



# Semi-Markov modelling in a Gibbs sampling algorithm for NIALM

Master's Thesis, Autumn 2013  
Department of Mathematics, div. of Mathematical Statistics  
Royal Institute of Technology, Stockholm

Supervisor: Timo Koski

Jean-Alexander Monin Nylund  
jamn@kth.se

### Abstract

Residential households in the EU are estimated to have a savings potential of around 27% [1]. The question yet remains on how to realize this savings potential. Non-Intrusive Appliance Load Monitoring (NIALM) aims to disaggregate the combination of household appliance energy signals with only measurements of the total household power load.

The core of this thesis has been the implementation of an extension to a Gibbs sampling model with Hidden Markov Models for energy disaggregation. The goal has been to improve overall performance, by including the duration times of electrical appliances in the probabilistic model.

The final algorithm was evaluated in comparison to the base algorithm, but results remained at the very best inconclusive, due to the model's inherent limitations.

The work was performed at the Swedish company Watty. Watty develops the first energy data analytic tool that can automate the energy efficiency process in buildings.

## Acknowledgements

I would like to thank my advisor, Timo Koski, for his support and comments during the writing of this thesis. I also wish to thank Felix Rios, who advised me during my work. Their help was most useful, and any errors that remain in this work are entirely mine.

To all of my friends and co-workers at Watty, I want to thank them for their continuous help, rewarding discussions and interesting input. I am especially indebted to Lars Lowe Sjöstrand, Jim Holmström, Lennart Liberg, Eva Andersson and Johan Laine who supported me with many technical and theoretical aspects in researching the developed algorithm. But foremost, I am grateful to all at Watty for having given me this opportunity, and turned these few months into such an exciting time.

I wish to thank my mother, Gunilla Monin Nylund, and my late father, Jean Monin, for their never ceasing support and encouragement. They inspired me to seek a higher education, something for which I can never thank them enough. To them, I dedicate this work.

Stockholm, December 30, 2013

Jean-Alexander Monin Nylund

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Purpose . . . . .	4
1.2	Outline . . . . .	4
<b>2</b>	<b>Background</b>	<b>5</b>
2.1	Short on graphical models . . . . .	5
2.2	Discrete-time Markov models . . . . .	6
2.2.1	Time-homogenous Markov chains . . . . .	6
2.2.2	Hidden Markov models . . . . .	7
2.2.3	Factorial hidden Markov models . . . . .	7
2.2.4	Semi-Markov models . . . . .	8
2.3	Gibbs sampling . . . . .	10
<b>3</b>	<b>Base Model</b>	<b>12</b>
3.1	Model assumptions . . . . .	12
3.1.1	Additive FHMM . . . . .	12
3.1.2	Gaussian emissions . . . . .	12
3.2	Approach . . . . .	13
3.2.1	State categorization . . . . .	13
3.2.2	Gibbs sampling with the fitted FHMM . . . . .	14
3.2.3	User-defined aspects of the Gibbs algorithm . . . . .	16
3.2.4	Gibbs sampling results . . . . .	16
3.2.5	Add-one smoothing . . . . .	17
<b>4</b>	<b>Method</b>	<b>18</b>
4.1	Model extension . . . . .	18
4.1.1	Explicit state duration . . . . .	18
4.2	State duration pre-analysis . . . . .	19
4.2.1	Fitting duration distributions . . . . .	20
4.3	Algorithm . . . . .	23
4.3.1	Quantization of prediction probabilities . . . . .	23
4.3.2	Computing conditional probabilities . . . . .	23
4.3.3	Conditional probability at chain edges . . . . .	26
<b>5</b>	<b>Results</b>	<b>27</b>
5.1	Experiment . . . . .	27
5.2	Binary classification test . . . . .	27
5.3	REDD-score . . . . .	28
5.4	Model convergence . . . . .	31
5.5	Energy disaggregation results . . . . .	33
5.5.1	Total energy use comparison . . . . .	33
5.5.2	Averaged energy usage time series . . . . .	35
5.5.3	Model behaviour at a single data point . . . . .	37

<b>6</b>	<b>Conclusions</b>	<b>42</b>
6.1	Binary classifications . . . . .	42
6.2	REDD score analysis . . . . .	42
6.3	Convergence analysis . . . . .	42
6.4	Total energy disaggregation results . . . . .	43
6.5	Comparison of model behaviour . . . . .	43
6.6	Accepting or rejecting the implemented model . . . . .	43
6.7	Further work . . . . .	44
<b>A</b>	<b>Proofs and calculations</b>	<b>45</b>
A.1	Conditional dependence in some graphical models . . . . .	45
	A.1.1 Chain structure . . . . .	45
	A.1.2 V-structure . . . . .	45
A.2	Factorizing the conditional probability distribution . . . . .	46
	A.2.1 FHMM . . . . .	46
	A.2.2 FHSMM . . . . .	47
<b>B</b>	<b>Data set</b>	<b>49</b>
B.1	Resampling REDD . . . . .	49
<b>C</b>	<b>ML distribution tables</b>	<b>50</b>
<b>D</b>	<b>Emission distributions</b>	<b>52</b>
<b>E</b>	<b>Duration distributions</b>	<b>56</b>

## 1 Introduction

The world's current thirst for energy, with its environmental impacts and negative external effects, have made energy technologies into an extremely relevant topic of research for the benefit of humanity's future existence. As of today, the world still mainly depends on non-renewable fossil fuels to cater for what in the end boils down to our every day needs.

While a great amount of research is directed toward the development of renewable resources, such as solar energy, wind power and the like, or even making our existing processes of energy extraction increasingly efficient, the gained benefits are still far from making our energy consumption sustainable, which raises concern towards a future when our current types of resources may not exist as abundantly as today.

Perhaps to no surprise, private households have been observed to have some of the largest capacities for improvement when it comes to efficient energy usage. Judging from a report by the Commission of the European Communities from 2006, residential households in the EU are estimated to have a savings potential of around 27% of their annual energy use. This amounts to a quarter of the full energy saving potential in all end-use sectors, including commercial buildings, transport and the manufacturing industry [1]. From this, the question still remains on how to best realize this savings potential.

In the domain of efficient energy usage lies the idea of saving energy by finding and providing detailed information about power consumption, to help make better decisions about energy spendings. This is something that can also be done at the individual household level by providing information on individual's consumption patterns.

Appliance load monitoring (ALM) is an ongoing field of study that deals in this matter, with the aim of providing direct feedback on household energy consumption at an individual appliance resolution. ALM has had two major approaches, that are described as *Intrusive* and *Non-Intrusive* ALM (shortened as IALM and NIALM). The IALM approach is also referred to as distributed sensing, as it focuses on measuring all household appliances separately. While field surveys and direct measurements of individual appliances have been and still are the most straight forward methods to acquire accurate energy usage data, the need for a multitude of sensors and time consuming installations have made IALM methods financially unapproachable [2]. Instead, some focus has shifted toward methods that automatically infer appliance energy signals with use of fewer sensors. Non-Intrusive Appliance Load Monitoring (NIALM), the principles of which the invention is attributed to Hart, Kern and Scheppe of MIT in the 1980's, aims to provide the desired functionality; to ably decompose one whole aggregated household power load signal into its underlying electrical appliances.

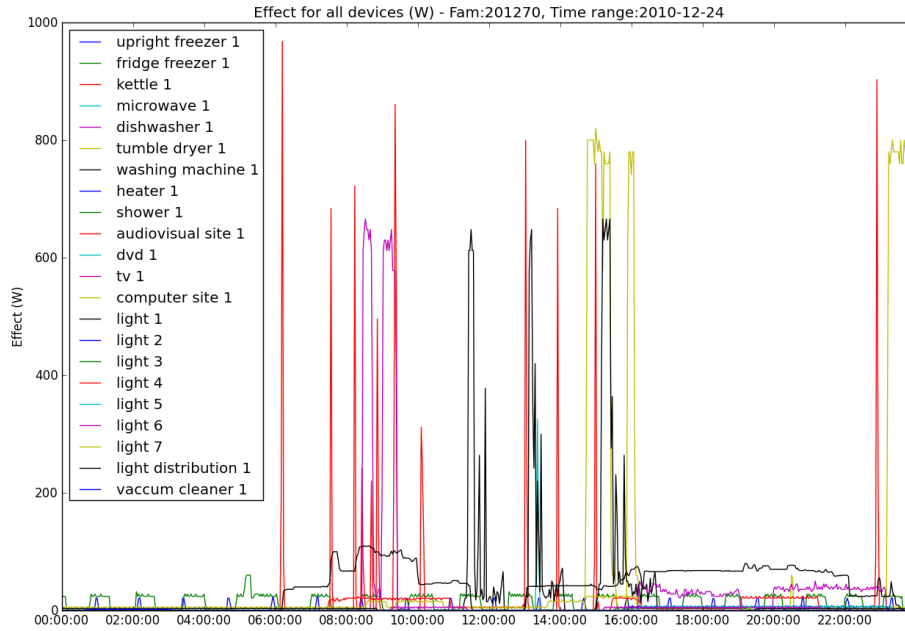
To estimate the component power loads of a household using nothing more than measurements of the total power load is typically a complex task, as is suggested by Figure 1 which depicts the individual load curves in a household with around 20 appliances, as well as the aggregated signal of the same household for comparison. At the core of NIALM is the idea that individual appliances have recognizable energy signals, also referred to as *signatures* [13]. The NIALM approach can vary depending on the appliance type, as well as what frequency has been used to sample the energy signal. While more information is obtained

from high resolution signals, the downside is the requirement of expensive measurement apparatus to sample at the required frequencies, which can be up to several kilohertz. For this reason, this thesis has been centred on NIALM at low or ultra-low frequency sampling (order of magnitudes around 1 to  $10^{-1}$  Hz).

