Control of Thermostatic Loads Using Moving Horizon Estimation of Individual Load States

Evangelos Vrettos
Power Systems Laboratory
ETH Zurich, Switzerland
vrettos@eeh.ee.ethz.ch

Johanna L. Mathieu
Electrical and Computer Engineering
University of Michigan, USA
jlmath@umich.edu

Göran Andersson
Power Systems Laboratory
ETH Zurich, Switzerland
andersson@eeh.ee.ethz.ch

***Abstract*—Recent research has shown that thermostatically controlled loads (TCLs) can provide power system services. However, a key challenge is to achieve coordinated control of large populations of resources using existing communication and control infrastructure or with minimal addition of new infrastructure. In this paper, we assume that we only have access to realistic measurements, i.e. data from residential smart meters every 15 minutes and noisy real-time measurements of the aggregate power consumption of TCLs from distribution substations. Our contribution is to develop a moving horizon state estimator (MHSE) to estimate the states of individual stochastic TCLs from these measurements. This is in contrast to previous work that focused on estimating the states of aggregate system models. The proposed MHSE is benchmarked against a simpler model-based prediction. We also propose a scalable closed-loop control structure that uses the MHSE method to provide frequency control with TCL populations. We demonstrate our results via a number of case studies with different TCL aggregations, process and measurement noise characteristics, and controller forcing levels. Our simulations show that the MHSE generally provides accurate state estimates and improves the controller performance.**

***Index Terms*—Thermostatically controlled loads, state estimation, demand response, frequency control, hybrid systems**

I. Introduction

The increasing penetration of fluctuating renewable energy sources in the electricity grid calls for more ancillary services, such as frequency control [1]. Although these services traditionally come from conventional generators, there is a recent interest in using the flexibility of loads for the same purpose [2]. The active participation of loads in power systems is expected to reduce the cost of ancillary services, increase their efficiency, and/or reduce their environmental footprint.

Recent research has focused on the use of aggregations of thermostatically controlled loads (TCLs), such as space heaters, air conditioners, refrigerators, and electric water heaters, to provide power system services [2]. TCLs are well-suited to this task because they are hybrid systems that operate with hysteresis controllers and so they have inherent operational flexibility. This means that their power consumption can be shifted in time, within limits, without consumers noticing. Additionally, their dynamics are relatively simple and they are easy to control. However, coordinating the operation of a large population of TCLs is challenging, and may require new control algorithms as well as significant investments in sensing, communication, and control infrastructure.

While some of the literature on direct load control with TCLs uses state-free models, e.g., input-output models [3], most uses dynamical system models and assumes access to state measurements, e.g., [4]–[6]. Recent research has focused on state estimation to reduce the implementation costs of direct load control schemes. For example, [7], [8] proposed a modeling method for TCL aggregations based on Markov chains, and applied a Kalman Filter for state estimation using partial state or aggregate power measurements. Ref. [9] used a similar modeling approach and proposed a moving horizon state estimation (MHSE) method to estimate aggregate model states, but without considering measurement noise. Ref. [10] proposed a four state aggregate system model, similar to that in [11], and a particle filter to estimate aggregate TCL states. A state estimation scheme for a model of aggregated TCLs based on partial differential equations was proposed in [12]. However, these approaches rely on aggregate load models, and estimate the distribution of TCLs in a normalized temperature state space, which do not perfectly capture the states of individual loads.

In contrast to the above work, we attempt to estimate *individual TCL states*, i.e. temperatures and on/off modes. This is expected to improve control performance because it will give the controller more information about the effect of both local and external control actions on the aggregate power consumption. Stochastic hybrid system state estimation problems are often solved with 'multiple model' estimation schemes that involve a filter for each mode [13]. However, with large numbers of modes, as in our problem, these approaches are intractable [14]. Therefore, we transform our hybrid system into a mixed logical dynamical (MLD) system [15] and propose a novel MHSE method that can be cast as a mixed-integer linear program (MILP). Our method can be used together with a broadcast controller to enable provision of fast time scale services with high accuracy requirements, for example, load frequency control (LFC), with minimal investment in new infrastructure. This paper extends our previous work [16], which focused exclusively on the effect of process noise. In this paper, we investigate the realistic case that includes both process and measurement noise.

The contributions of this work are threefold. First, we develop a MHSE method that allows us to estimate individual TCL load states with realistic data in the presence of both process and measurement noise. Second, we benchmark the MHSE method against a simpler model-based prediction approach, which gives us insight into the effect of noise and forcing levels on estimation performance. Third, we propose a scalable hierarchical closed-loop control structure for LFC provision with TCL aggregations. We report simulation results considering a realistic setup with real-time aggregate power

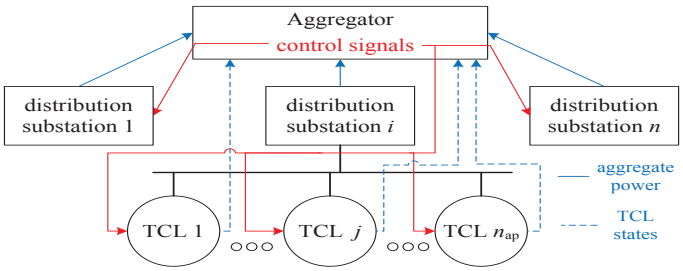


Fig. 1. The proposed control and communication architecture.

measurements from several distribution substations, and low-frequency data packages with TCL state measurements.

Section II describes the control and communication architecture, Section III introduces the MLD model and the control scheme, and Section IV presents the MHSE method. Section V explains our investigations, and simulation results are reported in Sections VI and VII, while Section VIII concludes.

II. Problem Description

Consider an aggregator controlling a TCL population to provide power system services. Ideally the aggregator would have access to high frequency state measurements from each individual TCL, but most smart meters have data transmission limitations. Therefore, we assume that residential smart meters are only able to transmit data to the aggregator every 15 minutes, though they are able to collect and store TCL state measurements every 10 seconds [17]. We also assume that the aggregator has access to aggregate power measurements from distribution substations every 10 seconds, which is a reasonable sample rate for frequency control provision. Using these measurements together with predictions of the uncontrolled load at each substation, we can estimate the aggregate power consumption of the controlled TCLs. Therefore, within each 15 minute interval we must estimate the TCLs' states using noisy aggregate power measurements along with high frequency state measurements collected in the previous interval. The benefit of this approach is that it requires little/no new infrastructure – the only things required are smart meters at each household, a method of measuring and communicating thermostat states (for example, via a simple home energy management system), and power meters at distribution, i.e. medium-low voltage, substations.

