalman filter observations incorporate the aggregate state measurements y_k^{S} and the aggregate power measurements y_k^{P} using $y_{k|t} = C_{k|t} x + v_{k|t}$ where

$$y_{k|t} = \begin{bmatrix} \gamma_{k|t}^{\mathrm{S}} \, y_k^{\mathrm{S}} \\ \gamma_{k|t}^{\mathrm{P}} \, y_k^{\mathrm{P}} \end{bmatrix}, \; C_{k|t} = \begin{bmatrix} \gamma_{k|t}^{\mathrm{S}} \, C^{\mathrm{S}} \\ \gamma_{k|t}^{\mathrm{P}} \, C^{\mathrm{P}} \end{bmatrix}, \text{ and } v_{k|t} = \begin{bmatrix} \gamma_{k|t}^{\mathrm{S}} \, v_k^{\mathrm{S}} \\ \gamma_{k|t}^{\mathrm{P}} \, v_k^{\mathrm{P}} \end{bmatrix}.$$

The measurement noise $v_{k|t}$ has covariance $V_{k|t}$, which is a block diagonal matrix composed of the aggregate state and aggregate power measurement noise covariances $\gamma_{k|t}^{\mathrm{S}} V^{\mathrm{S}}$ and $\gamma_{k|t}^{\mathrm{P}} V^{\mathrm{P}}$. These covariances are assumed to correspond to zero-mean, normal distributions. The indicator values ensure that measurements that have not arrived have no effect on the state estimate, and the resulting zero components of $y_{k|t}$, $C_{k|t}$, $v_{k|t}$, and $V_{k|t}$ can be removed to reduce the dimension of the computations with no effect on the estimate. The quantities $\widetilde{y}_{k|t}$, $\widetilde{C}_{k|t}$, $\widetilde{v}_{k|t}$, and $\widetilde{V}_{k|t}$ correspond to the reduced matrices.

To perform state estimation, we initialize the recalculation horizon using the state estimate and error covariance from the calculation at time $t-1$: $\widehat{x}_{t-N_t^{\mathrm{kf}}-1|t} = \widehat{x}_{t-N_t^{\mathrm{kf}}-1|t-1}$ and $H_{t-N_t^{\mathrm{kf}}-1|t} = H_{t-N_t^{\mathrm{kf}}-1|t-1}$. Previous state estimates are then recalculated to incorporate newly arrived measurements, and a new state estimate is generated for time-step $k = t$. For $k \in \mathcal{K}_t^{\mathrm{kf}}$:

$$\widehat{x}_{k|t}^{-} = A\widehat{x}_{k-1|t} + B\widehat{u}_{k-1} \tag{4a}$$

$$H_{k|t}^{-} = AH_{k-1|t}A^{\top} + W \tag{4b}$$

$$K_{k|t} = H_{k|t}^{-}\widetilde{C}_{k|t}^{\top}\left(\widetilde{C}_{k|t}H_{k|t}^{-}\widetilde{C}_{k|t}^{\top} + \widetilde{V}_{k|t}\right)^{\dagger} \tag{4c}$$

$$\widehat{x}_{k|t} = \widehat{x}_{k|t}^{-} + K_{k|t}\left(\widetilde{y}_{k|t} - \widetilde{C}_{k|t}\widehat{x}_{k|t}^{-}\right) \tag{4d}$$

$$H_{k|t} = H_{k|t}^{-} - K_{k|t}\widetilde{C}_{k|t}H_{k|t}^{-}. \tag{4e}$$

The aggregate model in (4a) generates an *a priori* state prediction $\widehat{x}_{k|t}^{-}$ and error covariance $H_{k|t}^{-}$ in (4b), where W is the process noise covariance of a zero-mean, normal distribution. Section V explains the estimated input $\widehat{u}_k$ in more detail. The Kalman gain $K_{k|t}$ is calculated in (4c) based on available observation values with $\dagger$ denoting a pseudo-inverse, which we use for numerical reasons. The *a posteriori* state estimate $\widehat{x}_{k|t}$ and error covariance $H_{k|t}$ are then calculated in (4d) and (4e). The output of the algorithm is the new state estimate for the current time-step t, i.e., $\widehat{x}_t = \widehat{x}_{t|t}$, which fully incorporates all measurements that have arrived. Note that the process noise is not necessarily zero-mean and normally-distributed, which results in a sub-optimal filter. We assume that the measurement noise and process noise are independent from each other and independent in time.