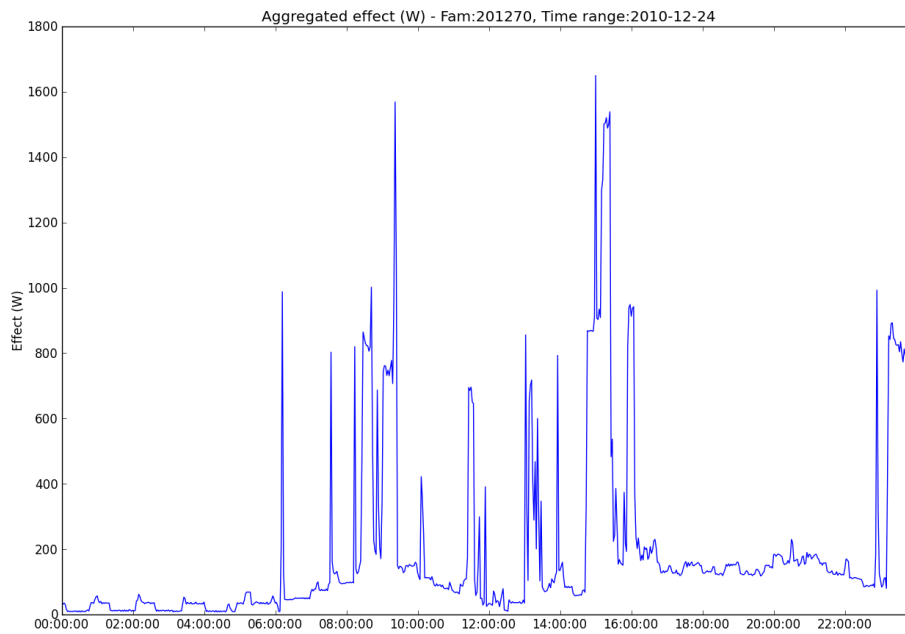
The work behind this thesis was performed at a Swedish company that specializes in developing NIALM methods using machine learning technologies. The work was to its largest extent centred around Factorial Hidden Markov Models (FHMM) and Gibbs sampling for power signal disaggregation. FHMMs were introduced by Ghahramani and Jordan, 1997, as an extension to the basic Hidden Markov Model (HMM), and has proven useful in areas such as speech recognition (Virtanen, 2006) and audio separation (Roweis, 2001).

One major weakness of conventional HMMs in the domain of NIALM is the less than accurate representation of how long typical appliances are used on a day to day basis. The *ON* or *OFF*-duration of an appliance is further referred to as its *state duration*, and for the regular HMM this duration will always be modelled as less probable the more time increases. Such state duration density could, as mentioned, in many cases be deemed inappropriate to model real life time series of appliances. Indeed, for household appliance state modelling purposes, one can easily present simple heuristic arguments to show that many household appliances have some state durations that are more likely than others. For example, the time a person spends watching television could be modelled after viewing habits, or a microwave oven's average run-time might be guessed as two minutes. In general, many appliances would not seem to exhibit strictly decreasing probability density functions for their state duration distributions.

The idea behind the implementation of this thesis has been taken from Rabiner's tutorial on HMMs [16], where explicit state durations are suggested as a possible improvement in HMMs. For this reason, the Hidden Semi-Markov Model (HSMM) is tried as a possible candidate to correctly model state durations. Semi-Markov chains try to include a representation of state duration and could as such be deemed a better choice than Markov models for appliance state durations.



(a) Individual appliance measurements



(b) Total load measurement

Figure 1: The energy disaggregation problem.



## 1.1 Purpose

The work described in this thesis was carried out at Watty. Watty is a Swedish company that develops the first energy data analysis tool that can automate the energy efficiency process in buildings. Their solution uses the total energy use in a building to evaluate its underlying systems and compare it to alternatives.

The work carried out at Watty has been divided between two areas; working as a part of the team responsible for the development and implementation of Watty's base model for energy disaggregation, and developing the algorithm discussed in this thesis.

The purpose of this thesis is to evaluate the results of a semi-Markov model extension on Watty's base model for NIALM, with the final aim being to improve overall energy disaggregation performance. Specifically, an algorithm that takes into account state durations (also referred to as state lengths) of appliances is researched and then added to the model. The subsequent performance results are compared with the original model's performance results and used for deciding whether or not to adopt the devised algorithm.

## 1.2 Outline

This thesis is organized as follows; first, section 2 gives the necessary introduction to Markov modelling as it is used in the rest of this work, while section 3 covers the functionality, structure and underlying assumptions of the original model used by the company at which the thesis work was performed.

Next, section 4 provides description of the developed algorithm, as well as the steps that lead to its design. Finally, sections 5 and 6 summarize the collection of results obtained from different model runs.

Additional results, along with description of the datasets used are displayed in appendices at the end of this thesis.

## 2 Background

This section contains the necessary theory that is implemented in the model in section 3. Section 2.1 provides a short introduction on graphical models, in preparation for section 2.2 and 2.3, which lay For a more thorough treatment of the Markov models presented below, see [4] and [16].

### 2.1 Short on graphical models

**Definition 1.** *The random variables  $X$  and  $Y$  are **conditionally independent** given  $Z$ , denoted  $X \perp Y|Z$ , if and only if the conditional joint probability can be written as a product of conditional marginal probabilities, or formally,*

$$X \perp Y|Z \iff P(X, Y|Z) = P(X|Z)P(Y|Z). \quad (1)$$

**Definition 2.** *The **chain rule** for probability theory says that any joint probability distribution of the random variables  $X_1, X_2, \dots, X_n$  can be calculated as*

$$P(X_n, \dots, X_1) = P(X_n|X_{n-1}, \dots, X_1) \cdot P(X_{n-1}, \dots, X_1) \quad (2)$$

$$= \prod_{k=1}^n P(X_k | \bigcap_{j=1}^{k-1} X_j). \quad (3)$$

A *probabilistic graphical model* is a way to represent a joint probability distribution by representing random variables as nodes, and dependence as node edges in a graph. A graphical model is illustrated in Figure 2.

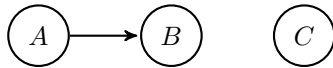


Figure 2: Directed graphical model.

In this figure, the nodes  $A$  and  $B$  are connected by the edge  $AB$ , representing that the random variables  $A$  and  $B$  are not independent of each other. The node  $C$  however remains unconnected and this random variable is thus independent of all other random variables in the model. The joint probability distribution is therefore

$$P(A, B, C) = P(C|A, B)P(B|A)P(A) \quad (4)$$

$$= P(C)P(B|A)P(A). \quad (5)$$

The graphical model displayed in Figure 2 is a *directed graphical model* (DGM), also known as a *Bayesian network*, where its edges are represented as arrows of direction. The directions indicate a believed causal relationships between the random variables, such as " $B$  is caused by  $A$ ".

Some definitions are useful when working with DGMs:

**Definition 3.** *The **parents** of a node are the set of nodes that direct **into** it.*

**Definition 4.** *The **children** of a node are the set of nodes that direct **away** from it.*

**Definition 5.** A **directed cycle** is a sequence of connected nodes such that we can get back to any node in the sequence by following the directed edges.

**Definition 6.** A **directed graphical model (DGM)** is graphical model with causal directions, but with no directed cycles.

For more on graphical models, see [10] and [15].

## 2.2 Discrete-time Markov models

For convenience, throughout this thesis the notation  $X_t$  will be used to signify  $X(t)$ , for a stochastic variable  $X$  depending on the variable  $t$ .

**Definition 7.** In probability theory, a sequence of stochastic time based events,  $\{X_n; n \in \mathbb{N}_0\}$ , with state space  $\mathbf{E}$  is said to have the (first-order) **Markov property** if

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n),$$

$\forall n$ , where  $x_n \in \mathbf{E}$  denotes the outcome of the  $n$ :th event in the sequence.

Informally put, the conditional probability distribution of future states of the process depend only upon the present state. For a finite or countable state space, the process is referred to as a (discrete-time) **Markov chain**.

**Definition 8.** The **transition probabilities** at the event  $n$ ,  $p_{i,j}(n)$ , of a Markov chain with state space  $\mathbf{E}$ , are defined as

$$p_{i,j}(n) = P(X_{n+1} = j | X_n = i), \quad i, j \in \mathbf{E}. \quad (6)$$

So  $p_{i,j}(n)$  is the probability to move from state  $i$  at event  $n$ , to state  $j$  at event  $n + 1$ . The conditional distribution of these probabilities can be represented in matrix form, as defined below.

**Definition 9.** The **transition matrix**  $\mathbf{P}(n)$  of a Markov chain with state space  $\mathbf{E} = \{1, 2, \dots\}$ , is defined as

$$\mathbf{P}(n) = \begin{bmatrix} p_{1,1}(n) & p_{1,2}(n) & \dots \\ p_{2,1}(n) & p_{2,2}(n) & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

### 2.2.1 Time-homogenous Markov chains

In the case where the transition probabilities are independent of time,

$$P(X_{n+1} = i | X_n = j) = P(X_n = i | X_{n-1} = j), \quad i, j \in \mathbf{E}, \quad \forall n, \quad (7)$$

the Markov chain is said to be *time-homogenous*. For such systems, the transition probabilities and transition matrix simplify to

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & \dots \\ p_{2,1} & p_{2,2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}. \quad (8)$$

### 2.2.2 Hidden Markov models

In *hidden Markov models* (HMM) the state  $X_n$  of the Markov model with state space  $\mathbf{E}$  is not directly observable. Instead, the observed stochastic process  $Y_n$  is a probabilistic function of the (underlying) state,

$$Y_n = f(X_n) \in \mathcal{O}, \quad (9)$$

where  $\mathcal{O}$  is the set of possible observations, also referred to as *emissions*. These emissions are conditionally independent given the underlying state sequence such that

$$P(Y_0, Y_1, \dots, Y_{n-1}, Y_n | X_0, X_1, \dots, X_{n-1}, X_n) \quad (10)$$

$$= \prod_{i=0}^n P(Y_i | X_i). \quad (11)$$

A probabilistic graphical model representation of a HMM is displayed in Figure 3. The model displayed is a directed graphical model (DGM), with  $2 \times T$  nodes, for  $T$  number of events.