The proposed control and communication architecture is shown in Fig. 1. Solid lines represent high frequency communication flows (every 10 seconds), whereas dashed lines show low frequency communication flows (every 15 minutes). The blue lines represent the measurements, whereas the red lines the control signals. To keep the estimation problem tractable, we propose a hierarchical architecture where the estimation is performed independently for each substation. However, the control is performed centrally and signals are broadcast to all TCLs based on the individual load state estimates.

III. Modeling & Control

A. TCL Modeling

We use the two-state hybrid TCL model developed in [18], [19] to model individual TCLs. Denote the TCL temperature and the on/off mode at time step t by $x_{\text{c},t} \in \mathbb{R}$ and $x_{\text{l},t} \in \{0, 1\}$, respectively. A heating TCL's stochastic discrete-time dynamics can be expressed as

$$x_{\text{c},t+1} = a x_{\text{c},t} + b x_{\text{l},t} + f T_{\alpha,t} + w_t \,, \tag{1}$$

$$x_{\text{l},t+1} = \begin{cases} 0 & \text{if } x_{\text{c},t+1} \geq M \\ 1 & \text{if } x_{\text{c},t+1} \leq m \\ x_{\text{l},t} & \text{otherwise} \end{cases} , \tag{2}$$

where $a = e^{-\Delta t/(CR)}$, $b = (1-a)RC_{\text{p}}P_{\text{n}}$, $f = (1-a)$, Δt is the discretization time step, C is the thermal capacitance, R is the thermal resistance, C_{p} is the coefficient of performance, P_{n} is the rated power, $T_{\alpha,t}$ is the ambient temperature, and w_t is the process noise, which includes plant-model mismatch and errors in predictions of external parameters such as $T_{\alpha,t}$ and consumer behavior. Additionally, $M = T_{\text{sp}} + 0.5T_{\text{db}}$ and $m = T_{\text{sp}} - 0.5T_{\text{db}}$ are the upper and lower temperature dead-band limits, respectively, where T_{sp} is the thermostat temperature set-point and T_{db} is the dead-band width.

By introducing the auxiliary binary variables $\delta_{1,t}$, $\delta_{2,t}$, $\delta_{3,t}$, and $\delta_{4,t}$, the stochastic hybrid system can be represented as an MLD system [15]

$$x_{t+1} = \underbrace{\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}}_{A} x_t + \underbrace{\begin{pmatrix} 0 & 0 & 0 & b \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{B} \delta_t + \underbrace{\begin{pmatrix} f \\ 0 \end{pmatrix} T_{\alpha,t}}_{F_t} + \begin{pmatrix} w_t \\ 0 \end{pmatrix} , \tag{3}$$

$$\tilde{E}_1 \delta_t \leq \tilde{E}_2 x_t + \tilde{E}_3 \,, \tag{4}$$

where $x_t := [x_{\text{c},t} \;\; x_{\text{l},t}]^T$, $\delta_t := [\delta_{1,t} \;\; \delta_{2,t} \;\; \delta_{3,t} \;\; \delta_{4,t}]^T$, $[\delta_{1,t} = 1] \leftrightarrow [x_{\text{c},t} \geq M]$, $[\delta_{2,t} = 1] \leftrightarrow [x_{\text{c},t} \leq m]$, $\delta_{3,t} = x_{\text{l},t+1}$, and $\delta_{4,t} = x_{\text{l},t}$. The mixed-integer linear inequalities (4) represent the internal hysteresis controller of the TCL. Note that these equations are similar to (11a)-(11b) in [15], except we include two additional terms – ambient temperature and process noise – and we do not allow the power consumption to vary continuously. $\tilde{E}_1$, $\tilde{E}_2$ and $\tilde{E}_3$ are defined based on E_1, E_2, E_3, E_4, and E_5 from [15], [16] as follows: $\tilde{E}_2 = E_4$; $\tilde{E}_3 = E_1 + E_5$; $\tilde{E}_1(:,k) = E_2(:,k)$ for $k = \{1, 2, 3\}$, and $\tilde{E}_1(:,4) = E_2(:,4) + E_3$.

A heterogeneous aggregation of n_{ap} TCLs can be modeled by stacking together models of individual TCLs, leading to the following state-space representation

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{\delta}_t + \boldsymbol{F}_t + \boldsymbol{w}_t \,, \tag{5}$$

$$\boldsymbol{E}_1 \boldsymbol{\delta}_t \leq \boldsymbol{E}_2 \boldsymbol{x}_t + \boldsymbol{E}_3 \,, \tag{6}$$

where $\boldsymbol{x}_t$, $\boldsymbol{\delta}_t$, $\boldsymbol{w}_t$, $\boldsymbol{E}_3$ are stacked vectors, e.g., $\boldsymbol{x}_t = [(x_t^1)^T \ldots (x_t^{n_{\text{ap}}})^T]^T$, and $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{F}_t$, $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ are block diagonal matrices with the matrices of the individual TCLs on the diagonals.

The output depends upon if full state information or only noisy aggregate power measurements are available

$$y_t = \begin{cases} \boldsymbol{C}_1 \boldsymbol{x}_t \,, & \text{if } t = jT_m \,,\; j \in \mathbb{N} \\ \boldsymbol{C}_2 \boldsymbol{x}_t + v_t \,, & \text{otherwise} \end{cases} , \tag{7}$$

where $\boldsymbol{C}_1 = I$; $\boldsymbol{C}_2 = [0 \; P_{\text{n}}^1 \; 0 \; P_{\text{n}}^2 \ldots 0 \; P_{\text{n}}^{n_{\text{ap}}}]$, v_t is measurement noise; and T_m is the period of TCL-level measurements.