B. Variations of the Networked Kalman Filter

Because of IID transmission delays, the state measurement histories from individual TCLs do not arrive synchronously at the aggregator. We develop and investigate the performance of two methods to form aggregate state estimates from asynchronous TCL measurements for use within the networked Kalman filter. The method in Section IV-B1 describes an algorithm that uses N^{TCL} filters from Section IV-A, one for each TCL. The method in Section IV-B2 uses a single networked Kalman filter that incorporates aggregate state predictions generated from identified individual TCL models.

1) Estimator 1: Parallel Kalman Filter Estimator: This state estimator uses a set of N^{TCL} networked Kalman filters. To form the state measurement for each filter, we convert each TCL's most recent air temperature and on/off mode measurement into a vector x_t^j for $j \in \mathcal{J}^{\mathrm{TCL}}$ that is defined equivalently to x_t in Section III-B. The vector x_t^j is 1 in the discrete state that the TCL's temperature and on/off mode maps to and 0 elsewhere. Each filter uses its TCL state measurement and the aggregate power measurement to generate a state estimate $\widehat{x}_t^j$ for each TCL $j \in \mathcal{J}^{\mathrm{TCL}}$. We assume TCL state measurements are accurate, and so we use a near-zero covariance, i.e., $V^{\mathrm{S},j} \approx 0$, for the TCL state measurements within

each Kalman filter. An overall aggregate state estimate is formed as $\hat{x}_t = \left(\sum_{j=1}^{N^{\text{TCL}}} \widehat{x}_t^j\right)/N^{\text{TCL}}$. The main disadvantages of this method are its large computational requirement and the usage of only the most recent TCL measurements.

2) Estimator 2: Single Kalman Filter Using State Predictions: This state estimator consists of three components: a parameter identification algorithm, a bank of N^{TCL} identified two-state individual TCL models, and a single networked Kalman filter. As each TCL's measurement history arrives, they are used within a nonlinear least squares algorithm to identify the thermal parameters $\widehat{\Lambda}^{\text{a},j}$ and $\widehat{U}^{\text{a},j}$ of the TCL, assuming a two-state TCL model and that the set-point, deadband width, and outdoor temperature are known. The set of identified TCL models allow each TCL to be simulated where we assume $Q_t^{\text{a,j}} = 0$ for all j. Aggregate state predictions $x_t^\star$ are formed based on model predictions at each time-step.

A single networked Kalman filter treats the predictions $x_t^\star$ as measurements, allowing the individual TCL models to influence the state estimator. The measurement noise associated with the aggregate state predictions v_t^{S} is assumed to be zero-mean and normally-distributed, and the aggregate state measurement noise covariance V^{S} is generated using the historical errors. Note that the noise will not be normally-distributed in general, which results in a sub-optimal filter. The method has the disadvantage that it relies on the accuracy of the two-state model, which is discussed in Section VI-B.

V. Control Algorithm

The aggregator uses the predictive control approach described in [13] to counteract input delays. In this approach, the control algorithm generates an open-loop input sequence $U_t \in \mathbb{R}^{N^{\text{x}}/2 \times N^{\text{mpc}}}$ at each time-step based on the current state estimate. IID delays cause asynchronous input arrival at the TCLs. Using time-stamping, we assume that the smart meter (or TCL) can select the most recently generated input sequence that has arrived. The TCL then selects the input from that sequence that applies to the current time-step, or it uses a zero input. The controller does not know the implemented input at each TCL, and an estimated input is generated from the known delay statistics and is used within the aggregator's algorithms.

We develop a model predictive control (MPC) algorithm that considers a horizon of N^{mpc} time-steps ranging from the present time-step t to future time-step $t+N^{\text{mpc}}-1$ to generate U_t. The MPC algorithm uses the aggregate model to design inputs to track the desired aggregate demand $y_t^{P,\text{ref}}$.