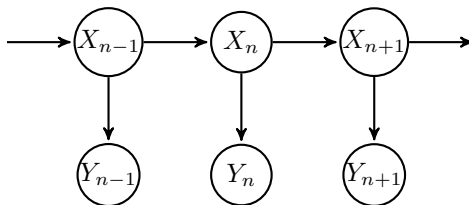


Figure 3: Hidden Markov model.

### 2.2.3 Factorial hidden Markov models

A generalization of HMMs was provided by Ghahramani and Jordan [5], in which the hidden state  $X_t$  is represented as a collection of independent state variables

$$X_t = (X_t^{(1)}, \dots, X_t^{(m)}, \dots, X_t^{(M)}),$$

corresponding to the underlying Markov chains in  $M$  individual HMMs, each with separate state space  $\mathbf{E}^{(m)}$  and emission process  $Y_t^{(m)}$ . The resulting observation from this model is a stochastic process  $\bar{Y}_t$  which is described as a function of the individual emissions,

$$\bar{Y}_t = f(Y_t^{(1)}, \dots, Y_t^{(M)}). \quad (12)$$

The model is referred to as a *factorial hidden Markov model* (FHMM), and its graphical model is illustrated in Figure 4. This model is a DGM, with  $((2 \times M) + 1) \times T$  nodes, for  $M$  appliances, and  $T$  time steps.

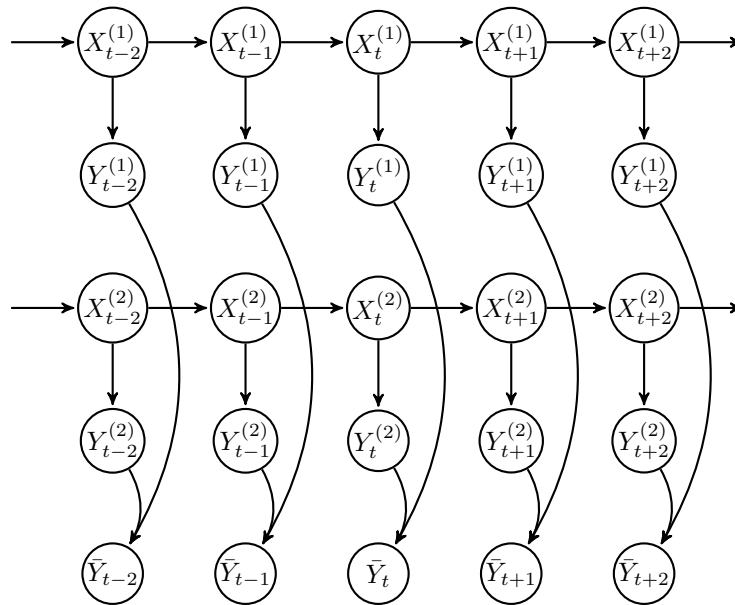


Figure 4: Factorial hidden Markov models.

**2.2.4 Semi-Markov models**

Markov processes can be extended by letting the duration between two following states  $X_n$  and  $X_{n+1}$  be stochastically modelled. A *Markov renewal process* denotes the sequence of states as  $(X_n, T_n)$  where each state,  $X_n$ , from a Markov chain, has an associated time of occurrence,  $T_n$ . Such a process is still Markovian; meaning that the Markov property is preserved throughout the sequence, and is illustrated in Figure 5.

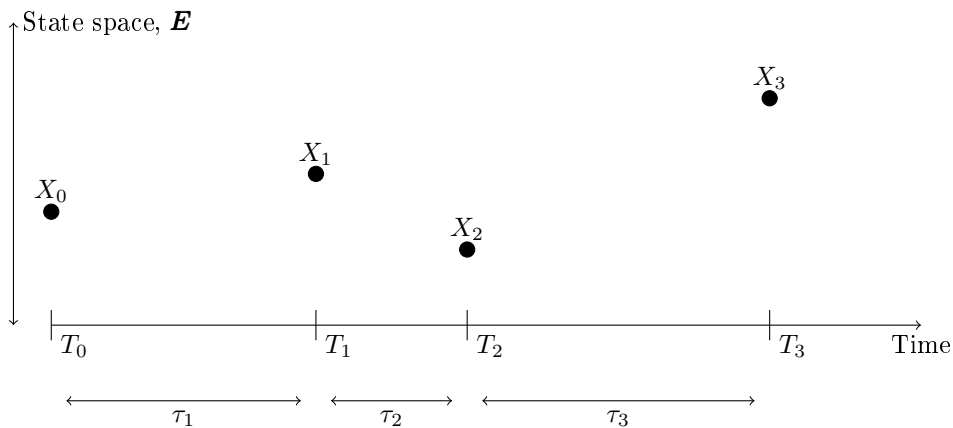


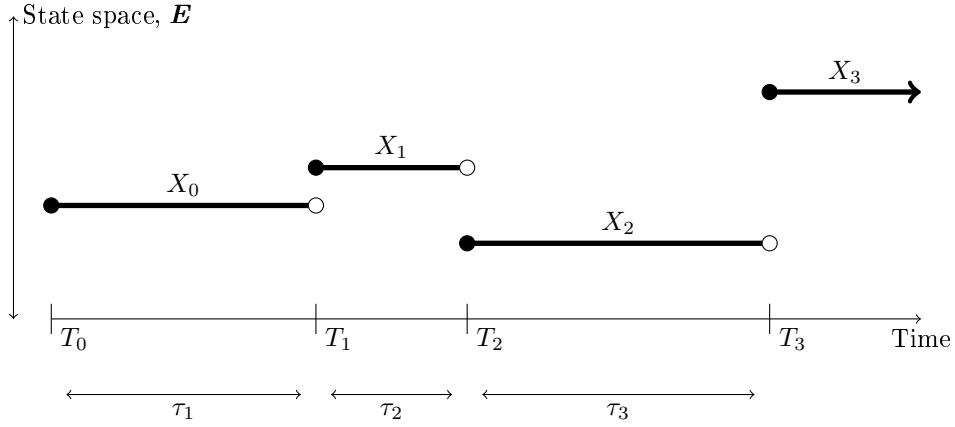
Figure 5: Markov renewal process.

Using the notation for Markov renewal processes, we define the *semi-Markov process* below.

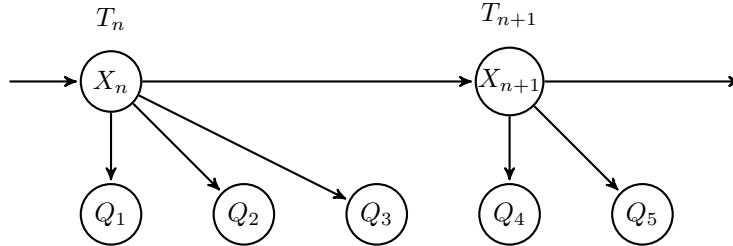
**Definition 10.** For a Markov chain with states  $X_n$  occurring at associated random times  $T_n$ , the *semi-Markov process*  $Q_t$  is defined as

$$Q_t = X_n, \quad t \in [T_n, T_{n+1}). \quad (13)$$

Intuitively, one may think of the semi-Markov process as observations of emission sequences for every occurring state in a Markov renewal process, see Figure 6.



(a) Analogy of semi-Markov process to Markov renewal process.



(b) A semi-Markov process, here  $Q_1 = Q_2 = Q_3 = X_n$ , and  $Q_4 = Q_5 = X_{n+1}$ .

Figure 6: Semi-Markov process.

In order to simplify the notations for future usage, we introduce the notion of the *generalized state* from [14],

$$G_n = (X_n, L_n), \quad (14)$$

$$Q_t = f(G_n), \quad (15)$$

which includes a stochastic duration variable  $L_n = T_{n+1} - T_n$ .

The process  $Q_t$  is not Markovian, but the underlying state sequence  $X_n$  in the Markov renewal process is. Therefore, since  $Q_t$  is dependent on  $X_n$ , the semi-Markov process will only transition between different states when its underlying Markov renewal process does. Using the generalized state notation displayed above, it is easily shown that the ensuing transitions in a semi-Markov process are also memoryless, since given a sequence of generalized states, the

conditional probability distribution of the following state is

$$P(Q_{T_{n+1}} = q_{T_{n+1}} | G_0, G_1, \dots, G_n), \quad (16)$$

using Definition 10, expression 16 translates as

$$= P(X_{n+1} = q_{T_{n+1}} | G_0, G_1, \dots, G_n) \quad (17)$$

expanding notation,

$$= P(X_{n+1} = q_{T_{n+1}} | X_0, X_1, \dots, X_n = x_n, L_0, L_1, \dots, L_n) \quad (18)$$

using  $L_n = T_{n+1} - T_n$ ,

$$= P(X_{n+1} = q_{T_{n+1}} | X_0, X_1, \dots, X_n = x_n, T_1, T_2, \dots, T_{n+1} = t_{n+1}) \quad (19)$$

using Markov property of the Markov renewal process,

$$= P(X_{n+1} = q_{t_{n+1}} | X_n = x_n) \quad (20)$$

change of notation,  $\tau = t_{n+1}$ ,

$$= P(X_{n+1} = q_\tau | X_n = x_n) \quad (21)$$

using Definition 10,

$$= P(Q_\tau = q_\tau | Q_{\tau-1} = q_{\tau-1}), \quad (22)$$

which holds if, and only if,  $q_\tau \neq q_{\tau-1}$ .