B. Control Design

The controller broadcasts signals to the TCLs, which attempt to track a power trajectory. We extend the closed-loop rule-based control algorithm in [6], [16] to handle the

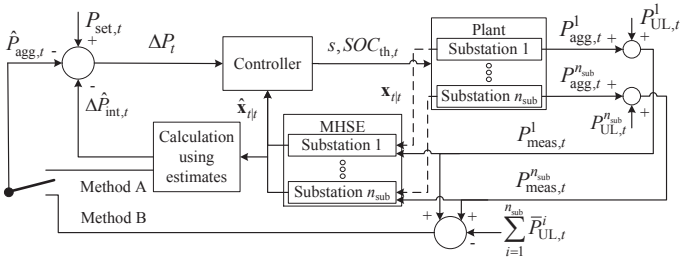


Fig. 2. Control loop including moving horizon state estimation (MHSE).

hierarchical architecture described in Section II. At each time step t, the controller calculates the required change in power

$$\Delta P_t = P_{\text{set},t} - \hat{P}_{\text{agg},t} - \Delta\hat{P}_{\text{int},t} \ , \tag{8}$$

where $P_{\text{set},t}$ is the desired set-point, $\hat{P}_{\text{agg},t}$ is the measured/estimated TCL aggregate power, and $\Delta\hat{P}_{\text{int},t}$ is the estimated change in power resulting from the TCLs' internal hysteresis controller actions, computable from TCL state estimates $\hat{\mathbf{x}}_{\text{l},t}$. Let $\mathbf{P}_{\text{n}}$ denote the vector of TCL power ratings; n_{sub} the number of substations in the network; $P^i_{\text{agg},t}$ the actual TCL aggregate power, $\bar{P}^i_{\text{UL},t}$ the predicted uncontrolled load, $P^i_{\text{UL},t} = \bar{P}^i_{\text{UL},t} + v_t$ the actual uncontrolled load, and $P^i_{\text{meas},t} = P^i_{\text{agg},t} + P^i_{\text{UL},t}$ the noisy aggregate power measurement, all at substation i. There are two ways to calculate $\hat{P}_{\text{agg},t}$: Method A) based on the state estimates, i.e. $\hat{P}_{\text{agg},t} = \mathbf{P}_{\text{n}}\hat{\mathbf{x}}_{\text{l},t}$, or Method B) based on the aggregate power measurements, i.e. $\hat{P}_{\text{agg},t} = \sum_{i=1}^{n_{\text{sub}}} \left(P^i_{\text{meas},t} - \bar{P}^i_{\text{UL},t}\right)$.

If $\Delta P_t < 0$ additional off switching is required, whereas if $\Delta P_t > 0$ additional on switching is required. The TCLs that will be switched are determined according to a priority list based on their estimated state of charge

$$\hat{SOC}_t = \frac{\hat{x}_{\text{c},t} - m}{M - m} \, , \tag{9}$$

where $\hat{x}_{\text{c},t}$ is the TCL's estimated temperature. At each time step, the controller broadcasts a pair $[SOC_{\text{th},t}, s_t]$, where $SOC_{\text{th},t} \in [0,1]$ is the $\hat{SOC}_t$ of the last TCL that enters the priority list, and $s_t \in \{0,1\}$ is a signal indicating whether an increase in consumption ($s_t = 1$) or a decrease in consumption ($s_t = 0$) is required. The TCLs that are outside of their dead-band are not controllable and ignore the control signal, whereas the rest respond based on their SOC.

Control actions effectively tighten TCL dead-bands. Define a TCL's temperature threshold as $x_{\text{th},t} = SOC_{\text{th},t}(M-m)+m$. Set $\tilde{M} = x_{\text{th},t}$ and $\tilde{m} = m$ if $s_t = 0$, and $\tilde{M} = M$ and $\tilde{m} = x_{\text{th},t}$ if $s_t = 1$. Replacing M and m with $\tilde{M}$ and $\tilde{m}$ in the logical relations of Section III-A, we get $[\delta_{1,t} = 1] \leftrightarrow [x_{\text{c},t} \geq \tilde{M}]$ and $[\delta_{2,t} = 1] \leftrightarrow [x_{\text{c},t} \leq \tilde{m}]$. Therefore, the external control actions can be directly incorporated into the MLD framework. The only difference is that the matrices $\tilde{E}_1$ and $\tilde{E}_3$ now depend on $\tilde{M}$ and $\tilde{m}$, and so are time-varying.

The control loop including MHSE is shown in Fig. 2. Note that the aggregator applies the MHSE method independently for each substation i.

IV. State Estimation Problem

We propose an MHSE method to estimate TCL states when current TCL state measurements are unavailable. At each time step $t \neq jT_m$, $j \in \mathbb{N}$ we solve a multi-period MILP with TCL temperatures, on/off modes, auxiliary binary variables, process and measurement noise as optimization variables. Define the optimization vector as $\boldsymbol{x}^{\text{opt}}_t := [\hat{\boldsymbol{x}}_{t-N+1|t}, \hat{\boldsymbol{\psi}}_{t-N+1|t}, \ldots, \hat{\boldsymbol{\psi}}_{t|t}]$, where $\hat{\boldsymbol{\psi}}_{k|t} := [\hat{\boldsymbol{\delta}}_{k|t}, \hat{\boldsymbol{w}}_{k|t}, \hat{v}_{k|t}]$, $\hat{\boldsymbol{\delta}}_{k|t} \in \{0,1\}^{4n_{\text{ap}}}$, $\hat{\boldsymbol{w}}_{k|t} \in \mathbb{R}^{n_{\text{ap}}}$, $\hat{v}_{k|t} \in \mathbb{R}$, $k \in [t-N+1, t]$, N as the estimation horizon, and $\hat{\cdot}_{k|t}$ as the estimate of $\cdot$ at time step k using measurements up to time step t. Note that $\hat{\boldsymbol{x}}_{t-N+2|t}, \ldots, \hat{\boldsymbol{x}}_{t|t}$ can be determined from $\hat{\boldsymbol{\delta}}_{k|t}$, $k \in [t-N+1, t]$, and therefore these additional optimization variables are not needed. With this notation, the estimation problem is

