

Identification of Hot Water End-use Process of Electric Water Heaters from Energy Measurements*

Adil Khurram, Luis Duffaut Espinosa^{†,‡} and Mads Almassalkhi
Department of Electrical and Biomedical Engineering
University of Vermont, Burlington, VT 05405 USA
{akhurram, lduffaut, malmassa}@uvm.edu

Roland Malhamé
Department of Electrical Engineering
Polytechnique Montréal, Montreal, QC H3T-1J4, CA
roland.malhamé@polymtl.ca

Abstract—This paper presents an algorithm for the identification of parameters for a stochastic hot water end-use process that drives a homogeneous population of thermostatically controlled electric water heaters (EWH). Usually, only metered interval consumption data (kWh) is collected and the hot water end-use process is unobservable to utility and aggregators. However, the availability of EWHs for demand response (DR) is closely coupled with the hot water end-use process. In this context, the hot water end-use process is modeled as a two-state Markov chain (Use / No use), which causes the thermostatic ON-OFF switching process to behave as a Markov renewal process (MRP). A set of first passage-time problems is developed to obtain the moments of the transition probability densities of the MRP. These problems are addressed by establishing a system of coupled partial differential equations characterizing the temperature evolution of the EWH population. A key quantity in the methodology for estimating the parameters is the total time an EWH is ON within a period of interest. It is referred to as the total busy time. Total busy time in this approach is a random variable for which analytical expressions of the moments are developed as a function of the metered window length. The latter expressions become the basis of a hot water demand model identification algorithm which is validated using agent-based simulations of EWHs.

Index Terms—Electric water heaters, demand response, estimation, Markov renewal processes, first passage time.

I. INTRODUCTION

Connectivity is becoming ubiquitous and, with smart appliances today, it is technically and economically feasible to leverage available distributed energy resources (DERs), such as “smart” air-conditioners and electric water heaters (EWH) to provide ancillary services using demand dispatch [1]. In demand dispatch, the DERs are aggregated, coordinated, and dispatched to provide grid services while taking into account local quality of service (QoS) requirements for the end-users. For example, the QoS requirement for EWHs dictates the temperature range specifying the device’s thermostatic controller parameters needed to ensure that water in the tank is maintained within a desirable temperature range. Together with QoS requirements, the DER end-use dynamics, such as hot water extraction rates from an EWH, place limits on the feasible range of demand dispatch schemes. This is because QoS specifications and end-use dynamics together determine the nominal power consumption of the device. Since QoS

specification (e.g., temperature range) do not change over time, they are static and relatively simple to model. However, end-use is generally a stochastic process and strongly affects nominal power consumption of a DER. Thus, having an accurate estimate of the uncontrollable end-use process can be valuable to predict performance of demand dispatch schemes.

Several DER control architectures have been proposed in the literature for demand dispatch schemes and a variety of local/distributed control policies to model the aggregate response (or flexibility) of DER populations [2]–[6]. Their overall aim is to utilize the flexibility available from the DERs while preserving QoS. However, these aggregate models assume that the underlying end-use process is known *a priori*, which in a practical setting is generally not the case. In other words, the expected capability of these demand dispatch schemes strongly depends on knowing the end-use process. In this paper, we propose a method to overcome this challenge by estimating representative end-use process parameters for a population of EWHs only from measured electric utility meter data (kWh) and physical water heater parameters.

Aggregate models of DERs generally assume the underlying end-use process to be white noise with drift that represents the fluctuations in energy due to uncontrollable end-user events as in [3]. The aggregated heat loss or gain due to end-user events in the case of heating and cooling loads such as electric space heaters and refrigerators can be accurately represented by the white noise process. However, the energy losses in an EWH are only due to: (i) water extraction from the tank and (ii) standing losses due to ambient conditions. Therefore, white noise is not an adequate representation of the end-use process for EWHs [7]. In [8], the end-use process is described by a Poisson random pulse (PRP) with randomized pulse amplitudes and widths, which is representative of the physics of water extraction processes [7]. While a hot water end-use process consists of varying water extraction rates, this paper focuses on a two state continuous-time Markov chain with constant water extraction rates as a starting point for the challenging stochastic parameter estimation problem [9]. Relatively constant periods of hot water extraction rates is a reasonable assumption due to the correlated nature of human activities over the hours of a day.

The literature on estimating end-use process parameters include data-driven methods [10], where historical electric meter data is used to develop regression-based models for load forecasting purposes. However, these “business as usual”

[†]Corresponding author.

*This work was supported by the U.S. DOE ARPA-E award DE-AR0000694.

[‡]This author was supported by NSF grant CMMI-1839387.

regression-based models are insufficient to predict both DER behavior and the effect on QoS when subjected to demand dispatch [9]. A method to generate hot water profile based on average energy consumption per activity such as bath and laundry is developed in [11]. Statistical models using time use data of daily activities of household members gathered from surveys such as American Time Use Survey (ATUS) data or time use data by Statistics Sweden (SCB), to predict the controlled load behavior have been developed to tackle this modeling gap [12]–[15]. These household activity data inform a model about the interaction between humans and their appliances. Then, the models are aggregated to predict the energy consumption of all residential households. While this approach is promising, it relies on high fidelity data, which is usually unavailable and not generalizable.

The impact of utility control on the load behavior can be seen from Fig. 1 which shows the aggregate power consumption of approximately 1700 EWHs in Vermont over several days. Daily profiles of power consumption usually consist of a morning peak, between 7 am and 9 am in Fig. 1 and an evening peak. In this figure, the utility turned OFF all water heaters between 2pm to 6pm resulting in cold load pick-up setting a peak right after 6pm. This peak is significantly different than the one observed in the morning and is due to the type of demand response program deployed by the utility. Note that the aggregate demand profile of EWHs can be divided into hourly periods of relatively constant demand as characterized by constant water extraction rates. Therefore, capturing the underlying water extraction process is helpful to predict even the controlled load dynamics as shown herein.