It will prove useful to refer to generalized states as *state blocks*, and as such one is able to talk about a *state block of length L* in order to refer to a generalized state  $G$  with duration  $L$ .

### 2.3 Gibbs sampling

*Gibbs sampling* is a Markov chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations from a specified multivariate probability distribution. MCMC algorithms were invented as methods that simulate complex dynamic systems, by generating samples from some constructed Markov chain. The crucial step in using MCMC algorithms is to construct the Markov chain such that its equilibrium distribution is the same as the distribution from the original system that is being simulated. An early example of MCMC usage was the simulation of a liquid in equilibrium with its gas phase, by Metropolis et al. (1953), in what has become known as the *Metropolis algorithm*.

The Gibbs sampling algorithm was introduced by Geman and Geman (1984), as a special case of the *Metropolis-Hastings* algorithm, a generalization of the Metropolis algorithm mentioned above, devised by Hastings (1970). The algorithm generates samples by sampling from the conditional distribution of the desired equilibrium distribution, and is illustrated with one iteration below:

For a given multivariate probability distribution,

$$P(X_1, X_2, \dots, X_{n-1}, X_n), \quad (23)$$

we denote the  $i$ :th sample from this distribution as  $\hat{X}^{(i)} = (x_0^{(i)}, x_1^{(i)}, \dots, x_n^{(i)})$ .

1. First, an initial sample

$$\hat{X}^{(0)} = (x_0^{(0)}, x_1^{(0)}, \dots, x_n^{(0)})$$

is selected from the distribution in (23).

2. Next, each variable  $x_j^{(1)}$  is sampled from the conditional distribution

$$P(X_j | x_0^{(1)}, x_1^{(1)}, \dots, x_{j-1}^{(1)}, x_{j+1}^{(0)}, \dots, x_n^{(0)}),$$

that is, the variable's distribution conditioned on all other variables in the joint distribution (23). After a variable has been sampled, this most recent value is used in the conditional distributions of the remaining variables.

3. Once all variables have been sampled individually, the resulting samples constitute one Gibbs sample  $\hat{X}^{(1)} = (x_0^{(1)}, x_1^{(1)}, \dots, x_n^{(1)})$ .
4. The procedure is then repeated with  $\hat{X}^{(1)}$  as the new initial sample.

For a sufficiently large amount of samples, the simulated distribution converges to the joint distribution that was being sought for. Because reaching the stationary distribution of the Markov chain can possibly be time consuming, it is common to ignore the first few Gibbs samples, in what is referred to as the *burn-in period*. Also, in order to avoid auto-correlation between successive samples, one may decide to generate the joint distribution by only keeping samples generated after a certain number of iterations in the sampler algorithm, in what is known as choosing the *step-size* of the Gibbs sampler. Determining a good burn-in period and step-size is a task difficult in itself.

For more on Gibbs samplers and MCMC algorithms see [3], [6] and [15].



### 3 Base Model

The model's purpose is to describe the underlying processes which generate the measured total electric load in individual households. Section 3.1 overviews the assumptions in the initial model, and section 3.2 describes the implementation details.

#### 3.1 Model assumptions

##### 3.1.1 Additive FHMM

The initial model used in the disaggregation algorithm is a FHMM as described in section 2.2.3. The observed variable  $\bar{Y}_t$ , the total power load, is characterized as a function of the individual emissions of the household's electrical appliances,  $Y_t^{(1)}, \dots, Y_t^{(m)}, \dots, Y_t^{(M)}$ , such that

$$\bar{Y}_t = f(Y_t^{(1)}, \dots, Y_t^{(m)}, \dots, Y_t^{(M)}), \quad (24)$$

for a household with  $M$  individual appliances. More specifically, the total power load is described as the sum of appliance emissions,

$$\bar{Y}_t = \sum_{m=1}^M Y_t^{(m)}, \quad (25)$$

and is as such referred to as an *additive* FHMM, see [12]. The underlying Markov models are assumed to each have the finite state space  $\mathbf{E} = \{0, 1\}$ , where 0 and 1 correspond to *OFF*- and *ON*-states respectively. Further, the processes are assumed time-homogenous, see (7). Then, as previously described in section 2.2, each appliance behaves as a Markov chain  $X_t^{(m)}$ , with transition matrix

$$\mathbf{P}^{(m)} = \begin{bmatrix} p_{0,0}^{(m)} & p_{0,1}^{(m)} \\ p_{1,0}^{(m)} & p_{1,1}^{(m)} \end{bmatrix}. \quad (26)$$

An example of the dependences in the initial model for a household with two underlying appliances is displayed in Figure 7 below.

##### 3.1.2 Gaussian emissions

The appliance emissions,  $Y^{(m)}$ , are the values of the individual energy load in appliance  $m$ . They are in the initial model assumed to be normally distributed depending on each possible state of the device. Since, each device is modelled as having only an *OFF* and *ON* state, the emission is described as

$$Y^{(m)} \sim \begin{cases} N(\mu_0^{(m)}, \sigma_0^{(m)}), & \text{if } X^{(m)} = \text{OFF} \\ N(\mu_1^{(m)}, \sigma_1^{(m)}), & \text{if } X^{(m)} = \text{ON} \end{cases}, \quad (27)$$

where  $\mu_0^{(m)}, \sigma_0^{(m)}, \mu_1^{(m)}, \sigma_1^{(m)}$ , are the means and standard deviations of the emissions for *OFF*- and *ON*-states respectively.

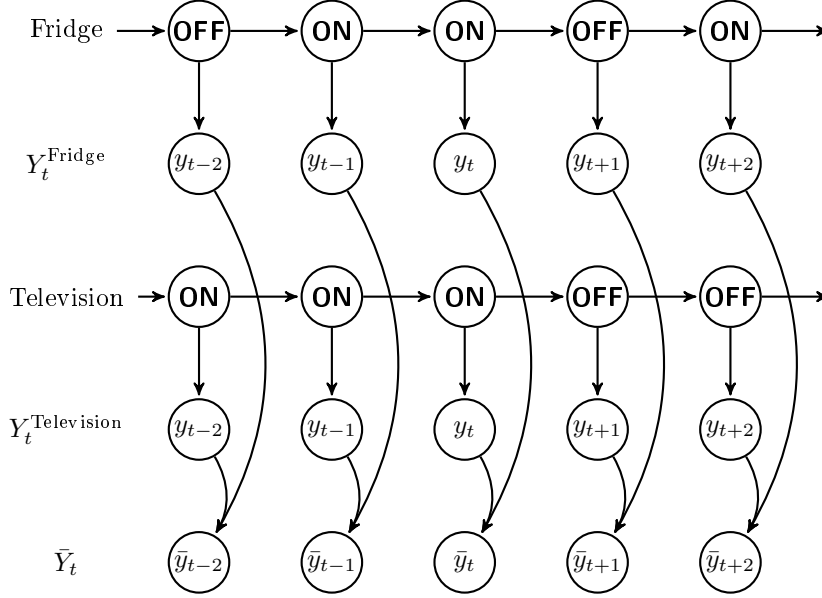


Figure 7: An example of the initial FHMM with two underlying appliances.

## 3.2 Approach

A supervised machine learning approach is used to tackle the disaggregation problem. First, a *training data set* of power load measurements over time from various appliances is used for fitting the FHMM. In essence, this means that

- (i) the power load measurements are categorized into a discrete set of states,
- (ii) transition matrices are obtained for every appliance, through simple counting of state changes,
- (iii) emission parameters of the different states are calculated, (means and variances).

For the base model, this training data set is composed of the daily supervised measurements for each family.

Second, this fitted model is used to disaggregate an aggregate power load signal from a *test data set*. This is done using a Gibbs sampling algorithm which aims to generate the most probable samples of state and emission sequences in order to describe the constituents of the aggregated power load signal, given our fitted model. Supervised machine learning approaches are discussed more in depth in [15].

### 3.2.1 State categorization

The original model approach categorizes each measurements as either an *OFF*- or *ON*-measurement, depending on a threshold criterion. The threshold criterion for an appliance  $m$ ,  $\theta^{(m)}$ , is defined as

$$\theta^{(m)} = \frac{1}{2n} \sum_{i=1}^n y_i^{+(m)}, \quad (28)$$

where  $y^+$  are all power load measurements that are greater than zero. The categorization of states is then given as follows,

$$X^{(m)} \sim \begin{cases} OFF, & \text{if } y^{(m)} \leq \theta^{(m)} \\ ON, & \text{if } y^{(m)} > \theta^{(m)}. \end{cases} \quad (29)$$

An example of this employed categorization is displayed in Figure 8. All appliance distributions, including a table of the empirical counts are displayed in appendix D.