$$\begin{aligned}
\min_{\boldsymbol{x}^{\text{opt}}_t} \ & \sum_{k=t-N+1}^{t} m_1 \ |\hat{y}_{k|t} - y_k| + m_2 \ ||Q^{-1}\hat{\boldsymbol{w}}_{k|t}||_1 \\
& + m_3 \ |R^{-1}\hat{v}_{k|t}| + \sum_{k=t-N+1}^{t-1} m_4 \ ||Q^{-1}\left(\hat{\boldsymbol{x}}_{\text{c},k|t} - \hat{\boldsymbol{x}}_{\text{c},k|t-1}\right)||_1 \\
& + m_5 \sum_{i=1}^{n_{\text{ap}}} \left|\hat{W}^i_{k|t} - W^i\right| + m_6 \ \left|\hat{V}_{k|t} - V\right| \, , \qquad (10) \\
\text{s.t.} \quad & \hat{\boldsymbol{x}}_{k+1|t} = \boldsymbol{A}\hat{\boldsymbol{x}}_{k|t} + \boldsymbol{B}\hat{\boldsymbol{\delta}}_{k|t} + \boldsymbol{F}_t + \hat{\boldsymbol{w}}_{k|t} \, , \qquad (11) \\
& \hat{y}_{k|t} = \boldsymbol{C}_2\hat{\boldsymbol{x}}_{k|t} + \hat{v}_{k|t} \, , \qquad (12) \\
& \boldsymbol{E}_1\hat{\boldsymbol{\delta}}_{k|t} \leq \boldsymbol{E}_2\hat{\boldsymbol{x}}_{k|t} + \boldsymbol{E}_3 \, , \qquad (13) \\
& \hat{\boldsymbol{x}}_{k|t} = \boldsymbol{x}_{k|t} \ , \quad \forall t \in [jT_m + 1, jT_m + N - 1] \, , \\
& \qquad\qquad \forall k \in [t-N+1, jT_m] \, , \qquad (14)
\end{aligned}$$

where Q^{-1} and R^{-1} are the inverses of the process and measurement noise covariance matrices, respectively, $m_1 - m_6$ are weighting factors, W^i is a known statistic on the process noise of TCL i (e.g., the mean value), V is a known statistic on the measurement noise, and $\hat{W}^i_{k|t}$ and $\hat{V}_{k|t}$ are the current estimates of those statistics computed from $\hat{\boldsymbol{w}}_{k|t}$ and $\hat{v}_{k|t}$.

The first term of (10) minimizes the difference between the output calculated from the estimated states and the measured aggregate power. The second and third terms penalize the process and measurement noise. The fourth term is needed to link the current estimation problem to the results of the previous estimation problems. Note that only the continuous states are included in the fourth term. The fifth and sixth terms require that the current noise statistics are close to their known values. Equations (11)-(13) describe the TCL hybrid dynamics. For $t \in [jT_m + 1, jT_m + N - 1]$, $j \in \mathbb{N}$, noise-free TCL state measurements are available, which are taken into account by introducing the equality constraint (14), and setting $m_4 = 0$ in (10). For $t \geq jT_m + N$, (14) is not considered and $m_4 \neq 0$.

Note that the terms of (10) that are related to either w or v are normalized by either Q or R, which are assumed known. This ensures that the numeric values of the penalties on process and measurement noise are in the same range, which is essential for good performance. However, no assumption on the probability distributions of the noises is needed. The MHSE performance can be improved by appropriately tuning the weighting factors $m_1 - m_6$. For example, for longer estimation horizons, higher m_5 and m_6 might be desired. In addition, m_4 can be time-varying, i.e. a function of the quality of recent estimates. If no noise statistics are available, the proposed method can be applied by dropping the last two terms, and tuning m_2 and m_4 based on the numeric range of

TABLE I.
SPACE HEATER PARAMETERS, ADAPTED FROM [7], [8]

$P_{\text{n}} \sim U(5,9)$ kW	$C \sim U(5,9)$ kWh/°C	$R \sim U(1.5,2.5)$ °C/kW
$C_{\text{p}} \sim U(2,3)$	$T_{\text{sp}} \sim U(19,23)$ °C	$T_{\text{db}} \sim U(0.25,1)$ °C

w and m_3 based on the numeric range of v.

The MHSE problem is a large MILP, even for relatively small TCL aggregations and estimation horizons. Note that a 1-norm minimization is used in (10), instead of least squares minimization as in [15], [20], [21], because it can be reformulated into a set of linear inequalities. Adopting the hierarchical architecture of Section II, the estimation problem can be solved independently and in parallel for each substation, which makes it tractable in real-time since the number of consumers per substation and/or feeder is low, e.g., 20 to 40 consumers.

V. Investigation Setup

We demonstrate the performance of the MHSE method via two case studies. In case study A, we investigate the estimation quality under different process and measurement noise levels, and different controller forcing levels, using an open-loop controller. This analysis provides insights on the noise levels that could be handled by the estimator in real-world applications. In case study B, we demonstrate how the MHSE method can improve the performance of closed-loop LFC with TCL aggregations. For both case studies, we use space heaters and parameterize them by drawing their parameters from the uniform probability distributions of Table I.

We assume that w_t and v_t follow zero-mean normal distributions with known variances (the same for all TCLs) and no autocorrelation. Therefore, the last two terms of (10) become

$$m_5 \sum_{i=1}^{n_{\text{ap}}} \frac{1}{\sigma_{\text{w}}^i} \left| \hat{\mu}_{\text{w},k|t}^i - \mu_{\text{w}}^i \right| + m_6 \frac{1}{\sigma_{\text{v}}} \left| \hat{\mu}_{\text{v},k|t} - \mu_{\text{v}} \right| , \quad (15)$$

where σ_{w}^i is the process noise standard deviation for TCL i, σ_{v} is the measurement noise standard deviation, and the noise means $\mu_{\text{w}}^i = \mu_{\text{v}} = 0$. In this case, Q is diagonal.

Measurements of TCL states are assumed to be available every $T_m = 15$ minutes. We fix $m_1 = 10^8$, and so the first term of (10) is a soft constraint, and we set $m_2 - m_6$ equal to 1. The choice of m_4 is critical for the convergence of the estimator [21], and an empirical investigation showed that $m_4 = 1$ leads to a reasonable performance. The theoretically optimal estimation period N depends on the aggregation size, and could be chosen applying the observability tests proposed in [15]. However, in practical application, N would likely be determined by real-time computational limitations, and so we empirically select $N = 10$ based on simulation results. All simulations are done with a time step of 10 seconds in MATLAB using a 4 core machine (2.83 GHz) with 8 GB RAM, and the MHSE problem was solved using CPLEX.