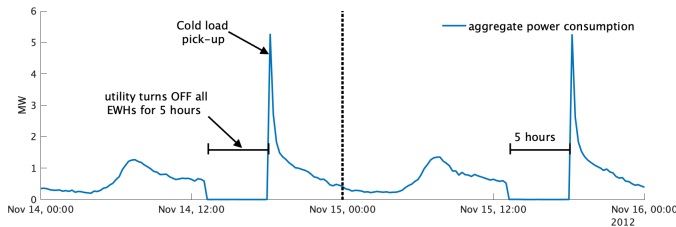


Figure 1. Average power consumption of 1700 real EWHs in VT

This paper builds on and extends the estimation methods presented in [9], [16], [17] where physics-based models of electric space heaters are employed. The estimation scheme first introduces a total busy-time random variable defined as the total ON time of the heater within fixed-time intervals. The rationale for defining this quantity is that by splitting the power consumption data of an electric water heater into periods of stationary statistics, one can relate the total thermostat ON time to the underlying likewise stationary water extraction statistics. Recursive relations for the moments of the total busy time are then used to develop an estimation algorithm for calculating the parameters of the stochastic hot water end-use process [9]. This paper extends the estimation scheme to the physically-based models of electric water heaters with two key contributions:

(1) The analytical results from [7], which were only valid for low water extraction rates have now been generalized to the case of arbitrary water extraction rates. This includes:

(i) The generalization of the coupled Kolmogorov equations representing the aggregate dynamics of a homogeneous group of EWHs to the case of arbitrary water extraction statistics; and (ii) the development of an adequate set of first passage-time probability density functions that are used to obtain the moments of the total ON time over fixed time windows.

(2) From the analytical contribution, a practically relevant identification procedure is developed and validated for estimating stochastic parameters of the unmeasured, hot water end-use process based only on interval meter readings and physical (tank) parameters for a homogeneous population of EWHs.

II. OVERVIEW OF IDENTIFICATION PROCEDURE

This section provides an overview of the inputs and outputs of the identification procedure, as illustrated in Fig. 2.

A. Availability of the metered data

In general, gathering data on hot water end-use processes requires expensive, device-level flow meters. In rare cases, sensors may be available to measure water extraction rates for the entire residence (all water) or device-level (hot water only) [18]. Furthermore, it can be seen from Fig. 1 that the power consumption of electric water heaters driven by the end-use process vary significantly throughout a day since it is a non-stationary random process. However, it can be considered stationary during durations of near-constant electric demand [7], [17], e.g. the morning peak between 7am and 9am in Fig. 1. Thus, we classify the daily kWh meter data into periods of statistically stationary hot water usage and propose the estimation strategy on one such period. This strategy can easily be generalized to multiple distinct periods that make up a representative day or a week.

B. The electric water heater model

The EWH considered herein consists of a first-order, simplified model with just a single equivalent heating element and an “average,” lumped temperature state. The hot water is extracted from the top of the tank and the cold water enters from the bottom. The temperature dynamics are governed by the following ordinary differential equation (ODE),

$$\frac{dx(t)}{dt} = \frac{P^{\text{rate}}m(t)}{c\rho L\eta} - \frac{(x(t) - x_a)}{\tau_L} - \frac{(x(t) - x_{\text{in}})}{60L}\bar{w}(t), \quad (1)$$

where $x(t)$ is the average temperature of the electric water heater, x_{in} is the temperature of the cold water entering through the tank inlet, x_a is the ambient temperature, $c = 4.186$ [kJ/kg-°C] is the specific heat capacity of 50°C water, $\rho = 0.988$ [kg/liters] is the density of hot water, L [liters] is the capacity of the water heater tank, P^{rate} is the rated power in kW of the heating element, η is the heat transfer efficiency, τ_L is the time constant representing the standing losses. The uncontrollable rate at which hot water is extracted from the tank is given by $\bar{w}(t) := w(t)q(t)$ [liters/min], where $q(t) \in \{0, 1\}$ is the logic state for the hot water usage process, i.e., $q(t) = 1$, if water is extracted from the tank at rate $w(t)$ [liters/min] at time t ; else $\bar{w}(t) = q(t) = 0$. The EWH operates in thermostat mode and $m(t) \in \{0, 1\}$ represents the physical state of the mechanical relay (open \equiv 1) at time t . The thermostat control

logic maintains the temperature within the user-specified, fixed dead-band $[x_-, x_+]$, $x_- < x_+$. The logic switches from ON ($m(t) = 1$) to OFF ($m(t) = 0$) at the upper boundary (x_+) and from OFF to ON at the lower boundary (x_-).

This paper considers the case of a homogeneous group of electric water heaters whose physical parameters and dead-band settings are known from manufacturer specifications and user-preferences, respectively. Energy measurements are then used as proxies for the time an EWH is ON within a time window of interest and referred to as the total busy time. The moments of the total busy time random variable are derived in Section V. The proposed estimation strategy, shown in Fig.2, takes as input the energy measurements and computes the total busy time within successive time windows of interest. Statistics of the total busy time along with the physical parameters of a homogeneous group of EWH are used to estimate the parameters of the unobservable end-use process, which is mathematically described in the next section.

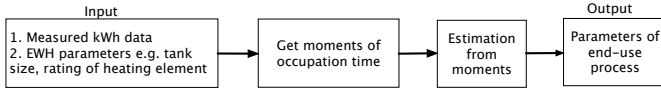


Figure 2. Overview of the end-use process identification problem

III. MODELING OF THE HOT WATER END-USE PROCESS AND ESTIMATION

This section describes the modeling of water heater end-use process and the corresponding Markov renewal process (MRP) defined at the switching instants of the thermostat. A set of first passage time problems are then presented to determine the transition density functions of the MRP.