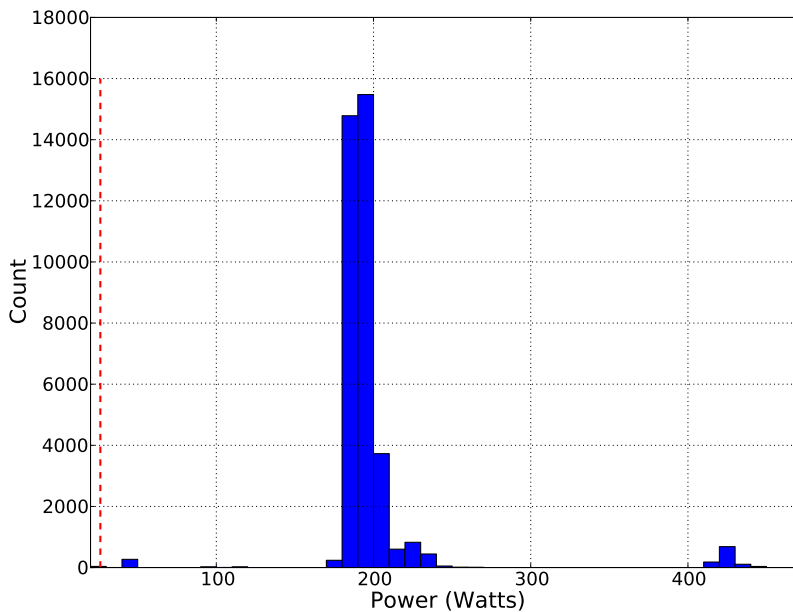


Figure 8: Non-zero power measurements (Watts) for one family’s refrigerator. Bin size: 10 W. The dotted line indicates the threshold value,  $\theta$ , that categorizes the power measurements as either *ON* or *OFF*.

### 3.2.2 Gibbs sampling with the fitted FHMM

The Gibbs sampling method produces samples by successively sampling from the conditional distributions of each individual random variables that constitute the model, given all other random variables in the joint probability distribution, see section 2.3. In the FHMM case, the produced samples are simulations from the joint probability distribution of all state variables, given the total emission, for each appliance and time step,

$$P(\mathbf{X}|\bar{\mathbf{Y}}), \quad (30)$$

where

$$\begin{aligned} \mathbf{X} = & X_0^{(1)}, X_1^{(1)}, \dots, X_{T-2}^{(1)}, X_{T-1}^{(1)}, \\ & X_0^{(2)}, X_1^{(2)}, \dots, X_{T-2}^{(2)}, X_{T-1}^{(2)}, \\ & \dots, \\ & X_0^{(M-1)}, X_1^{(M-1)}, \dots, X_{T-2}^{(M-1)}, X_{T-1}^{(M-1)}, \\ & X_0^{(M)}, X_1^{(M)}, \dots, X_{T-2}^{(M)}, X_{T-1}^{(M)}, \end{aligned}$$

is the set of state variables, and

$$\bar{\mathbf{Y}} = \bar{Y}_0, \bar{Y}_1, \dots, \bar{Y}_{T-2}, \bar{Y}_{T-1},$$

is the set of total emissions, (the household's power load measurements).

Thus, for every sub-iteration in the Gibbs sampling algorithm, each state variable  $X_t^{(m)}$  is updated with a sample from the conditional probability distribution,

$$P(X_t^{(m)} | \mathbf{X} \setminus X_t^{(m)}, \bar{\mathbf{Y}}). \quad (31)$$

At each time step  $t$ , any state variable  $X_t^{(m)}$  is only conditionally dependent of its immediate previous and following state,  $X_{t-1}^{(m)}$  and  $X_{t+1}^{(m)}$ , due to the Markov chain property. Also, a state  $X_t^{(m)}$  is conditionally dependent of its emission  $Y_t^{(m)}$ , and of the measured total emission  $\bar{Y}_t$ , because of the respectively direct and indirect causal connections that exist in the underlying chain structure  $X_t^{(m)} \rightarrow Y_t^{(m)} \rightarrow \bar{Y}_t$ , see section A.1.1. Further, because all emissions  $Y_t^{(m)}$  sum together to the total emission  $\bar{Y}_t$ , these variables form a so called v-structure, see appendix A.1.2. This leads to all  $Y_t^{(m)}$  at a time step to be conditionally dependent, given  $\bar{Y}_t$ .

Finally, for this reason, not given the individual emission chains  $\mathbf{Y}$ , a state  $X_t^{(m)}$  is not conditionally independent of the other state variables at  $t$ , given  $\bar{\mathbf{Y}}$ .

From this reasoning, it can now be seen that these variables are all part of what is referred to as the *Markov blanket* of  $X_t^{(m)}$ .

**Definition 11.** *The **Markov blanket** of a node  $X$  is defined as the set of nodes that render  $X$  conditionally independent of all the other nodes in the graph. [10]*

The dependence structure is highlighted in Figure 9.

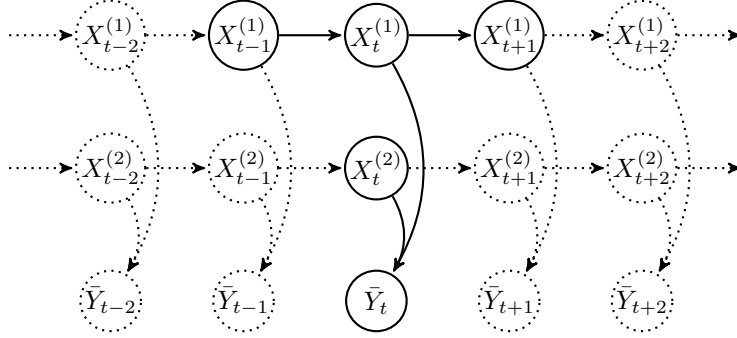


Figure 9: Dependence in a FHMM graphical model with two appliances, when sampling the conditional distribution (31) for state  $X_t^{(1)}$ . Dependence relations are illustrated with solid lines.

Following the above reasoning, the conditional probability in (31) simplifies to evaluating the conditional probability of  $X_t^{(m)}$  given only its Markov blanket,

$$P(X_t^{(m)} | X_{t-1}^{(m)}, X_{t+1}^{(m)}, X_t^{(1)}, \dots, X_t^{(m-1)}, X_t^{(m+1)}, \dots, X_t^{(M)}, \bar{Y}_t), \quad (32)$$

and with  $\mathbf{V} = X_t^{(1)}, \dots, X_t^{(m-1)}, X_t^{(m+1)}, \dots, X_t^{(M)}$ ,

$$P(X_t^{(m)} | X_{t-1}^{(m)}, X_{t+1}^{(m)}, \mathbf{V}, \bar{Y}_t). \quad (33)$$

It can be shown (see appendix A.2.1), that this conditional probability may be factorized as

$$\text{Constant} \times P(X_{t+1}^{(m)} | X_t^{(m)}) P(X_t^{(m)} | X_{t-1}^{(m)}) P(\bar{Y}_t | X_t^{(m)}, \mathbf{V}). \quad (34)$$

This factorization expresses the original conditional distribution (31) as a product of the conditional transition probabilities, and the conditional emission probabilities, which are easily sampled from the prior distributions in the fitted model.

### 3.2.3 User-defined aspects of the Gibbs algorithm

Parts of the Gibbs sampler algorithm are user-defined, and/or specific to our model:

- (i) The initial values are set by sampling the state variables from uniform probability distributions over the state space.
- (ii) In each sub-iteration, the state variables are updated for one appliance at a time, in order of increasing time. The chosen order of appliances is randomized.

### 3.2.4 Gibbs sampling results

The collection of obtained Gibbs samples forms simulated distributions of an individual appliance's emission  $Y_t^{(m)}$  at each time step. Ultimately, the estimated appliance emissions  $\hat{y}_t$  are calculated as the expected value of this distribution,

$$\hat{y}_t^{(m)} = E[Y_t^{(m)}]. \quad (35)$$

This method corresponds to taking the expected Gibbs sample, and is a common practice in Gibbs sampling, see [3] and [6].

### 3.2.5 Add-one smoothing

In cases where there is no data in the training data set, for a certain appliance state, (such as an appliance never being turned on for example), the resulting fitted model will never have the ability to predict this state transition, should it occur in the test data set. This is known as the *zero count problem* or the *sparse data problem*, and it is analogous as an illustration of the philosophical problem of induction [15].

In order to counter this limitation of the model's state space, a commonly practised technique known as *add-one smoothing* is applied. This solution adds one to all empirical counts in the model, thereby fitting the model to enable predictions of all possible samples in the state space.

For a model with state space  $\mathbf{E} = \{OFF, ON\}$  the  $ON$  transition probabilities, being fitted with the count of  $ON \rightarrow OFF$  transitions  $N_1$  and the  $ON \rightarrow ON$  transitions  $N_2$ , as

$$p_{ON,OFF} = \frac{N_1}{N_1 + N_2}, \quad p_{ON,ON} = \frac{N_2}{N_1 + N_2}, \quad (36)$$

become, with add-one smoothing,

$$p_{ON,OFF} = \frac{N_1 + 1}{N_1 + N_2 + 2}, \quad p_{ON,ON} = \frac{N_2 + 1}{N_1 + N_2 + 2}. \quad (37)$$

It is not difficult to see that if  $N_1$  is 0, and  $N_2$  is large,

$$p_{ON,OFF} \rightarrow 0, \quad p_{ON,ON} \rightarrow 1, \quad (38)$$

which demonstrates that the effect of add-one smoothing diminishes for a large training data set.

## 4 Method

This section accounts for the implementations that were performed during the course of this work. In subsection 4.1, the implementations which are the basis for this thesis are presented. Section 4.2 reviews the preparatory analysis of state durations in the data sets, and section 4.3 describes the algorithmic modifications on the original model.

### 4.1 Model extension

The major weakness of conventional HMMs is the modelling of state duration [16]. The inherent duration,  $l$ , of a state  $i$  given the base model, is geometrically distributed with probability mass function

$$P_i(l) = (p_{i,i})^{l-1}(1 - p_{i,i}). \quad (39)$$

As mentioned in the section 1, this state duration density could in some cases be deemed inappropriate to model real life appliance data. Instead, it is believed that the state duration of an everyday household appliance can be modelled according to some unknown distribution. As such, a semi-Markov model is proposed as a replacement for the underlying Markov model in the base model, and the new model is subsequently referred to as a factorial hidden semi-Markov model (FHSMM).