VI. Benchmarking the MHSE Method

Consider a population of 20 TCLs connected at a substation along with other uncontrolled loads, and assume a TCL coincident load equal to 20% of the total substation load P_{sub}. We investigate several options for process and measurement noise: $\sigma_{\text{w}} = \{5 \cdot 10^{-4},\ 10^{-3},\ 5 \cdot 10^{-3}\}$, referred to as "low", "medium", and "high" process noise, and $\sigma_{\text{v}} = \{0,\ 0.02,\ 0.05,\ 0.1\} \cdot P_{\text{sub}}$, referred to as "zero", "low", "medium", and "high" measurement noise. The measurement noise is given as a percentage of P_{sub} since it represents the prediction errors of the uncontrolled demand.

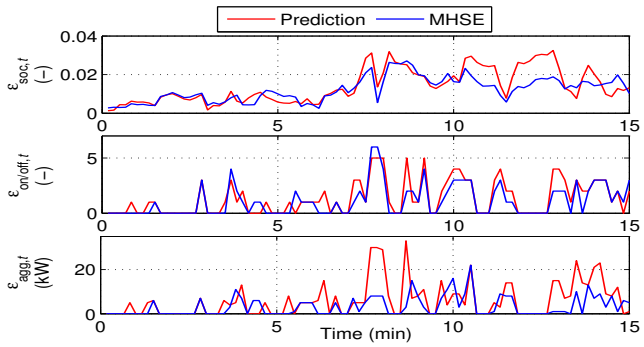


Fig. 3. Estimation errors for MHSE and the model-based predictor.

We compare the MHSE method against a simple model-based prediction approach, where the states are computed from (5), (6) assuming $w_t = 0$. We run simulations for 15 minutes and use as indicators the estimation errors of temperature, on/off mode, and aggregate power consumption at each time step. The controller broadcasts the random control signals

$$s_t = |s_{t-1} - r_{\text{s}}|, \quad SOC_{\text{th},t} = SOC_{\text{th},t-1} - r_{\text{soc}}(-1)^{s_t}, \quad (16)$$

where $r_{\text{s}} \in \{0,1\}$ is a discrete random variable that takes the value 1 with probability P_{s}, and $r_{\text{soc}} \in [0, R_{\text{soc}}]$ is a continuous uniformly distributed random variable. The interpretation of P_{s} is the probability with which the control direction changes between two consecutive time steps. The variable R_{soc} describes the change in the control signal magnitude between two consecutive time steps. Therefore, higher values of P_{s} and/or R_{soc} result in more aggressive control actions. In this paper, two different forcing levels are investigated: (a) a low forcing scenario with $P_{\text{s}} = 0.1$, and (b) a high forcing scenario with $P_{\text{s}} = 0.5$. In both cases, we fix $R_{\text{soc}} = 0.025$.

Depending on the measurement noise level, a group of TCLs with similar power ratings might not be distinguishable by the MHSE. However, what matters most is to estimate the aggregate power and the SOC of the group. For this reason, we cluster the TCLs by their power ratings and use these clusters to assess the estimation quality. Denote the number of clusters by n_{cl}, cluster i by CL_i, and define its SOC as

$$SOC_{\text{CL}_i,t} = \frac{\sum_{j \in \text{CL}_i} (x_{\text{c},t}^j - m^j)}{\sum_{j \in \text{CL}_i} (M^j - m^j)} . \quad (17)$$

The estimated SOC of cluster i can be defined similarly using $\hat{x}_{\text{c},t}^j$. We define the SOC and on/off mode estimation errors as

$$\varepsilon_{\text{soc},t} = \sum_{i=1}^{n_{\text{cl}}} \left| SOC_{\text{CL}_i,t} - \hat{SOC}_{\text{CL}_i,t} \right| , \quad (18)$$

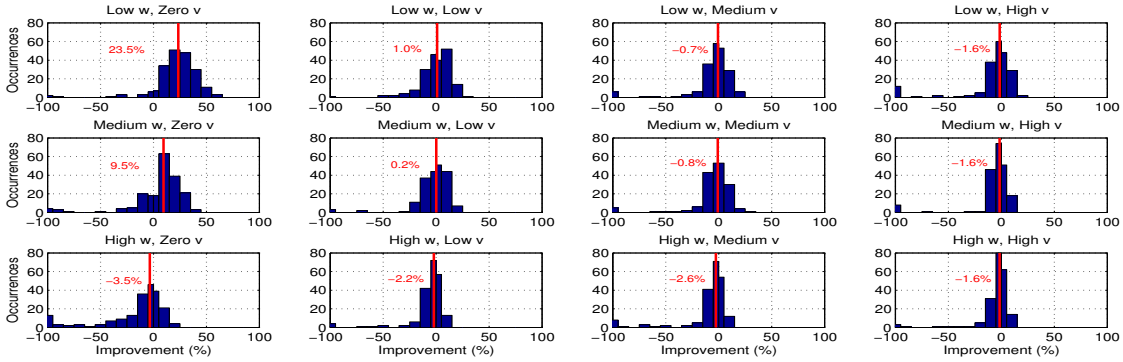
$$\varepsilon_{\text{on/off},t} = \sum_{i=1}^{n_{\text{cl}}} \left| \sum_{j \in \text{CL}_i} u_t^j - \sum_{j \in \text{CL}_i} \hat{u}_t^j \right| . \quad (19)$$

We define the TCL aggregate power estimation error as

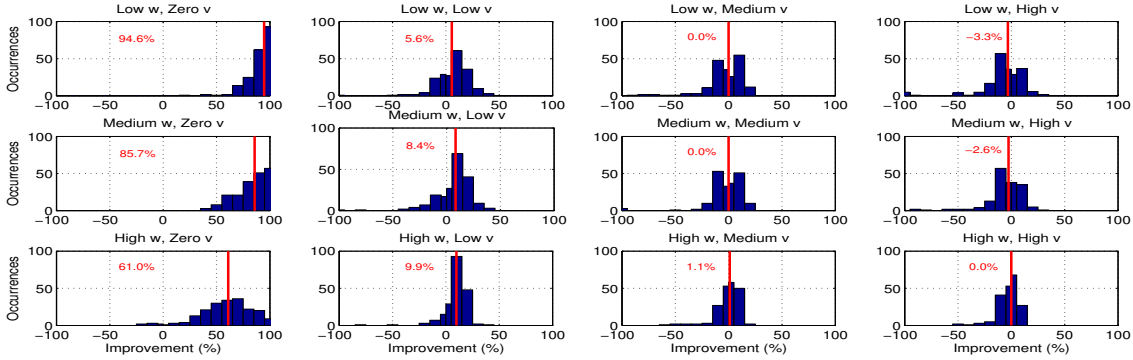
$$\varepsilon_{\text{agg},t} = \boldsymbol{C}_2 \boldsymbol{x}_t - \boldsymbol{C}_2 \hat{\boldsymbol{x}}_t . \quad (20)$$