A. Modeling water heater end-use process

Consider the rate of hot water extraction from the tank of an individual EWH to be constant (i.e., $w(t) \equiv W$). Then, the hot water end-use process is either in demand ($q(t) = 1$) or not in demand ($q(t) = 0$), and evolves according to what is assumed to be a two-state ($\{0, 1\}$) Markov chain model $q(t)$ (see [7]). The corresponding time invariant transition probabilities are given by,

$$P(q(t+h) = 1 | q(t) = 0) = \lambda_0 h + o(h) \quad (2)$$

$$P(q(t+h) = 0 | q(t) = 1) = \lambda_1 h + o(h) \quad (3)$$

where $h > 0$ is a small time increment. The electric water heater operating under this simplified demand process maintains the temperature within the deadband by the operation of the thermostat switch. A Markov renewal process, $y(t)$ is defined by recording the thermostatic switching instants ($m(t)$) which occur at the edges of the deadband [7]. The Markov renewal process, $y(t)$ consists of four states $\{0, 0', 1, 1'\}$, as illustrated in Fig. 3, where 1 represents the onset of a power consumption without hot water use, $1'$ indicates the onset of a power consumption in the presence of hot water use. Similarly, 0 denotes the onset of a power interruption with no hot water use and, finally, $0'$ indicates the onset of a power interruption but with hot water use. The transitions between states can only occur at the edges of the deadband, i.e. the switching instants of the EWH's thermostat transitions from 1 to 0, if

$x(t) = x_+$ without hot water use and transitions from 1 to $0'$, if $x(t) = x_+$ with hot water use. The remaining transitions follow in a similar fashion. Note that Fig. 3 also includes the first passage time probability density functions, $g_{ij}(t)$, which are defined as,

$$g_{ij} d\tau := P[t \leq \bar{t} < t + d\tau, y(\bar{t}) = j], \quad (4)$$

$\forall i, j = \{1, 1', 0, 1'\}$, where \bar{t} is the first time the MRP $y(t)$ switches to state j given that $y(t)$ has just switched to state i at $t = 0$. The transitions between the states 1 and $1'$ and 0 and $0'$ are not possible since the MRP $y(t)$ as defined, switches state only when the thermostat changes state, at which time the end-use state $q(t)$ is also recorded. More specifically, $g_{11'}(t) = g_{1'1}(t) = g_{00'}(t) = g_{0'0}(t) = 0$. In the next subsection, the statistical evolution of the ensemble of homogeneous EWHs is obtained from two coupled Kolmogorov equations with boundary conditions. These equations are then used to express the g_{ij} probability density functions.

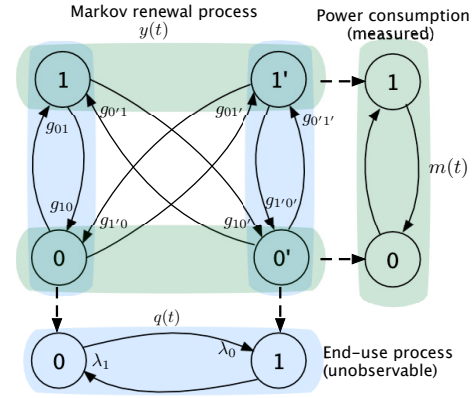


Figure 3. State transition diagram of the Markov renewal process

B. Partial differential equation description of load dynamics

The probability density functions, $f_i, f_{i'}$, associated with the Markov process consisting of the continuous state $x(t)$, the discrete state $m(t)$ and the hot water end-use process $q(t)$ are defined by,

$$f_i(\lambda, t) d\lambda = P[\lambda \leq x(t) \leq \lambda + d\lambda, m(t) = i, q(t) = 0] \quad (5)$$

$$f_{i'}(\lambda, t) d\lambda = P[\lambda \leq x(t) \leq \lambda + d\lambda, m(t) = i, q(t) = 1] \quad (6)$$

for $i = \{1, 0\}$. The probability density functions satisfy the following system of coupled partial differential equations [7],

$$\frac{\partial}{\partial t} \vec{f}_i(x, t) = V_i \frac{\partial}{\partial x} \vec{f}_i(x, t) + \Lambda^\top \vec{f}_i(x, t) \quad (7)$$

where,

$$\vec{f}_i(x, t) = \begin{bmatrix} f_i(x, t) \\ f_{i'}(x, t) \end{bmatrix}, V_i = \text{diag}\{v_i, v_{i'}\}, \quad (8)$$

$$\Lambda = \begin{bmatrix} \lambda_0 & -\lambda_0 \\ -\lambda_1 & \lambda_1 \end{bmatrix}, v_i = \frac{x_- - x_a}{\tau} - \left(\frac{P^{\text{rate}}}{cpL\eta} \right) i,$$

$$v_{i'} = \frac{x_- - x_a}{\tau} - \left(\frac{P^{\text{rate}}}{cpL\eta} \right) i + A,$$

for all $i \in \{0, 1\}$, where $A := \frac{x_- - x_{in}}{60L} W$ is the heat loss from the tank due to hot water extraction. In A , note that variable $x(t)$ has been replaced with constant x_- in order to make the

analysis more tractable. This is an acceptable approximation since the exact loss rate should not vary too much over a small (a few degrees) temperature deadband. The *cooling rates* are represented by, $v_1, v_{1'} \leq 0$ when thermostat is ON, and $v_0, v_{0'}$ when the thermostat is OFF. The conditional transition probability functions $g_{ij}(t)$ are obtained by the first passages of the $x(t)$ temperature process to the x_- or x_+ boundaries. The corresponding transition probability functions are given by (see [7])

$$g_{ij}(t) = v_i f_i(x_+, t), \quad (9)$$

$$g_{ji}(t) = v_j f_j(x_-, t), \quad \forall i = \{1, 1'\}, j = \{0, 0'\}. \quad (10)$$

In the next subsection, we use (7), (9) and (10) to derive a set of recursive equations yielding analytical expressions of any order moments of the $g_{ij}(t)$ first passage-time densities.