The claim is that semi-Markov models outperform Markov models in terms of representation of the true underlying state sequences for household appliances. In order to model these changes, we introduce an extra layer of nodes (state variables), as well as some changes in the notation of the base model.

For a FHSMM, we let the individual emissions (24) represent emissions from the semi-Markov process,  $\{Q_t; t \in \mathbb{N}_0\}$ , such that

$$\begin{aligned} \mathbf{Q} = & Q_0^{(1)}, Q_1^{(1)}, \dots, Q_{T-2}^{(1)}, Q_{T-1}^{(1)}, \\ & Q_0^{(2)}, Q_1^{(2)}, \dots, Q_{T-2}^{(2)}, Q_{T-1}^{(2)}, \\ & \dots, \\ & Q_0^{(M-1)}, Q_1^{(M-1)}, \dots, Q_{T-2}^{(M-1)}, Q_{T-1}^{(M-1)}, \\ & Q_0^{(M)}, Q_1^{(M)}, \dots, Q_{T-2}^{(M)}, Q_{T-1}^{(M)}, \end{aligned}$$

is the set that represent all of the underlying states that decide the total emission process,  $\{\tilde{Y}_t; t \in \mathbb{N}_0\}$ . As seen in section 2.2.4, these states are described as emissions from the generalized state processes,  $\{G_n^{(m)}; n \in \mathbb{N}_0\}$ , with  $X_n^{(m)}$  henceforth representing the state variable in the generalized state process for appliance  $m$ , see equations (14) and (15).

#### 4.1.1 Explicit state duration

Using the semi-Markov extension of the FHMM, the aim is to model the occurrence of state transitions by using appropriate duration densities instead of the inherent duration distribution. A generative approach for specifying explicit state durations was described in [16], as follows:

1. Start with an initial state  $Q_0 = q$ , from a specified initial state distribution.
2. Sample a duration  $l$  from this state's duration density  $P_q(L)$ .
3. Generate the observation sequence  $Y_0, Y_1 \dots, Y_{l-1}$ , from the joint observation density  $P(Y_0, Y_1 \dots, Y_{l-1} | Q_0 = q)$ . Because of the conditional independence in HMMs described in section 2.2.2, we may sample each of these observations  $Y_t$  separately, from  $P(Y_t | Q_t = q)$ .
4. Last, a new state  $Q_l$  is sampled from a transition matrix, corrected for no chances of self-transitions,  $p_{q,q} = 0, \forall x$ . Steps 2-4 are repeated a desired amount of times.

As this method is designed toward generating a chain of semi-Markov emissions, it is not compatible per default with the Gibbs sampling method described in section 3.2.2. As described in section 2.3, the Gibbs sampling method should aim to generate samples from an equilibrium distribution of state sequences that corresponds to a given sequence of total emissions  $\bar{\mathbf{Y}}$ . In this case, the equilibrium distribution is believed to be a semi-Markov chain.

This approach has been applied in the algorithm proposed below, and is described more in depth in section 4.3.

1. Given a set of total observations  $\bar{\mathbf{Y}}$ , start with an initial sample from the conditional distribution  $P(\mathbf{Q} | \bar{\mathbf{Y}})$
2. Begin a Gibbs sampling process, where each iteration samples a state  $q_t$  at time index  $t$ , by sampling a corresponding generalized state  $G_i$  from the conditional distribution of generalized states at index  $t$ ,  $P(G_i | \mathbf{Q} \setminus q_t, \bar{\mathbf{Y}})$ . The state  $q_i$  here corresponds to the state variable of the generalized state  $G_i$ .

In the calculation of the conditional distribution  $P(G_i | \mathbf{Q} \setminus q_i, \mathbf{Y}, \bar{\mathbf{Y}})$ , the semi-Markov property prohibits the underlying states from self-transitioning, similarly to step 4 in the first method described above, see section 4.3.2 for further detail.

## 4.2 State duration pre-analysis

A previous study by Hyungsul et al. [9] advocates the idea that geometric durations are unsuitable to model *ON*-state durations, and considers the gamma distribution  $\Gamma(k, \theta)$ , with density function

$$f(x) = \frac{1}{\Gamma(k)} x^{k-1} \frac{1}{\theta^k} e^{-x/\theta}, \quad x > 0, \quad (40)$$

for a duration  $x$ , as a better model. This report further suggests that *OFF*-durations are better modelled as bimodal distributions, due to most electrical appliances often being off for a longer period during night time.

The data set was pre-processed (see appendix B) and split into periods of maximum 24 hours for each separate household. Thereafter, distributions of *ON*- and *OFF*-durations were plotted for a range of the most common of appliances in the various data sets. Considering the possibility that some duration data overlaps between periods, the duration data at the beginning and end of



each period was discarded (referred to as *censoring* the observations at the boundaries in [8]). As such, the obtained distributions are believed to accurately represent state durations that occur within the 24 hour periods, and will be referred to as "daily" data.

By discarding the beginning and end duration data, one also avoids the problem of fitting the distributions with state durations that are of the length of a full day. State durations with length of a full day or more are possible occurrences in the data sets, and was observed to be due to two causes: Either from missing data in the data set, or due to some appliances that are turned on very rarely, and therefore have extreme *OFF*-durations, spanning up to multiple days. It should be noted that this second type of duration is not possible to model, as a result of the data pre-processing and subsequent choice in period length.

The daily duration distributions are displayed in appendix E.

#### 4.2.1 Fitting duration distributions

Similarly to [9], distributions were fitted from the data using maximum likelihood estimation (MLE).

In order to decrease the impact of extreme tail values on the ML estimation, the distribution was cut off at the 95:th percentile and all higher measurements were discarded. The effect of this is illustrated in figure 10 below.

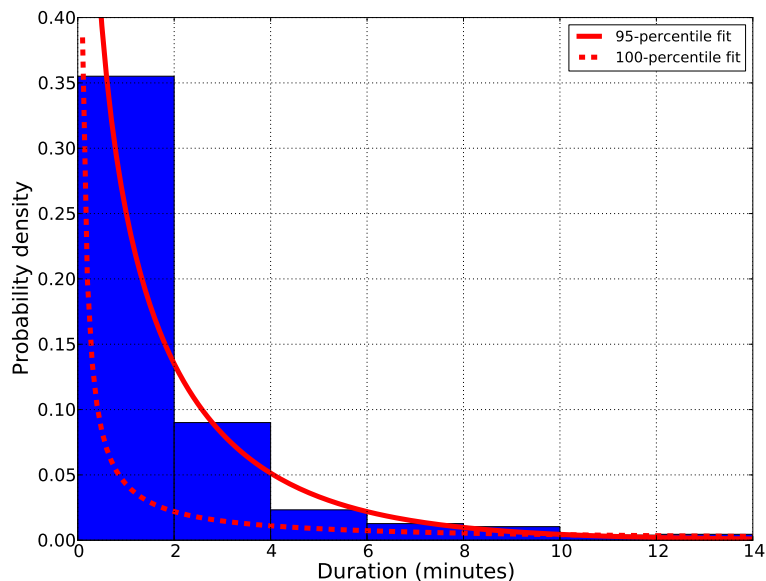


Figure 10: Probability densities from ML estimations pre-analysis. The solid red curve is the density function for the duration distribution fitted up to the 95th percentile. The dotted red curve is the gamma density function when fitting the same appliance with all duration data. The appliance fitted is the microwave 1 from REDD.

For comparison, the ML exponential distribution,  $Exp(\lambda)$ , with probability density function

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad x > 0, \quad (41)$$

for a duration  $x$ , was also estimated, as this distribution is the continuous analogue of the inherent geometrical distribution. The fitness of these distributions on the data is compared using the log-likelihood ratio (LLR),

$$LLR = \log \frac{\max_{k,\theta} P(\text{durations} | \text{Gamma}(k, \theta))}{\max_{\lambda} P(\text{durations} | \text{Exp}(\lambda))}. \quad (42)$$

Some estimated ML parameters, as well as LLR results are shown in Table 1, for the full datasets see Tables 4-5 in appendix C. The amount of samples used to estimate the ML distribution is also displayed. As can be seen, this sample amount varies greatly between some appliances, a result of the preprocessing discussed in the beginning of this section. For the MLE, a higher sample count of the empirical distribution warrants a better fit to the true underlying distribution, and therefore gives a better estimate of the underlying distribution parameters displayed in the tables.

As can be seen, many appliance durations with high sample count have large, positive LLR values, which then motivates the gamma distribution as a choice of explicit state duration distribution. An example of state distribution histograms with fitted gamma distribution is displayed in Figure 11.

For results presented in section 5, only appliances with high sample count and high LLR was chosen to evaluate the FHSMM. The selected appliances are the ones displayed in Table 1. Their histograms with fitted gamma distributions are displayed in appendix E.

<b>ON-duration distributions</b>						
Label	Households	Samples	$\lambda$	$k$	$\theta$	LLR
microwave 1	4	3848	0.574	0.602	2.896	3.27E+03
stove 1	4	3537	0.572	0.510	3.431	3.10E+03
refrigerator 1	5	2164	0.059	3.301	5.139	6.08E+05
lighting 1	6	1180	0.156	0.496	12.951	4.33E+04
lighting 3	4	854	0.182	0.152	36.211	2.20E+04
furance 1	3	656	0.100	0.219	45.550	6.17E+04

(a)

<b>OFF-duration distributions</b>						
Label	Households	Samples	$\lambda$	$k$	$\theta$	LLR
microwave 1	4	3809	0.284	0.233	15.081	3.39E+04
stove 1	4	3544	0.556	0.410	4.389	3.40E+03
refrigerator 1	5	2145	0.033	4.119	7.333	1.94E+06
lighting 1	6	1189	0.084	0.352	33.644	1.61E+05
lighting 3	4	808	0.082	0.116	105.631	1.16E+05
furance 1	3	661	0.101	0.182	54.217	6.07E+04

(b)

Table 1: REDD selection of appliance duration distributions ML parameters, sorted by total number of samples.