Within this section, k is used to indicate the time-step of the MPC horizon. Inputs corresponding to time-step k are produced at time $t = k - N^{\text{mpc}} + 1, \ldots, k$, resulting in a total of N^{mpc} inputs for each time-step. The matrix $\mathcal{U}_k = \begin{bmatrix} u_{k|k} & \ldots & u_{k|k-N^{\text{mpc}}+1} \end{bmatrix}$ denotes the set of input vectors that apply to time-step k.

As in [27], we form an input estimate $\widehat{u}_t$ based on previously transmitted input sequences. Ref. [27] attempts to estimate the single input within an actuator whereas we form the input estimate $\widehat{u}_t$ as the weighted sum of possible inputs and their probability of being implemented. The probabilities, which are the elements of the vector $\mathcal{P}$, are based on known input delay statistics and the TCL input selection logic. Section V-B details the construction of $\widehat{u}_t$, $\mathcal{U}_k$, and $\mathcal{P}$. The following section presents the MPC formulation, which is a finite-horizon, linear quadratic output regulator with input and state constraints that is implemented using [28].

A. MPC Formulation

To set N^{mpc}, we first fix a parameter $p^{\max}$. The value $1 - p^{\max}$ is the probability that no valid input is available at the TCL, and we explain this further in Section V-B. The MPC algorithm is initialized using the current state estimate $x_t = \widehat{x}_t$, the current aggregate demand request $y_t^{P,\text{ref}}$, and any previously transmitted inputs that apply to time-steps within the horizon. The full formulation is

$$\underset{u,\delta}{\text{minimize}} \sum_{k=t}^{t+N^{\text{mpc}}-1} \Big[c^y \left(y_k^{\text{err}}\right)^2 + c^\delta (\delta_k^- + \delta_k^+) + \sum_{m=k-N^{\text{mpc}}+1}^{k} c^u (u_{k|m}^\top \; u_{k|m}) \Big] \quad (5)$$

$$\text{s.t.} \quad x_{k+1} = A\, x_k + B\, \widehat{u}_k \quad (6)$$

$$\widehat{u}_k = \mathcal{U}_k \mathcal{P} \quad (7)$$

$$y_k^{\text{err}} = y_k^{P,\text{ref}} - C^{\text{P}} x_k \quad (8)$$

$$u_{k|m}^i \le x_k^i \qquad i \in \{1, \ldots, N^{\text{x}}/2\} \quad (9)$$

$$-u_{k|m}^i \le x_k^{N^{\text{x}}+1-i} \qquad i \in \{1, \ldots, N^{\text{x}}/2\} \quad (10)$$

$$0 - \delta_k^- \le x_k \le 1 + \delta_k^+ \quad (11)$$

$$0 \le \delta_k^-, \delta_k^+. \quad (12)$$

The objective function (5) minimizes the total cost over the horizon where the costs c^y, c^u, and c^δ penalize the tracking error y_k^{err}, control effort, and soft constraint violations δ_k^+ and δ_k^-. The dynamic model in (6) corresponds to (3a) excluding the process noise and using the estimated input from (7). The tracking error is calculated in (8) using a persistent value of the current aggregate demand request, i.e., $y_k^{P,\text{ref}} = y_t^{P,\text{ref}}$ for all k in the MPC horizon. The input constraints (9) and (10) limit the fraction of TCLs to switch from a particular bin to be less than the fraction of TCLs within that bin. The soft state constraint (11) is satisfied regardless of the initial value provided from the unconstrained state estimator, and (12) restricts the soft constraint violations to positive values.

B. Constructing Input Estimates

The vector $\mathcal{P}$ weights each of the inputs in $\mathcal{U}_k$ based on their probability of being implemented. The probabilities within $\mathcal{P}$ are fixed during an MPC calculation. However, the probabilities could be recomputed between MPC calculations if the delay distribution changes, e.g., due to different traffic levels at various times of the day.