C. High water usage case and the first passage time problems

The previous section introduced the coupled system of PDEs that govern the time evolution of the probability density functions $\vec{f}_i(x, t)$ away from the edges of the thermostat deadband. It should be noted here that $v_1 \leq 0$ and $v_{1'} \leq 0$ represents the case of low hot water use in which the temperature of the tank increases in the presence of hot water use. However, during periods of high hot water use, the tank temperature decreases instead and is characterized by $v_1 \leq 0$ and $v_{1'} > 0$. The first passage time analysis for the case of lower hot water use was developed in [7]. In this section, the focus is on the more important case of high hot water use in which the temperature decreases in the presence of hot water use even though the thermostat is ON which is represented by $v_1 \leq 0, v_{1'} > 0$ in the system of coupled PDEs.

Theorem 1: Let $\vec{m}_k(x) := [m_k^{(1)}(x), m_k^{(1')}(x)]^\top$ be the vector of moments of the k^{th} order corresponding to the vector probability density function: $\vec{f}_1(x, t), k \geq 0$. The vectors $\vec{m}_k(x)$ satisfy the following recursive system of linear ordinary differential equations (ODE):

$$\frac{d}{dx} \vec{m}_0(x) = V_1^{-1} \Lambda^\top \vec{m}_0(x) + V_1^{-1} \vec{f}_1(x, 0), \quad (11)$$

and for all $k \geq 1$, we have

$$\frac{d}{dx} \vec{m}_k(x) = V_1^{-1} \Lambda^\top \vec{m}_k(x) - k V_1^{-1} \vec{m}_{k-1}(x), \quad (12)$$

with the absorbing boundary conditions $\vec{f}_1(x_+, t) = \vec{0}$ and $\vec{f}_1(x_0, t) = \vec{0}$, where $-\infty < x_0 < x_-$, $\vec{0} \in \mathbb{R}^{2 \times 1}$ and the initial condition,

$$\vec{f}_1(x, 0^+) = \begin{bmatrix} \delta(x - x_-) \\ 0 \end{bmatrix} \quad (13)$$

for the first passage time in 1 and

$$\vec{f}_{1'}(x, 0^+) = \begin{bmatrix} 0 \\ \delta(x - x_-) \end{bmatrix} \quad (14)$$

for the first passage time in $1'$. Further defining,

$$\vec{\Gamma}_k(x)^\top = [\vec{m}_0(x) \quad \vec{m}_1(x) \quad \dots \quad \vec{m}_k(x)], \quad (15)$$

it obeys the following ODE,

$$\frac{d}{dx} \vec{\Gamma}_k(x) = A_k \vec{\Gamma}_k(x) + B_k u_k(x) \quad (16)$$

$$A_k = \begin{bmatrix} V_1^{-1} \Lambda^\top & 0 & 0 & \dots & 0 \\ V_1^{-1} & V_1^{-1} \Lambda^\top & 0 & \dots & 0 \\ 0 & 2V_1^{-1} & V_1^{-1} \Lambda^\top & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \dots & \dots & kV_1^{-1} & V_1^{-1} \Lambda^\top \end{bmatrix}$$

$$B_k = \begin{bmatrix} V_1^{-1} & 0 & \dots & 0 \\ 0 & I & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & I \end{bmatrix}, u(x) = \begin{bmatrix} \vec{f}(x, 0) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

with the condition,

$$\lim_{x_0 \rightarrow -\infty} \vec{\Gamma}(x_0) = 0, \quad -\infty < x_0 < x_-. \quad (17)$$

Proof : See Appendix A ■

It should be noted here that in the case under consideration of high water usage, MRP $y(t)$ can only exit from 1 and $1'$ in state 0, owing to the fact that temperature always decreases in the presence of water demand. Furthermore, one can derive a similar system of equations for the moments starting at 0 or $0'$. It is omitted for lack of space.

The previous theorem provides initial conditions, boundary conditions and a system of linear ODEs to carry out first passage time computations under high water extraction rates. By solving the system of ODEs in (16) one can derive analytical expressions of the moments of the first passage time densities g_{ij} . First passage time process is conceptually depicted in Fig. 4. It follows the temperature of an EWH and corresponding MRP states visited, as it enters the lower edge (x_-) in state 1 and transitions to 0 at the top edge (x_+) of the deadband. Within the deadband, temperature decreases with rate $-v_{1'}$ when water is being extracted from the tank and increases otherwise with rate $-v_1$. The particular set-up of Fig. 4 is used to obtain the moments of g_{10} and $g_{10'}$.

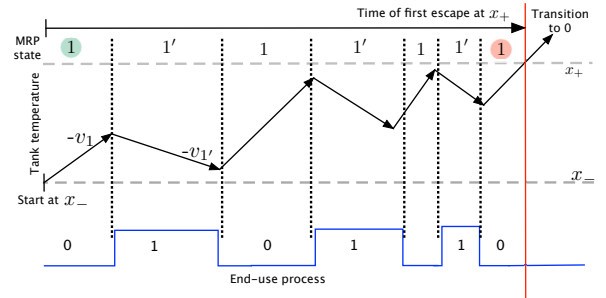


Figure 4. This figure illustrates the first passage time process starting at the lower edge of the deadband (x_-) in state 1 (thermostat ON and without hot water use). The first passage time corresponds to the first time the temperature reaches the upper boundary (x_+) of the deadband.