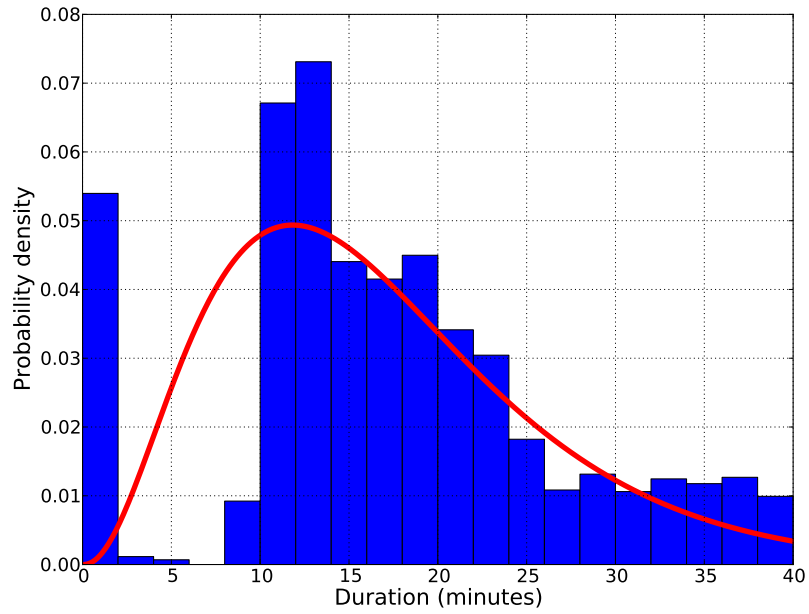
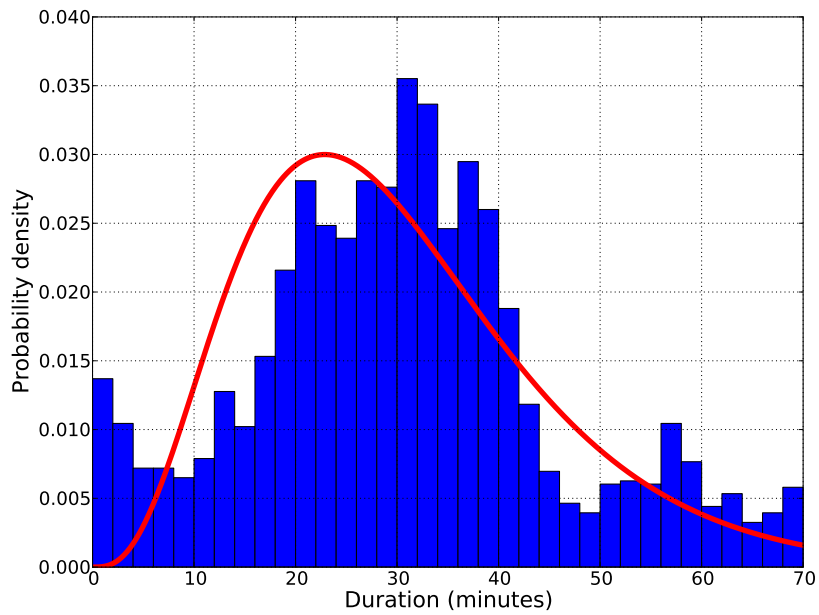
(a) Refrigerator, *ON*-state distribution.(b) *OFF*-state distribution.

Figure 11: Refrigerator duration times in minutes, REDD dataset.

### 4.3 Algorithm

Below is a brief overview of the steps in the algorithm used to account for state durations.

1. Find the posteriori distributions of state durations in each appliance.
2. Compute the conditional probabilities, for each appliance's state node in the network, from the probability densities of the distributions fitted in step 1.
3. Sample from this conditional distribution.

Step 1 has already been discussed in section 4.2.1, and so this section goes into the details of step 2 and 3.

#### 4.3.1 Quantization of prediction probabilities

Because the model consists of discrete Markov chains, and the fitted duration distributions are continuous, the duration probability densities were quantized into a discrete sets of probability masses.

$$\underbrace{P_q(L=l)}_{\text{discrete}} = \underbrace{P_q(l-\delta \leq l_1 < l+\delta)}_{\text{continuous}}, \quad (43)$$

for state  $q$ , where  $\delta$  is a chosen bin size, selected as half the frequency of power measurements. All duration probabilities in section 4.3 are calculated using this assumption.

#### 4.3.2 Computing conditional probabilities

In the FHSMM case, the produced samples are simulations from the joint probability distribution of all state variables, given the total emission, for each appliance and time step,

$$P(\mathbf{Q}|\bar{\mathbf{Y}}), \quad (44)$$

where, for every sub-iteration in the Gibbs sampling algorithm, each state variable  $Q_t^{(m)}$  is updated with a sample from the conditional probability distribution,

$$P(Q_t^{(m)}|\mathbf{Q}\setminus Q_t^{(m)}, \bar{\mathbf{Y}}). \quad (45)$$

With  $\mathbf{G}^{(m)} = \{G_0^{(m)}, G_1^{(m)}, \dots\}$ ,  $\mathbf{G} = \{\mathbf{G}^{(m)} : m = 1, \dots, M\}$ , and  $\mathbf{G}^{-\mathbf{m}} = \mathbf{G}\setminus\mathbf{G}^{(m)}$ , we may see that expression (45) is equivalent to

$$P(Q_t^{(m)}|\mathbf{G}^{-\mathbf{m}}, \mathbf{G}^{(m)}\setminus Q_t^{(m)}, \bar{\mathbf{Y}}). \quad (46)$$

Similarly to the FHMM, the state block  $G_n$  at time  $t$  of an appliance is not conditionally independent of the individual emission, nor of the other appliances state blocks at time  $t$ , given the total emission,  $\bar{Y}_t$ . However, given the semi-Markov property, the state variable  $Q_t^{(m)}$  will not be dependent solely of the immediate previous state, but of the whole previous state block.

Following the above reasoning, the conditional probability in (45) simplifies to evaluating the conditional probability of  $Q_t^{(m)}$  given the state block directly before it,  $A$ , the state block directly after it,  $B$ , the total emission,  $\bar{Y}_t$ , and all other appliance's state blocks occurring at time  $t$ , referred to as  $\mathbf{V}$ . The simplified expression is displayed below, as

$$P(Q_t^{(m)}|A, B, \mathbf{V}, \bar{Y}_t). \quad (47)$$

It can again be shown (see appendix A.2.2), that this conditional probability may be factorized as

$$\text{Constant} \times P(Q_t^{(m)}|A, B)P(\bar{Y}_t|Q_t^{(m)}, \mathbf{V}), \quad (48)$$

which expresses the original conditional distribution (31) as a product of

1. the conditional total emission probability given the sampled state  $Q_t^{(m)}$  and all other state blocks at time  $t$ ,
2. the conditional probability of the state  $Q_t^{(m)}$  given the previous and next state blocks.

For the first factor, we make the assumption that the individual emissions  $Y_t^{(m)}$  are only dependent of the state at time  $t$ . This in turn makes the total emissions,  $\bar{Y}_t$ , independent of previous and next state,

$$P(\bar{Y}_t|Q_t^{(m)}, \mathbf{V}) = P(\bar{Y}_t|Q_t^{(1)}, Q_t^{(2)}, \dots, Q_t^{(m)}, \dots, Q_t^{(M-1)}, Q_t^{(M)}), \quad (49)$$

and this conditional probability is in fact the same as the conditional emission probability of the base model, in (34). The problem reduces to evaluating the second factor;  $P(Q_t^{(m)}|A, B)$ .

It can easily be seen that this evaluation is heavily dependent of the configuration of states in the block sequence  $(A, Q_t^{(m)}, B)$ . As illustrated in Figure 12, sampling different outcomes of  $Q_t^{(m)}$  gives rise to different configurations of state sequences.

There are four distinct cases of state sequences,

$$(A, Q_t^{(m)}, B) = (\dots, q_{t-1}, q_t, q_{t+1}, \dots),$$

three of them which are displayed in Figures 12b-12d:

1.  $\underbrace{q_{t-i} = \dots = q_{t-1}}_A \neq q_t \neq \underbrace{q_{t+1} = \dots = q_{t+j}}_B$ ,
2.  $\underbrace{q_{t-i} = \dots = q_{t-1}}_A = q_t \neq \underbrace{q_{t+1} = \dots = q_{t+j}}_B$ ,
3.  $\underbrace{q_{t-i} = \dots = q_{t-1}}_A \neq q_t = \underbrace{q_{t+1} = \dots = q_{t+j}}_B$ ,
4.  $\underbrace{q_{t-i} = \dots = q_{t-1}}_A = q_t = \underbrace{q_{t+1} = \dots = q_{t+j}}_B$ ,

for some  $i, j \in \mathbb{Z}$ .