The mean absolute error (MAE) over the simulation horizon N_{sim} is calculated as $\text{MAE} = (1/N_{\text{sim}}) \cdot \sum_{t=1}^{N_{\text{sim}}} |\varepsilon_t|$, where ε_t is any of $\varepsilon_{\text{soc},t}$, $\varepsilon_{\text{on/off},t}$, and $\varepsilon_{\text{agg},t}$.

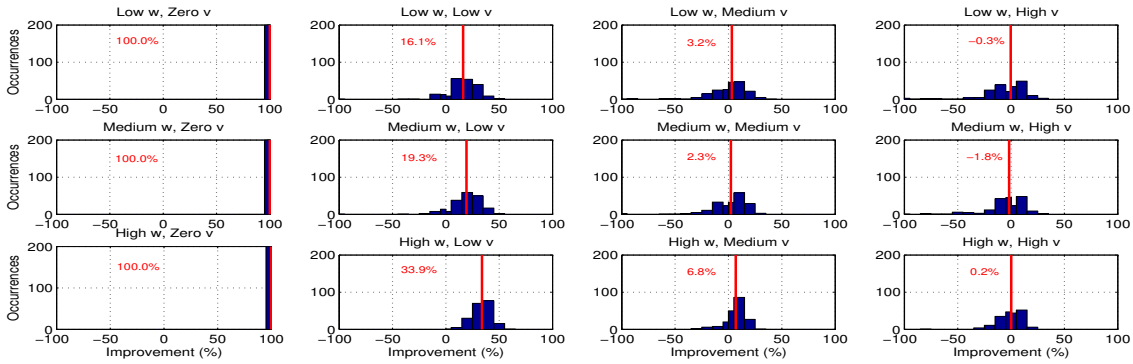
Figure 3 compares the MHSE method to the model-based prediction approach for a case with medium process noise,



(a) Histograms of SOC estimation improvement.



(b) Histograms of on/off mode estimation improvement.



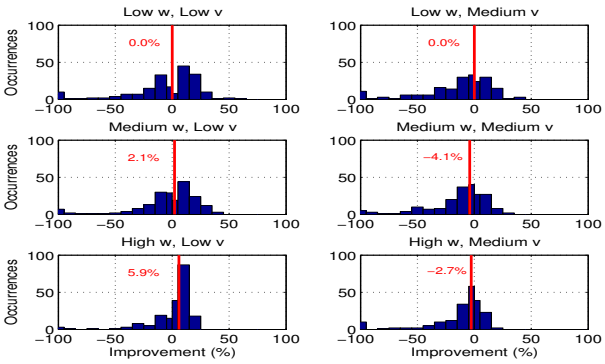
(c) Histograms of aggregate power estimation improvement.

Fig. 4. Comparison of MHSE method versus the model-based prediction approach for high forcing control. Red lines indicate medians.

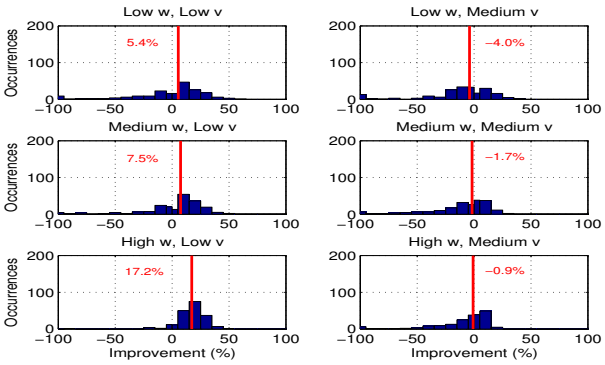
low measurement noise, high forcing, and $n_{\text{cl}} = 5$ clusters. The MHSE method provides better estimates most of the time, and reduces MAE_{soc} by 13%, $\text{MAE}_{\text{on/off}}$ by 19%, and MAE_{agg} by 47%. To more-fully benchmark the MHSE method, we repeat this analysis for each of the twelve combinations of process and measurement noise, and for each of the two forcing levels. For each case, we run 200 simulations with randomly generated TCL populations and noise realizations. Figure 4 shows histograms of estimation improvement for high forcing, where the red lines correspond to the medians.

With the exception of the cases $\{$low w, zero $v\}$ and $\{$medium w, zero $v\}$, the MHSE method does not improve the SOC estimates. This is because the TCL temperatures cannot be observed directly, but only through the on/off mode estimates. However, even with poor SOC estimates, the on/off modes and aggregate power can be estimated well for most of the cases. If $v = 0$, the MHSE method drastically improves the estimation quality especially for aggregate power, as also found in [16]. For all cases with low v and for the case $\{$high w, medium $v\}$, the estimates of the on/off modes and aggregate power generally improve, resulting in positive medians in the range $1.1\%-33.9\%$. Interestingly, for the case $\{$high w, low $v\}$ the MHSE method achieves better aggregate power estimates for all 200 scenarios, and better on/off mode estimates for 87% of them. On the other hand, for all cases with high v, the MHSE method performance is poor.

Figure 5 shows the results for on/off mode and aggregate power estimation for low controller forcing. For SOC estimates and for all cases with zero or high v, we observed patterns similar to those seen in the high forcing case; therefore, those histograms are omitted due to space limitations. The MHSE method outperforms the model-based prediction only for the cases $\{$medium w, low $v\}$, and $\{$high w, low $v\}$. Also, note that the resulting improvement is worse than in the respective high forcing cases. This provides useful intuition: the higher the forcing level, the more information is retrieved from



(a) Histograms of on/off mode estimation improvement.



(b) Histograms of aggregate power estimation improvement.