IV. ILLUSTRATIVE EXAMPLE

In this section, an illustrative example is presented for the first passage time moment calculations described in the previous section. These moments lead to approximate analytical expressions for the transition probability density functions g_{ij} that are essential to relate the statistics of the power consumption data to the end-use process.

A. Solution of first passage time problems

Consider a homogeneous group of electric water heaters with cooling rates $v_1, v_{1'}$ and $v_0, v_{0'}$ and simplified case of two moments, i.e., $\vec{m}_0(x)$ and $\vec{m}_1(x)$. The linear system (16) can be written for the first passage time in 1s as,

$$\frac{d}{dx} \vec{\Gamma}_1(x) = A_1 \vec{\Gamma}_1(x) - B_1 \vec{u}(x) \quad (18)$$

$$A_1 = \begin{bmatrix} \frac{\lambda_0}{v_1} & -\frac{\lambda_1}{v_1} & 0 & 0 \\ -\frac{\lambda_0}{v_{1'}} & \frac{\lambda_1}{v_{1'}} & 0 & 0 \\ -\frac{1}{v_1} & 0 & \frac{\lambda_0}{v_1} & -\frac{\lambda_1}{v_1} \\ 0 & -\frac{1}{v_{1'}} & -\frac{\lambda_0}{v_{1'}} & \frac{\lambda_1}{v_{1'}} \end{bmatrix}, B_1 = \begin{bmatrix} V^{-1} & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix}, \quad (19)$$

$$\vec{u}(x) = \begin{bmatrix} \vec{f}_1(x, 0) \\ \vec{0} \end{bmatrix}, V_1 = \begin{bmatrix} v_1 & 0 \\ 0 & v_{1'} \end{bmatrix} \quad (20)$$

$$\vec{m}_0(x) = \begin{bmatrix} m_0^{(1)}(x) \\ m_0^{(1')}(x) \end{bmatrix}, \vec{m}_1(x) = \begin{bmatrix} m_1^{(1)}(x) \\ m_1^{(1')}(x) \end{bmatrix}, \quad (21)$$

$$\vec{f}(x, 0) = \begin{bmatrix} \delta(x - x_-) \\ 0 \end{bmatrix}, \Lambda^\top = \begin{bmatrix} -\lambda_0 & \lambda_1 \\ \lambda_0 & -\lambda_1 \end{bmatrix}, \quad (22)$$

where zero vector $\vec{0} \in \mathbb{R}^{2 \times 1}$, zero matrix $\mathbf{0} \in \mathbb{R}^{2 \times 2}$, and I is identity matrix. This system has two repeated eigenvalues:

$$\gamma_1 = \frac{\lambda_0 + \lambda_1}{v_1 v_{1'}} \left(\frac{\lambda_0}{\lambda_0 + \lambda_1} v_{1'} + \frac{\lambda_1}{\lambda_0 + \lambda_1} v_1 \right) \quad (23)$$

$$\gamma_2 = 0. \quad (24)$$

The average heating rate should be positive or equivalently the average cooling rate should be negative, because the probability flux should escape entirely from the upper boundary ($x = x_+$). The mathematical consequence of this fact is that the nonzero eigenvalue (γ_1) of the system should be positive ($\gamma_1 > 0$) and the average cooling rate

$$\left(\frac{\lambda_0}{\lambda_0 + \lambda_1} v_{1'} + \frac{\lambda_1}{\lambda_0 + \lambda_1} v_1 \right) < 0, \quad (25)$$

which indeed implies that $\gamma_1 > 0$. Solving the initial value problem described in the previous section results in the following zero order and first order moments,

$$\vec{m}_0(x_+) = \left[-\frac{1}{v_1}, 0 \right]^\top \quad (26)$$

$$\vec{m}_1(x_+) = \left[\left(\frac{1}{v_1} \right) \frac{\Delta}{\frac{\lambda_1}{\lambda_0 + \lambda_1} v_1 + \frac{\lambda_0}{\lambda_0 + \lambda_1} v_{1'}}, 0 \right]^\top \quad (27)$$

Similarly, first passage time calculations can be performed for the $1'$ state. For the remaining states ($0, 0'$) a similar procedure is used and its details are omitted here.

B. Approximation of g_{ij} by moment matching

The conditional probability density functions (pdfs) g_{ij} are approximated by the moment matching techniques, in which the pdfs are represented by the approximated functions \hat{g}_{ij} whose moments match those obtained from the solution of the first passage time problems. In this paper, only the zero and first order moments are considered resulting in the following Markovian-type (exponential) approximation of the pdf,

$$\hat{g}_{ij}(t) = \frac{m_0^{ij}}{m_1^{ij}} \exp \left\{ -\frac{m_0^{ij}}{m_1^{ij}} t \right\} \quad (28)$$

where m_0^{ij}, m_1^{ij} are the appropriate zero order and first order moments of g_{ij} . In Laplace domain,

$$\hat{G}_{ij}(s) = \left(\frac{(m_0^{ij})^2}{m_1^{ij}} \right) \left(s + \frac{m_0^{ij}}{m_1^{ij}} \right)^{-1} \quad (29)$$

This result can be extended to generalized phase-type distributions by matching any number of moments depending upon the desired accuracy and is the topic of ongoing work [19].