The inputs in $\mathcal{U}_k$ become "older" as we go from left to right in the matrix, and we calculate the elements of $\mathcal{P}$ using the corresponding input vector's location in $\mathcal{U}_k$. Due to its input selection logic, a TCL uses the input corresponding to the leftmost column of $\mathcal{U}_k$ that has arrived. The probability of using a column depends on two events: 1) the column must have arrived, and 2) every column to its left within $\mathcal{U}_k$ must not have arrived. We generate the elements of $\mathcal{P}$ based on these two events, the necessary delays for these events to occur,

TABLE II
SIMULATION PARAMETERS

Parameter	Description	Value
N^{x}	Number of State Bins [-]	100
N^{TCL}	Number of TCLs [-]	10,000
R^{P}	Aggregate Power Noise Covariance [kW^2]	$\mathcal{N}(0, 4\mathrm{E}6)$
P^{avg}	Average Steady-State TCL Demand [kW]	6E3
Δt	Time-Step Duration [s]	2
$\Delta t^{S,s}$	TCL State Measurement Interval [s]	2
$\Delta t^{S,t}$	TCL State History Transmission Interval [s]	900
n^{steps}	Time-Steps in Simulation [-]	1800
$p^{\max}$	MPC Delay Probability Threshold [-]	0.999
c^{y}	MPC Output Cost [-]	1
c^{u}	MPC Input Cost [-]	1
c^{δ}	MPC Soft Constraint Cost [-]	1.01

and the probability of realizing these delays. Note that while assuming IID delays simplifies the following calculations, they are still possible without independence.

To simplify notation, denote the ith element of $\mathcal{P}$ as p^i, the corresponding column of $\mathcal{U}_k$ as u^i, the delay associated with the arrival of column u^i as τ^i, and the probability of the first and second necessary events as $p^{1,i}$ and $p^{2,i}$. Using the assumption of IID delays, $p^i = p^{1,i}p^{2,i}$. The input u^i arrives by time-step k if its delay is less than i

$$p^{1,i} = p(\tau^i < i) \qquad i = 1, \ldots, N^{\mathrm{mpc}}. \tag{13}$$

Assuming IID delays, the probability that all columns left of column i have not arrived by time-step k is

$$p^{2,i} = \prod_{n=1}^{i-1} p(\tau^n > i - n) \qquad i = 1, \ldots, N^{\mathrm{mpc}}. \tag{14}$$

The first column is used if its delay is less than one. The second column requires that its delay is less than two and the first column's delay is greater than one, and so on.

For an MPC calculation, columns within $\mathcal{U}_k$ whose right time index is less than t correspond to previously transmitted inputs, and they are set to the transmitted values. The remaining columns are free decision variables, but only a portion are transmitted in the input sequence. Specifically, the sequence U_t consists of the one column from each $\mathcal{U}_k$, $k \in \{t, \ldots, t + N^{\mathrm{mpc}} - 1\}$, whose right-hand time index corresponds to the current time t, i.e., every $u_{a|b}$ with $b = t$. The horizon length N^{mpc} is set such that the sum of elements in $\mathcal{P}$ is greater than $p^{\max}$ where N^{mpc} is the length of $\mathcal{P}$.

VI. CASE STUDIES

We summarize a series of simulations investigating 1) the impact of compensating for delays, and 2) the ability of the methods to provide frequency regulation despite communication delays and model error. Section VI-A details the simulation parameters, algorithm combinations, delay distributions, reference signal construction, and quantities used to evaluate the simulations. Section VI-B presents the simulation results.

A. Case Study Setup

Table II details the simulation parameters. We simulate a TCL population of 10,000 air conditioners. The average steady-state TCL demand is slightly different between the two- and three-state TCL model populations because of the parameters used, and the value in the table is approximate. We use a zero-mean, normal distribution to generate the aggregate power measurement noise. Similar to [6], [19], we set the noise variance assuming the average steady-state TCL demand is 15% of the demand served by the substation, and the standard deviation of the power measurement noise is set to 5% of the total substation load. We generate the aggregate model's process noise covariance using historical errors. The resulting process noise is neither zero-mean nor normally-distributed.