Fig. 5. Comparison of MHSE method versus the model-based prediction approach for low forcing control. Red lines indicate medians.

the aggregate power measurements improving the estimation quality. Overall, our results indicate that the MHSE method would perform the best in applications with high process and low measurement noise, irrespective of the forcing level.

VII. Application of MHSE in Frequency Control

We now investigate the value of MHSE in closed-loop control. Consider a population of n_{ap} TCLs evenly distributed among n_{sub} distribution substations. Based on Section VI, we select $\sigma_{\text{w}} = 5 \cdot 10^{-3}$ and $\sigma_{\text{v}} = 0.02 \cdot P_{\text{sub}}$, and fix the noise realizations. The desired set-point is a function of the baseline consumption of the TCL population ($P_{\text{b},t}$), the control band (α), and the control signal (Y_t)

$$P_{\text{set},t} = P_{\text{b},t} + \alpha Y_t P_{\text{b},t} \ . \tag{21}$$

We test two control signals: (a) an extract of the Swiss LFC signal from 2009 with $\alpha = 0.5$ and (b) the high-frequency component of the same signal obtained by applying a high-pass filter with cutoff frequency $1/30$ Hz and $\alpha = 3$. Both signals are shown in Fig. 6. Similar decompositions of the frequency signal into low and high frequency components were proposed in [22], [23]. The reason we investigate (b) is that a larger control band can be offered without energy constraint violations, and more forcing will be necessary, which is expected to improve the estimation/control performance based on Section VI. In each case, we compare a controller with MHSE against a controller with the model-based predictor, and consider Methods A and B to calculate $\hat{P}_{\text{agg},t}$. We also investigate a reference case with perfect TCL state information, which gives us a performance bound. Two

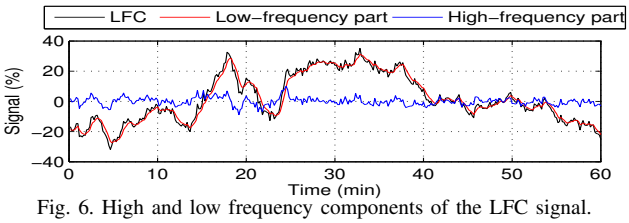


Fig. 6. High and low frequency components of the LFC signal.

TABLE II.
RMSE VALUES FOR SET-POINT TRACKING

	2 Substations		20 Substations	
Method	A	B	A	B
Standard LFC				
Reference (kW)	2.21	N/A	2.13	N/A
Prediction (kW)	11.84	16.40	31.66	36.67
MHSE (kW)	10.26	14.56	28.55	35.09
Improvement (%)	13.34	11.22	9.82	4.31
High frequency part of LFC				
Reference (kW)	2.29	N/A	2.18	N/A
Prediction (kW)	23.29	18.26	75.75	44.08
MHSE (kW)	18.13	15.67	55.18	40.18
Improvement (%)	22.16	14.18	27.16	8.85

aggregation sizes are considered: $n_{\text{ap}} = 40$ TCLs distributed among $n_{\text{sub}} = 2$ substations, and $n_{\text{ap}} = 400$ TCLs distributed among $n_{\text{sub}} = 20$ substations. All simulations are performed for 1 hour with a time step of 10 seconds.

We assess the controller performance by calculating the root mean square error (RMSE) between the desired set-point and the TCL aggregate power consumption, and we summarize the results in Table II. For all cases, the MHSE method outperforms the model-based predictor resulting in lower RMSE values. The improvement is more pronounced in the case of the high-frequency signal due to higher forcing, which is consistent with the results of Section VI. Interestingly, the best choice of method to calculate $\hat{P}_{\text{agg},t}$ depends on the signal. For the LFC signal Method A is preferable, whereas for the high-frequency signal Method B performs better. The reason is that the higher the frequency content of the signal, the higher the signal to (measurement) noise ratio; therefore, the aggregate power measurements $P^i_{\text{meas},t}$ can be trusted more. These observations are valid for both $n_{\text{sub}} = 2$ and $n_{\text{sub}} = 20$. The best results for each case are highlighted in grey in Table II. Overall, the MHSE method reduces the RMSE by approximately 14% for $n_{\text{sub}} = 2$ and 9% for $n_{\text{sub}} = 20$.

The tracking performance and the broadcasted control signals for $n_{\text{sub}} = 20$ are shown in Figs. 7 and 8. Note that the high-frequency signal results in a more aggressive s_t time series, and so SOC_{th} values are much closer to 0.5, compared to that of the LFC signal. Table III shows the average number of switching actions per TCL for the highlighted cases of Table II. For the LFC signal, MHSE slightly increases the number of switching actions compared to the model-based predictor; however, for the high-frequency signal, it significantly reduces the number of switching actions, in particular for $n_{\text{sub}} = 2$. In this case, MHSE not only improves controller performance, but also decreases the wear on TCLs.

Beyond filtering the LFC signal, there are additional ways of increasing forcing levels to improve MHSE/controller performance. For example, the TCL population could be divided into two groups, each tasked with following a high frequency signal that sum to the original LFC signal. Additionally, one

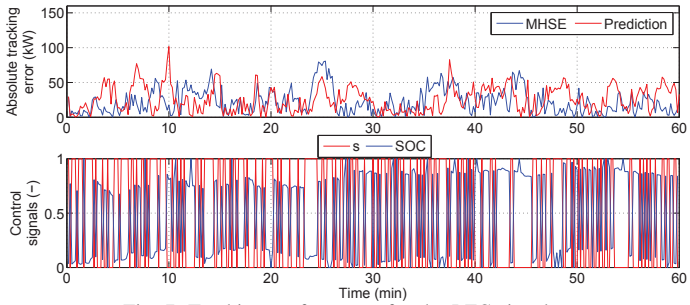


Fig. 7. Tracking performance for the LFC signal.

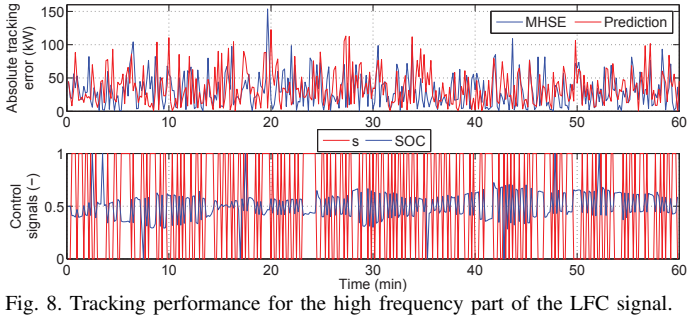


Fig. 8. Tracking performance for the high frequency part of the LFC signal.

could add an artificial signal to the LFC signal, and that signal could be balanced by another resource.