V. PARAMETER ESTIMATION FROM TOTAL BUSY TIME

In this section, we bring together the results from prior sections and propose an estimation strategy based on the available data. The MRP defined at the switching instants of the thermostat classifies the process into four states $\{1, 1', 0, 0'\}$. However, the available utility grade power consumption data cannot distinguish between 1 (thermostat ON, *without* hot water use) and $1'$ (thermostat ON, *with* hot water use) states since the hot water end-use process is not observed. Similarly, the states 0 and $0'$ are indistinguishable from measurements. However, in the stationary steady-state of the MRP, the total ON time random variables become identically distributed since the state at the start of a measurement interval becomes random with a common distribution. Therefore, in the stationary steady-state the states 1 and $1'$ are combined into an ON state with a density obtained by combining the 1 and $1'$ densities with weights $m_0^{(01)}$ and $m_0^{(01')}$ and similarly with states 0 and $0'$ which are probabilistically combined into an OFF state. The thermostat ON and OFF are represented by 1_A and 0_A , respectively, and the resulting process is a 1_A-0_A alternating renewal process (ARP). Subscript $(\cdot)_A$ is added to distinguish between the states of the MRP and the ARP. The stationary ARP statistics are then used to identify the parameters of the underlying hot water end-use process.

A. Moments of the total busy time

The parameters of 1_A-0_A ARP in its stationary steady-state which is blind to the initial MRP state at the start of power measurement windows, are now identifiable from the data available for estimation. Furthermore, let $\xi(t) = \int_0^t m(\tau) d\tau$ be the total time the thermostat is ON within a time period of length t . This variable $\xi(t)$ is also called *total busy time* random variable over an interval of length t . Recursive expressions for the moments of $\xi(t)$ in steady state were derived in [9] that do not require the knowledge of the state of the thermostat at the start of the window. The first-order and the second-order moments of $\xi(t)$, in the Laplace domain, are obtained after application of [Theorem 2 in 9] resulting in the following equilibrium distribution expressions,

$$E_{\text{eq}}[\xi(t)](s) = \frac{\mu_{1_A}}{\mu_{1_A} + \mu_{0_A}} \frac{1}{s^2}, \quad (30)$$

$$E_{\text{eq}}[\xi^2(t)](s) = \frac{\mu_{1_A}}{\mu_{1_A} + \mu_{0_A}} \frac{1}{s^2} - \frac{2}{(\mu_{1_A} + \mu_{0_A}) s^4} \frac{(1 - F_{0_A}(s))(1 - F_{1_A}(s))}{(1 - F_{0_A}(s)F_{1_A}(s))}, \quad (31)$$

where $F_i(s) = \mathcal{L}[f_i(t)]$, $i \in \{1_A, 0_A\}$, is the Laplace transform the pdf associated with 1_A and 0_A states, and $\mu_i = E[f_i(t)]$, $i \in$

$\{1_A, 0_A\}$ with $E[\cdot]$ being the usual expectation operator. A short description on computation of $F_i(s), \mu_i$ for $i \in \{1_A, 0_A\}$ is provided in Appendix B. We are finally in a position to estimate λ_0, λ_1 from the moments of $\xi(t)$ as is presented next.

B. Numerical validation of parameter estimation

The parameter estimation strategy is demonstrated on a period of relatively stationary water demand for example, the morning peak between 7am and 9am in Fig. 4. The case here is that of high water usage with the end-use parameters $\phi = \{\lambda_0, \lambda_1\}$, for which the transitions probability functions (g_{ij}) are derived in IV. Data for this type of estimation can be obtained by measuring the aggregated power consumption of a group of EWHs within the same period of interest over the course of several days, and then appended together. Following this line of thought, 10,000 EWHs are simulated for 16 hours, with tank size 250 litres, heating element rated at 4.5 kW, hot water rate of extraction of 5.4 litres/min, ambient and inlet temperature 21.1° C, thermostat set-point and deadband adjusted at 51° C and 6° C respectively. Aggregated power consumption of this group is measured where each 2 hour period is assumed to represent a single day. The simulated data then represents the aggregate power consumption over 8 days and in what follows, we show that this data at least in the simulation environment is sufficient to accurately estimate the unknown ϕ .

The proposed estimation problem determines $\phi^* = \{\lambda_0^*, \lambda_1^*\}$ that minimizes the loss function,

$$\phi^* = \underset{\phi}{\operatorname{argmin}} \|\hat{r}(t, \phi) - \tilde{r}(t, \phi)\|_2^2 \quad (32)$$

where $\hat{r}(t, \phi) = (\hat{E}_{\text{eq}}[\xi(t)](t, \phi), \hat{E}_{\text{eq}}[\xi^2(t)](t, \phi))^T$ is the empirical mean and second moment obtained from the data, $\tilde{r}(t, \phi) = (E_{\text{eq}}[\xi(t)](t, \phi), E_{\text{eq}}[\xi^2(t)](t, \phi))^T$ is the analytical mean and second order moment from (30), (31). The estimation problem (32) is solved using `lsqcurvefit` in Matlab for $t \in \{1, 2, 5, 15\}$ minutes. It can be seen from Table I that the estimated $\hat{\phi}$ are close to true ϕ . An immediate observation from the results in Table I, is that shorter windows result in the estimated parameters closer to the true values. This type of analysis will enable the utilities to collect appropriate metered data that results in reasonable estimates of the end-use consumption. Therefore, we show next the accuracy of the estimated $\hat{\phi}$ in the context of cold-load pickup of Fig. 1.

The objective now is to show a potential application of the estimation scheme in demand dispatch. Consider the same group of EWHs that generated the data for estimation under constant water demand. After 2 hours all EWHs are forced OFF for a period of 4 hours and subsequently allowed to turn back ON, mimicking the direct load control scheme from Fig. 1. Aggregate response of EWHs for the actual and estimated parameters is shown in Fig. 5 with $t = 2$ minutes. Clearly, the aggregated power demand and the mean tank temperature match well. Similar results are obtained for $t \in \{1, 5, 15\}$ minutes. The difference however, is in the transient response after EWHs are allowed to turn back ON, as shown in Fig. 6.