We conduct a series of case studies, varying: 1) the average delay, 2) the reference signal, 3) the model used to simulate individual TCLs within the plant, and 4) the estimator. The two reference signals, called the Reg-A and Reg-D reference, correspond to segments of published PJM traditional and dynamic frequency regulation signals from [29]. The reference signals are from May 4, 2014. The signals are interpolated to two second time-steps, and each signal is scaled so that the maximum demand change request corresponds to $\pm 20\%$ of the average steady-state aggregate TCL demand. Three delay distributions are used, referred by their average delay: 0, 10, and 20 seconds. With average delays of 0, we do not impose any measurement or input delays. With average delays of 10 and 20 seconds, IID delays τ are sampled from a discretized log-normal distribution $\tau = \lfloor \exp(\tau^\star) \rfloor$ where $\lfloor \cdot \rfloor$ rounds down and $\tau^\star$ is normally-distributed. In the cases with delays, the variance of the log-normal distribution is 0.25 sec^2. We use the chosen distribution to model delays that do not take negative values and are unbounded above. Recall that the estimator treats the delays deterministically after each measurement has arrived, and so the delay distribution does not impact the estimator performance. The controller uses the statistics of the delay distribution, and so alternative distributions would change the entries in $\mathcal{P}$, but these can be computed regardless of the distribution.

We compare four control setups referred to as Estimator 1-NC, Estimator 1-TS, Estimator 1-FC, and Estimator 2-FC. These setups vary both the complexity of delay compensation method and the estimator used. Estimators 1-FC and 2-FC (FC = full compensation) pair the controller from Section V with Estimator 1 and 2, respectively. Estimator 1-TS (TS = time stamping) pairs Estimator 1 with a simplified controller that does not include delay information, i.e., $N^{\mathrm{mpc}} = 1$ and $\mathcal{P} = 1$. Estimator 1-TS uses time-stamping where delays are compensated for in the estimator and TCLs use the most recently generated input that has arrived. Estimator 1-NC (NC = no compensation) pairs Estimator 1 with the controller, but it does not use time-stamping or selection logic at the TCLs. In this setup, the delays are not compensated for in Estimator 1, and TCLs use whichever input arrives first during a time-step or a zero input vector if no input arrived.

We quantify the results using two values 1) the normalized RMS tracking error (RMSE), and 2) the PJM score described below. We run ten instances of each case with different realizations of the random quantities and average the RMSE and PJM score across the instances for each case. The RMSE for a single case instance is

$$P^{\mathrm{RMSE}} = \frac{1}{P^{\mathrm{avg}}} \sqrt{\frac{1}{n^{\mathrm{steps}}} \sum_{t \in \mathcal{T}} \left(y_t^{P,\mathrm{real}} - y_t^{P,\mathrm{ref}} \right)^2} \tag{15}$$

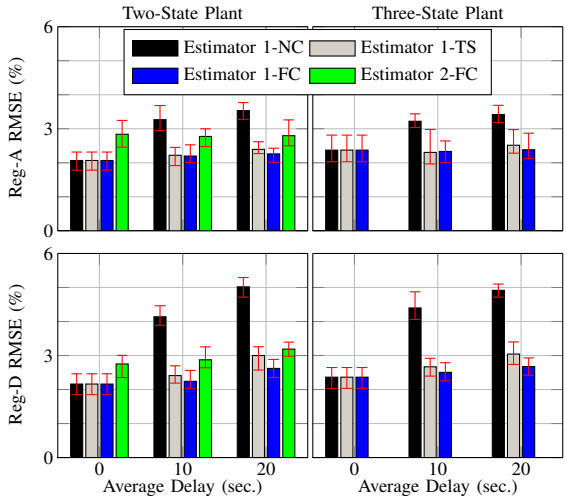
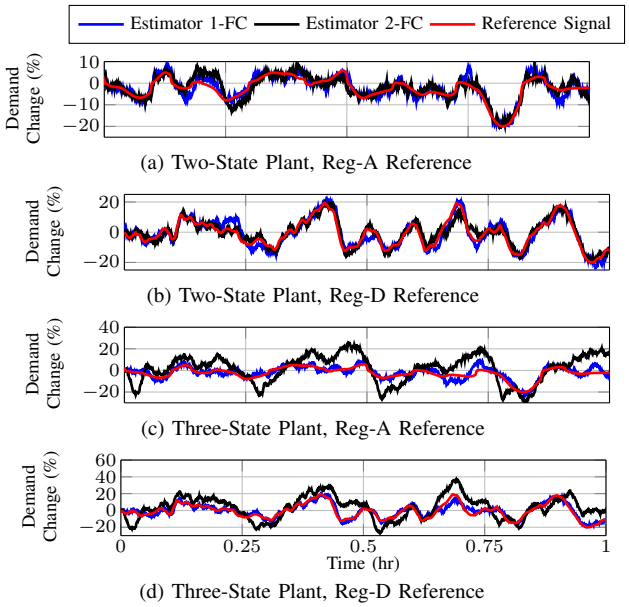