VIII. Concluding Remarks

In this paper, we presented a novel moving horizon state estimation (MHSE) method to estimate individual states of thermostatically controlled loads (TCLs). The method improves control performance in power trajectory tracking applications, and is designed to work with realistic measurements, i.e. real-time noisy aggregate power measurements from distribution substations and TCL state measurements from smart meters that arrive at lower frequency intervals. We demonstrated the MHSE performance under different process and measurement noise characteristics and controller forcing levels, and benchmarked it against a simpler model-based predictor.

Simulations showed that the proposed method provides accurate estimates for a certain spectrum of process and measurement noise. We also proposed a scalable hierarchical closed-loop control structure for load frequency control provision with TCLs, which uses the MHSE method. The controller/estimation performance was assessed for different frequency control signals, and for different aggregations sizes. Based on our simulations, MHSE generally reduces the controller tracking errors. Future work will investigate the estimator performance with different types of noise distributions, and in the presence of autocorrelation.

TABLE III.
AVERAGE SWITCHING ACTIONS PER TCL

	2 Substations	20 Substations
Standard LFC		
Reference	3.82	2.56
Prediction	3.93	2.57
MHSE	4.63	2.63
High frequency part of LFC		
Reference	10.40	9.57
Prediction	23.23	10.69
MHSE	17.68	9.43

References

[1] Y. Makarov, C. Loutan, J. Ma, and P. De Mello, "Operational impacts of wind generation on California power systems," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 1039–1050, 2009.

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

[3] C. Perfumo, E. Kofman, J. H. Braslavsky, and J. K. Ward, "Load management: Model-based control of aggregate power for populations of thermostatically controlled loads," *Energy Conversion and Management*, vol. 55, pp. 36–48, 2012.

[4] S. Koch, J. Mathieu, and D. Callaway, "Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services," in *Proceedings of the Power Systems Computation Conference*, Stockholm, Sweden, Aug. 2011.

[5] J. Kondoh, N. Lu, and D. Hammerstrom, "An evaluation of the water heater load potential for providing regulation service," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1309–1316, 2011.

[6] E. Vrettos, S. Koch, and G. Andersson, "Load frequency control by aggregations of thermally stratified electric water heaters," in *Proceedings of the IEEE PES Innovative Smart Grid Technologies Europe*, Berlin, Germany, Oct. 2012.

[7] J. Mathieu and D. Callaway, "State estimation and control of heterogeneous thermostatically controlled loads for load following," in *Proceedings of the Hawaii International Conference on Systems Science*, Wailea, HI, Jan. 2012, pp. 2002–2011.

[8] 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.

[9] E. Can Kara, Z. Kolter, M. Berges, B. Krogh, G. Hug, and T. Yuksel, "A moving horizon state estimator in the control of thermostatically controlled loads for demand response," in *Proceedings of the IEEE SmartGridComm Conference*, Vancouver, Canada, 2013.

[10] T. Borsche, F. Oldewurtel, and G. Andersson, "Minimizing communication cost for demand response using state estimation," in *Proceedings of the PowerTech Conference*, Grenoble, France, 2013.

[11] K. Kalsi, M. Elizondo, J. Fuller, S. Lu, and D. Chassin, "Development and validation of aggregated models for thermostatic controlled loads with demand response," in *Proceedings of the Hawaii International Conference on Systems Science*, Wailea, HI, 2012.

[12] S. Moura, J. Bendtsen, and V. Ruiz, "Observer design for boundary coupled PDEs: Application to thermostatically controlled loads in smart grids," in *Proceedings of the IEEE Conference on Decision and Control*, Florence, Italy, Dec. 2013.

[13] I. Hwang, H. Balakrishnan, and C. Tomlin, "State estimation for hybrid systems: applications to aircraft tracking," *IEE Proceedings of Control Theory and Applications*, vol. 153, no. 5, pp. 556–566, 2006.

[14] M. Hofbaur and B. Williams, "Hybrid estimation of complex systems," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, vol. 34, no. 5, pp. 2178–2191, 2004.

[15] A. Bemporad, G. Ferrari-Trecate, and M. Morari, "Observability and controllability of piecewise affine and hybrid systems," *IEEE Transactions on Automatic Control*, vol. 45, no. 10, pp. 1864–1876, Oct. 2000.

[16] E. Vrettos, J. L. Mathieu, and G. Andersson, "Demand response with moving horizon estimation of individual thermostatic load states from aggregate power measurements," in *Proceedings of the American Control Conference*, Portland, USA, June 2014.

[17] F. Chen, J. Dai, B. Wang, S. Sahu, M. Naphade, and C. T. Lu, "Activity analysis based on low sample rate smart meters," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, USA, Aug. 2011.

[18] C. Y. Chong and A. S. Debs, "Statistical synthesis of power system functional load models," in *Proceedings of the IEEE Conference on Decision and Control*, Fort Lauderdale, USA, Dec. 1979.

[19] S. Ihara and F. Schweppe, "Physically based modeling of cold load pickup," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 9, pp. 4142–4150, 1981.

[20] A. Bemporad, D. Mignone, and M. Morari, "Moving horizon estimation for hybrid systems and fault detection," in *Proceedings of the American Control Conference*, San Diego, USA, June 1999, pp. 2471–2475.

[21] G. Ferrari-Trecate, D. Mignone, and M. Morari, "Moving horizon estimation for hybrid systems," *IEEE Transactions on Automatic Control*, vol. 47, no. 10, pp. 1663–1676, Oct. 2002.

[22] Y. Makarov, P. Du, M. Kintner-Meyer, and H. Illian, "Sizing energy storage to accommodate high penetration of variable energy resources," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 34–40, Jan. 2012.

[23] S. Meyn, P. Barooah, A. Bušić, and J. Ehren, "Ancillary service to the grid from deferrable loads: the case for intelligent pool pumps in florida," in *Proceedings of the IEEE Conference on Decision and Control*, Florence, Italy, Dec. 2013.