The estimated $\hat{\phi}$ differs slightly from the true ϕ as seen in Table I even though the steady state response of the estimated

system is exact. This is because the estimated $\hat{\phi} = \{\hat{\lambda}_1, \hat{\lambda}_0\}$ corresponds to the same steady state of the end-use process as the actual ϕ . The difference in the transient response, as shown in Fig. 6, is apparent from the nonzero eigenvalue γ_1 in (23). For the window size of 15 minutes, it follows from (23) that the nonzero eigenvalue obtained from the estimated parameters is twice the eigenvalue obtained from true parameters. One possible explanation for this behavior is that shorter windows correspond to increasing the sampling frequency of the ARP. Therefore, several window sizes may result in the same occupation behavior of the ARP. Furthermore, the estimates can be improved by including the correlation information between occupation time of successive windows for estimation as in [9]. However, further work is required to fully characterize the impact of this aliasing-type effect observed here. Nonetheless, the estimated values are helpful to model steady-state demand and average EWH QoS under homogeneous conditions.

TABLE I
COMPARISON BETWEEN ESTIMATED AND ACTUAL PARAMETERS

Window size	Actual	1 min	2 min	5 min	15 min
λ_0	0.0014	0.0014	0.0016	0.0021	0.0029
λ_1	0.0083	0.0084	0.0095	0.0120	0.0170

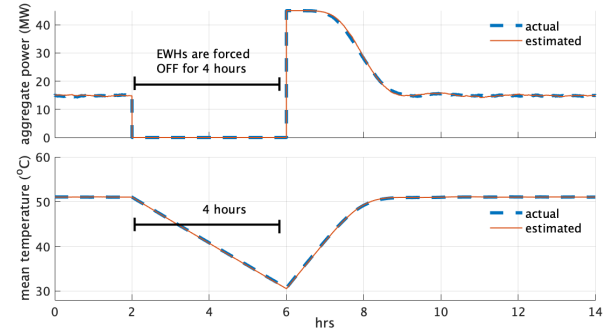


Figure 5. Aggregate power consumption of the group of EWHs is shown here when all EWHs are forced OFF for 4 hours. The window size for estimation is 2 minutes.

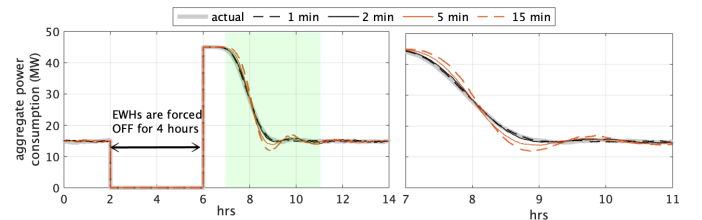


Figure 6. The aggregate power consumption of EWHs using the parameter estimates of ϕ obtained from different window sizes $t \in \{1, 2, 5, 15\}$ is plotted on the left and the shaded region is enhanced in the right plot. It shows that the aggregate power consumption differ in the oscillations before steady state is achieved.

VI. CONCLUSION AND FUTURE WORK

This paper develops preliminary results for an identification algorithm to estimate the parameters of an underlying hot water end-use process of electric water heaters from energy measurement. Unlike prior work in the area, which focused on low hot water extraction rates, this identification procedure has been generalized herein to include arbitrary extraction rates

and validated within a conventional DR setting for 10,000 EWHs. The estimated parameters serve to accurately model the dynamics of a homogeneous fleet of EWHs and is valuable for utilities to predict the controlled load behavior when subjected to demand dispatch.

Future work seeks to extend the procedure to estimate the water intensity rate, $w(t)$, to relax the homogeneous assumption on the fleet, and study the role of uncertainty in the physical EWH (tank) parameters on the end-use process estimates. Finally, we will incorporate actual interval meter data from a utility partner to estimate and optimize demand dispatch capability from a fleet of EWHs and compare against similar estimates from “black box” learning-based methods.

APPENDIX

A. Proof of theorems

Proof : (Theorem 1) Upon taking the Laplace transforms of the partial differential equations in (7) for $i = 1$ results in

$$\frac{\partial \vec{F}(x, s)}{\partial x} = V_1^{-1}(sI - \Lambda^\top)\vec{F}(x, s) - V_1^{-1}\vec{f}(x, 0). \quad (33)$$

Since the k^{th} order moment is defined as,

$$\vec{m}_k(x) = (-1)^k \left. \frac{\partial^k \vec{F}(x, s)}{\partial s^k} \right|_{s=0}, \quad \forall k = 0, 1, \dots, \quad (34)$$

therefore, setting $s = 0$ in (33) yields the linear first order ODE of zero order moment ($\vec{m}_0(x)$) as,

$$\frac{d\vec{m}_0(x)}{dx} = -V_1^{-1}\Lambda^\top\vec{m}_0(x) - V_1^{-1}\vec{f}(x, 0^+). \quad (35)$$

The first order moment is obtained by taking the derivative of (33) w.r.t. s and consequently setting $s = 0$ which results in,

$$\frac{d\vec{m}_1(x)}{dx} = -V_i^{-1}\Lambda^\top\vec{m}_1(x) - V_i^{-1}\vec{m}_0(x). \quad (36)$$