Fig. 2. The average RMSE across all simulation scenarios varying the plant (two-state and three-state TCL models), control setup, average delay, and reference (Reg-A and Reg-D). Error bars indicate the range of values achieved across the ten instances of each scenario.



(a) Two-State Plant, Reg-A Reference

(b) Two-State Plant, Reg-D Reference

(c) Three-State Plant, Reg-A Reference

(d) Three-State Plant, Reg-D Reference

Fig. 3. Time-series plots comparing the reference tracking of Estimator 1 and Estimator 2 with average delays of 20 seconds.

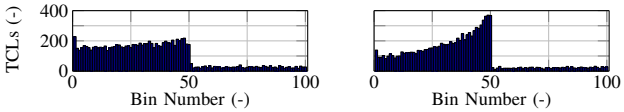
where P^{avg} is the average steady-state aggregate TCL demand and $y_t^{P,\text{real}}$ is the achieved aggregate demand. The PJM score is a value between 0 and 1 that is calculated using PJM's standards [30, pp. 52-54]. The score uses the correlation, delay, and difference in energy between the requested and actual signals. A passing score is ≥ 0.75, which could certify a resource to provide frequency regulation when tracking the PJM test signal. A score of ≥ 0.50 would be satisfactory to maintain certification for an arbitrary reference signal.

B. Simulation Results

Figure 2 summarizes the average RMSE for the four control setups, Table III summarizes the average PJM scores, and Fig. 3 provides time series for Estimators 1-FC and 2-FC assuming an average delay of 20 seconds. Note that Fig. 2 does not contain results for cases using Estimator 2-FC and three-state models as the plant. As shown in Fig. 3c and 3d,

TABLE III
AVERAGE PJM SCORES [-]

Control Setup	Avg. Delay	Two-State Plant Reg-A	Two-State Plant Reg-D	Three-State Plant Reg-A	Three-State Plant Reg-D
Estimator 1-NC	0	0.852	0.899	0.810	0.894
	10	0.727	0.821	0.735	0.835
	20	0.707	0.782	0.727	0.805
Estimator 1-TS	0	0.852	0.899	0.810	0.894
	10	0.846	0.908	0.821	0.901
	20	0.836	0.892	0.801	0.896
Estimator 1-FC	0	0.852	0.899	0.810	0.894
	10	0.849	0.908	0.814	0.901
	20	0.849	0.903	0.824	0.903
Estimator 2-FC	0	0.785	0.882	——	——
	10	0.799	0.892	——	——
	20	0.807	0.887	——	——



(a) Two-State Plant (b) Three-State Plant

Fig. 4. Comparison of discrete state bin distributions in steady-state for populations of TCLs represented by two-state and three-state models.

the resulting RMSE is high ($\sim 15-20\%$) in these cases, and we explain the cause below.

Note that with no delay, the RMSE and PJM scores of Estimators 1-NC, 1-TS, and 1-FC are all identical in Fig. 2 and Table III since all three methods are equivalent without delay. However, with increasing average delays, the RMSE increases significantly for Estimator 1-NC, increases slowly for Estimator 1-TS, and is roughly constant for Estimator 1-FC. This indicates that i) delay compensation is needed if delays are significant, ii) a simple compensation method reduces the effects of delays, and iii) more complex methods can further mitigate the effects of delays. Estimator 2-FC generally performs worse than Estimator 1-TS, and we discuss this below. Note that all methods produce average PJM scores over the 0.75 threshold, except Estimator 1-NC when used to provide Reg-A with average delays of 10 and 20 seconds.