Hence the k^{th} order moments is given by,

$$\frac{d\vec{m}_k(x)}{dx} = -V_i^{-1}\Lambda^\top\vec{m}_k(x) - kV_i^{-1}\vec{m}_{k-1}(x) \quad (37)$$

which can be written in the form of the system of ODE of (16). The system of (16) consists of $2k$ equations and $4k$ unknowns. However, recall from the discussion in section II-C that $g_{10'}(t) = g_{1'0'} = 0$. Therefore, $\vec{m}_k^{(1')} = \vec{0}$ which reduces the number of unknowns to $3k$. Furthermore, owing to the decrease in temperature in $1'$ state, $\Gamma_k(x_-) \neq 0$ and is unknown, since there will always be some probability flux that crosses x_- . However, in a properly designed EWH, the long term mean upward temperature drift is positive and, the temperature of the tank will eventually reach x_+ . To obtain the remaining $1k$ linearly independent equations, consider an arbitrary boundary x_- such that $-\infty < x_0 < x_-$ with the condition (17). Evaluating (16) at $x = x_0$ and $x_i = x_-$ gives the remaining equations necessary to obtain the moments of first passage time densities. ■

B. Calculation of F_{1A}, F_{0A}

Probability density function $F_{1A}(s), F_{0A}(s)$ corresponding to the $1_A, 0_A$ of ARP are given by,

$$F_{1A}(s) = G_{10}(s)m_0^{(01)} + G_{1'0}(s)m_0^{(01')}, \quad (38)$$

$$F_{0A}(s) = G_{01}(s)m_0^{(10)} + G_{01'}(s)m_0^{(1'0)}, \quad (39)$$

where, $G_{10}(s), G_{01}(s), G_{1'0}(s), G_{01'}(s)$ are the transition probability functions obtained by solving first passage time problems as derived in Section IV, and $m_0^{(ij)}$ is the zero order moment of g_{ij} . The mean μ_{1A} and μ_{0A} associated with F_{1A} and F_{0A} respectively follows from (38), (39) after taking the expectation,

$$\mu_{1A} = m_0^{(01)}m_1^{(10)} + m_0^{(01')}m_1^{(1'0)}, \quad (40)$$

$$\mu_{0A} = m_0^{(10)}m_1^{(01)} + m_0^{(1'0)}m_1^{(01')}. \quad (41)$$

REFERENCES

- [1] A. Brooks, E. Lu, D. Reicher, C. Spirakis, and B. Wehl, “Demand dispatch,” *IEEE Power and Energy Magazine*, vol. 8, no. 3, pp. 20–29, May 2010.
- [2] S. P. Meyn, P. Barooah, A. Busic, Y. Chen, and J. Ehren, “Ancillary service to the grid Using intelligent deferrable loads,” *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 2847–2862, Nov. 2015.
- [3] J. L. Mathieu, S. Koch, and D. S. 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.
- [4] A. C. Kizilkale and R. P. Malhamé, “Mean field based control of power system dispersed energy storage devices for peak load relief,” in *52nd IEEE Conference on Decision and Control*, Dec 2013, pp. 4971–4976.
- [5] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, “Aggregate flexibility of thermostatically controlled loads,” *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 189–198, Jan 2015.
- [6] M. Almassalkhi, L. Duffaut Espinosa, P. D. H. Hines, J. Frolik, S. Paudyal, and M. Amini, *Asynchronous coordination of distributed energy resources with packetized energy management*. New York, NY: Springer New York, 2018, pp. 333–361.
- [7] R. Malhamé, “A jump-driven markovian electric load model,” *Advances in Applied Probability*, vol. 22, no. 3, p. 564–586, 1990.
- [8] L. A. Duffaut Espinosa, M. Almassalkhi, P. Hines, and J. Frolik, “Aggregate modeling and coordination of diverse energy resources under packetized energy management,” in *56th IEEE Conference on Decision and Control*, December 2017, pp. 1394–1400.
- [9] S. El-Ferik and R. P. Malhamé, “Identification of alternating renewal electric load models from energy measurements,” *IEEE Transactions on Automatic Control*, vol. 39, no. 6, pp. 1184–1196, June 1994.
- [10] T. Hong and S. Fan, “Probabilistic electric load forecasting: A tutorial review,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 914 – 938, 2016.
- [11] J. Kondoh, N. Lu, and D. J. 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, Aug 2011.
- [12] B. J. Johnson, M. R. Starke, O. A. Abdelaziz, R. K. Jackson, and L. M. Tolbert, “A method for modeling household occupant behavior to simulate residential energy consumption,” in *ISGT 2014*, Feb 2014, pp. 1–5.
- [13] J. Widén and E. Wackelgard, “A high-resolution stochastic model of domestic activity patterns and electricity demand,” *Applied Energy*, vol. 87, no. 6, pp. 1880 – 1892, 2010.
- [14] M. Muratori, M. C. Roberts, R. Sioshansi, V. Marano, and G. Rizzoni, “A highly resolved modeling technique to simulate residential power demand,” *Applied Energy*, vol. 107, pp. 465 – 473, 2013.
- [15] N. Arghira, L. Hawarah, S. Ploix, and M. Jacomino, “Prediction of appliances energy use in smart homes,” *Energy*, vol. 48, no. 1, pp. 128 – 134, 2012.
- [16] R. Malhamé and C.-Y. Chong, “Electric load model synthesis by diffusion approximation of a high-order hybrid-state stochastic system,” *IEEE Transactions on Automatic Control*, vol. 30, no. 9, pp. 854–860, 1985.
- [17] S. Fan, “Unsupervised learning based on markov chain modeling of hot water demand processes,” Master’s thesis, École Polytechnique de Montréal, Cambridge, 2017.
- [18] Pecan Street, “Annual report FY 2017-2018,” Pecan Street Inc. 2018.
- [19] T. Osogami and M. Harchol-Balter, “Closed form solutions for mapping general distributions to quasi-minimal PH distributions,” *Performance Evaluation*, vol. 63, no. 6, pp. 524 – 552, 2006.