Focusing on Estimator 1-FC, PJM scores for the Reg-D reference are slightly better than those for the Reg-A reference, while the RMSE is slightly worse for the Reg-D signal. Because the trends are different for each performance metric, it is unclear whether the Reg-A or Reg-D cases are superior. Using the three-state TCL models within the plant resulted in slightly worse RMSE and PJM scores, but they are still above PJM's threshold, which is important since three-state models capture the TCL dynamics more accurately. While we used zero-mean heat injections to generate all of the results shown in this paper, simulations results (not shown here due to space limitations) indicate that Estimator 1-FC can adequately handle more realistic, biased heat injections. In contrast, Estimator 2-FC is unable to account for biased heat injections.

Estimator 2-FC's performance is dependent on the TCL model used within the plant. When two-state models are used within the plant, Estimator 2-FC's PJM scores are acceptable in all scenarios; however, it performs worse than Estimators 1-TS and 1-FC. Estimator 2-FC's performance is dependent

on the number of discrete states N^{x}, and diminishes if N^{x} is reduced, e.g., to 40, for computational reasons. Improvements to Estimator 2-FC may be achievable with more advanced parameter estimation methods.

When the three-state models are used within the plant, Estimator 2-FC is not able to provide effective frequency regulation because Estimator 2 uses two-state models to generate aggregate state predictions, which are treated as measurements within the state estimator. Figure 4 shows the steady-state distribution of TCLs within 100 discrete state bins using the two- and three-state plants. The distribution in the two-state plant (left) is fairly flat across the first 50 bins, which correspond to TCLs that are off. Alternatively, the distribution in the three-state plant (right) has a large concentration of TCLs around bin 50, which corresponds to the edge of the dead-band. This qualitative difference in their distributions means that the two-state models used in Estimator 2 cannot effectively predict the aggregate state when the plant consists of three-state models. Because the thermal mass temperature and environmental heat injections are difficult to measure, identifying the three-state model is difficult.

VII. Conclusions

In this paper, we developed a predictive controller and two estimators to mitigate the effects of communication delays within a residential demand response scenario. In simulations, we investigated the ability of the algorithms to control an aggregation of TCLs to track frequency regulation signals. Results show that both estimator-controller combinations are able to effectively provide frequency regulation with average delays of up to 20 seconds. The first estimator, which includes only an aggregate model and relies on a number of Kalman filters running in parallel is effective with both structural- and parameter-based modeling error. The second estimator assumes a specific TCL model and identifies parameters for those models. It is effective if the assumed TCL model structure matches the true model structure.

Future work should incorporate and address time-varying outdoor temperatures and time-varying, biased heat injections into the individual TCL models. Investigating a more effective parameter identification algorithm may allow Estimator 2 to be used in a wider range of scenarios. Modifying the controller to account for modeling error may improve tracking performance. Finally, accounting for non-normally-distributed measurement noise should be investigated.

References

[1] G. Strbac, "Demand side management: Benefits and challenges," *Energy Policy*, vol. 36, no. 12, pp. 4419–4426, 2008.

[2] P. Siano, "Demand response and smart grids – A survey," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, 2014.

[3] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Conversion and Management*, vol. 50, no. 5, pp. 1389–1400, 2009.

[4] S. Bashash and H. Fathy, "Modeling and control of aggregate air conditioning loads for robust renewable power management," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 4, pp. 1318–1327, 2013.

[5] W. Zhang, J. Lian, C.-Y. Chang, and K. Kalsi, "Aggregated modeling and control of air conditioning loads for demand response," *IEEE Transactions Power Systems*, vol. 28, no. 4, pp. 4655–4664, 2013.

[6] J. Mathieu, S. Koch, and D. Callaway, "State estimation and control of electric loads to manage real-time energy imbalance," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 430–440, 2013.

[7] T. Borsche, F. Oldewurtel, and G. Andersson, "Minimizing communication cost for demand response using state estimation," in *IEEE PowerTech – Grenoble*, 2013.

[8] E. Vrettos, J. L. Mathieu, and G. Andersson, "Control of thermostatic loads using moving horizon estimation of individual load states," in *Power Systems Computation Conference*, 2014.

[9] D. S. Callaway and I. A. Hiskens, "Achieving controllability of electric loads," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 184–199, 2011.

[10] SCE, "Southern California Edison company 2012-2014 demand response program portfolio," Southern California Edison, Tech. Rep. Application No. A.11-03-003 filed before the Public Utilities Commission of California, March 2011.

[11] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, vol. 52, pp. 213–234, 2013.

[12] J. H. Eto, J. Nelson-Hoffman, E. Parker, C. Bernier, P. Young, D. Sheehan, J. Kueck, and B. Kirby, "The demand response spinning reserve demonstration–measuring the speed and magnitude of aggregated demand response," in *45th Hawaii International Conference on System Science*, 2012.

[13] L. Zhang, H. Gao, and O. Kaynak, "Network-induced constraints in networked control systems – A survey," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 403–416, 2013.

[14] K. Wada and A. Yokoyama, "Load frequency control using distributed batteries on the demand side with communication characteristics," in *IEEE PES Conference on Innovative Smart Grid Technologies*, Oct 2012.

[15] A. Ghaffari, S. Moura, and M. Krstic, "Modeling, control, and stability analysis of heterogeneous thermostatically controlled load populations using partial differential equations," *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 137, no. 10, 2015.

[16] J. H. Braslavsky, C. Perfumo, and J. K. Ward, "Model-based feedback control of distributed air-conditioning loads for fast demand-side ancillary services," in *IEEE Conference on Decision and Control*, 2013, pp. 6274–6279.

[17] N. Gatsis and G. B. Giannakis, "Residential load control: Distributed scheduling and convergence with lost AMI messages," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 770–786, 2012.

[18] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Frequency regulation from flexible loads: Potential, economics, and implementation," in *American Control Conference*, 2014, pp. 65–72.

[19] G. S. Ledva, E. Vrettos, S. Mastellone, G. Andersson, and J. L. Mathieu, "Applying networked estimation and control algorithms to address communication bandwidth limitations and latencies in demand response," in *Hawaii International Conference on System Sciences*, 2015, pp. 2645–2654.

[20] R. C. Sonderegger, "Dynamic models of house heating based on equivalent thermal parameters," Ph.D. dissertation, Princeton Univ., NJ., 1978.

[21] N. Wilson, B. Wagner, and W. Colborne, "Equivalent thermal parameters for an occupied gas-heated house," *ASHRAE Transactons*, vol. 91, no. CONF-850606, 1985.

[22] S. Ihara and F. C. Schweppe, "Physically based modeling of cold load pickup," *IEEE Transactions on Power Apparatus and Systems*, no. 9, pp. 4142–4150, 1981.

[23] C.-Y. Chong and A. Debs, "Statistical synthesis of power system functional load models," in *IEEE Conference on Decision and Control*, 1979, pp. 264–269.

[24] K. Ogata, *Discrete-time control systems*. Prentice Hall Englewood Cliffs, NJ, 1995, vol. 2.

[25] (2015) Gridlab-D House Class Documentation. Online. [Online]. Available: http://gridlab-d.sourceforge.net/wiki/index.php/House

[26] L. Schenato, "Optimal sensor fusion for distributed sensors subject to random delay and packet loss," in *IEEE Conference on Decision and Control*, 2007, pp. 1547–1552.

[27] J. Fischer, A. Hekler, and U. D. Hanebeck, "State estimation in networked control systems," in *15th International Conference on Information Fusion (FUSION)*, 2012, pp. 1947–1954.

[28] J. Löfberg, "Yalmip : A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. [Online]. Available: http://users.isy.liu.se/johanl/yalmip

[29] PJM. (2015) Ancillary Services. reg-data-external-may-2014.xls. [Online]. Available: https://www.pjm.com/markets-and-operations/ancillary-services.aspx

[30] *PJM Manual 12: Balancing Operations*, PJM, 2014, revision 31.