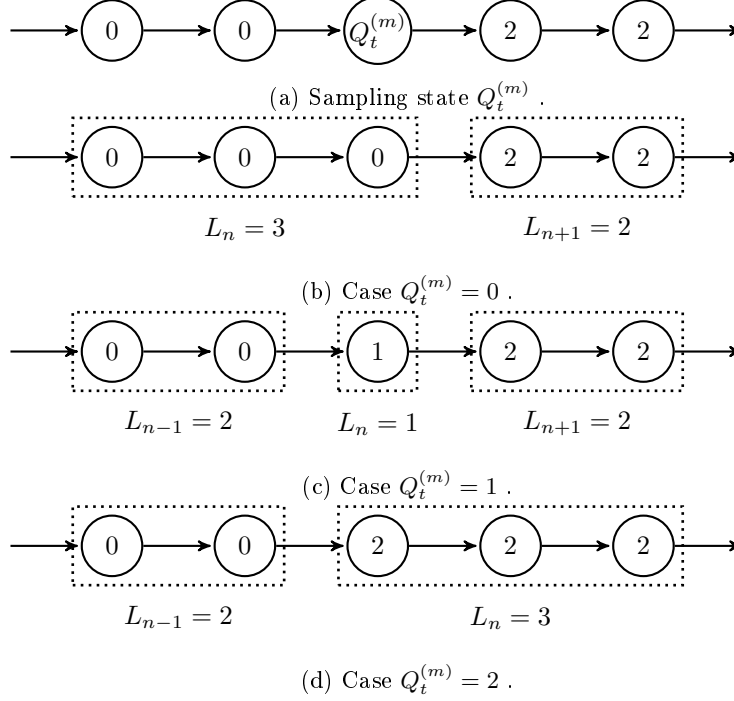


Figure 12: Example of segment configurations during sampling, for a model with state space  $\mathbf{E} = \{0, 1, 2\}$ .

These four cases are evaluated separately in accordance with section 2.2.4:

- If  $q_{t-1} \neq q_t$ , then there is a Markov transition from state  $q_{t-1}$  to state  $q_t$ . These transition probabilities are easily determined in the same manner as the base model.
- If  $q_t = q_{t+1} = \dots = q_{t+l-1} = q_{t+l}$ , (with  $q_t = x$ ), then this sequence of states originates from a common state block ( $X = x, L = l + 1$ ). The probability of this state block,

$$P(X = x, L = l + 1) = P_x(L = l + 1), \quad (50)$$

is obtained from a pre-fitted state duration distribution.

With transition probabilities denoted as in (8), and state duration probabilities denoted as in (50), the conditional probability  $P(Q_t^{(m)} = q|A, B)$ , for

$$\begin{aligned} A &= (X = a, L = i), \\ B &= (X = b, L = j), \end{aligned}$$

may be evaluated in each one of the separate cases, as

1.  $P_a(L = i) \cdot p_{a,q} \cdot P_q(L = 1) \cdot p_{q,b} \cdot P_b(L = j)$ ,

2.  $P_a(L = i + 1) \cdot p_{a,b} \cdot P_b(L = j)$ ,
3.  $P_a(L = i) \cdot p_{a,b} \cdot P_b(L = j + 1)$ ,
4.  $P_a(L = i + j + 1)$ .

### 4.3.3 Conditional probability at chain edges

As the sample FHSMM is produced over a finite interval  $[1, T]$ , for each daily data (see section 4.2), we refer to the times  $t = 1$  and  $t = T$  as the *edges* of the sample FHSMM. For the block sequences, as described in the previous section, there will always be some state blocks beginning, and some ending, at these edges. In such cases, an *edge block*  $G$ , with observed state and duration ( $X = x, L = l$ ) is not observed beyond the edge, and it cannot be determined whether its duration exceeds  $L = l$ . Therefore, its duration probability is calculated as

$$P_x(L \geq l), \quad (51)$$

which translates as, since the durations are discrete,

$$P_x(L \geq l) = 1 - P_x(L < l) \quad (52)$$

$$= 1 - \sum_{j=1}^{l-1} P_x(L = j). \quad (53)$$

## 5 Results

This section presents the performed experiments in section 5.1, along with their respective results. These are ordered by first presenting binary classification scores in section 5.2, then showing a REDD-score comparison in section 5.3, then prediction convergence results are displayed in section 5.4. Finally, section 5.5 displays results of how the models compared in appliance wise energy signal disaggregation.

### 5.1 Experiment

The experiment follows the procedure described in section 3.2. The FHMM and FHSMM were fitted using a training data set consisting of all data from five of the six households in REDD. This fitted model was then used to disaggregate aggregated data from the test data set, which consisted of all data from the single household not included in the training data set. The results were cross-validated, by repeating the above procedure while using different combinations of the six households.

The appliances used were taken as the first  $n$  appliances from the following set:  $\{\textit{lighting 1, refrigerator 1, microwave 1, stove 1, lighting 3, furnace 1}\}$ . The appliances in this set were chosen in accordance with section 4.2.1.

The experiment was subject to two types of parameter variations:

1. Generating 25, 50, 75 and 100 Gibbs samples, for a model fitted with 6 appliances.
2. Fitting the model with 1-6 appliances, for 100 Gibbs samples.

### 5.2 Binary classification test

Using binary classification, as described in [15], *ON*-states were categorized as positives, and *OFF*-states were categorized as negatives. The prediction of each state node in the sample was compared with the true state in the test data set.

This comparison has four types of outcomes illustrated in a so called *contingency table*, Table 2, below.

		True state	
		<i>ON</i>	<i>OFF</i>
Predicted state	<i>ON</i>	True positive (TP)	False positive (FP)
	<i>OFF</i>	False negative (FN)	True negative (TN)

Table 2: Contingency table of classification comparisons.

From this table, an appliance predicted as *OFF*, although it in reality was *ON*, would be a False negative (FN). The statistical measures ultimately evaluated were *Accuracy*, *Precision*, *Recall* and *F1 score*.



The accuracy of the classification is the proportion of correct classifications to all classifications in the contingency table. It is calculated as

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}. \quad (54)$$

The precision is the proportion of correctly identified *ON*-states to the total amount of classified *ON*-states. It is calculated as

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (55)$$

The recall is the proportion of correctly identified *ON*-states to the total amount of true existing *ON*-states. It is calculated as

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (56)$$

The F-score is the harmonic mean of precision and recall. It is a widely used measure, that measures the *ON*-state classification rate with respect to both recall and precision. It is calculated as

$$\text{F-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (57)$$

The classification was performed for all daily data in all households, after the experiments presented in section 5.1, and the resulting scores have been linearly interpolated and are displayed in Figures 13-14.

### 5.3 REDD-score

A performance measure used by Kolter and Johnsson [11] in their work on the REDD is a measure of the total energy correctly assigned. It will in this thesis be referred to as the REDD-score, and is defined as

$$\text{REDD-score} = 1 - \frac{\sum_{t=1}^T \sum_{m=1}^M |\hat{y}_t^{(m)} - y_t^{(m)}|}{2 \sum_{t=1}^T \sum_{m=1}^M y_t^{(m)}}. \quad (58)$$

Here,  $\hat{y}_t^{(m)}$  denotes the algorithm's prediction, as described in section 3.2.4.

This measure was evaluated for the experiments in section 5.1, and the results are displayed in Figure 15.

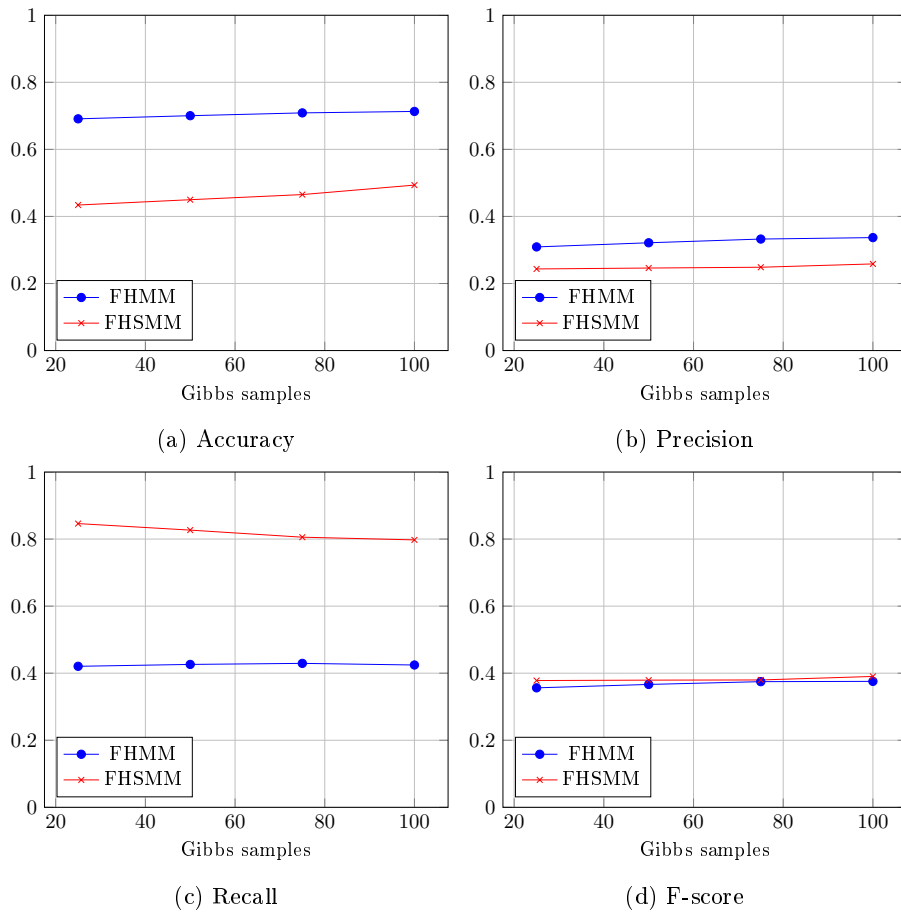


Figure 13: Binary classification scores with varying amount of Gibbs samples. REDD, 6 appliances.

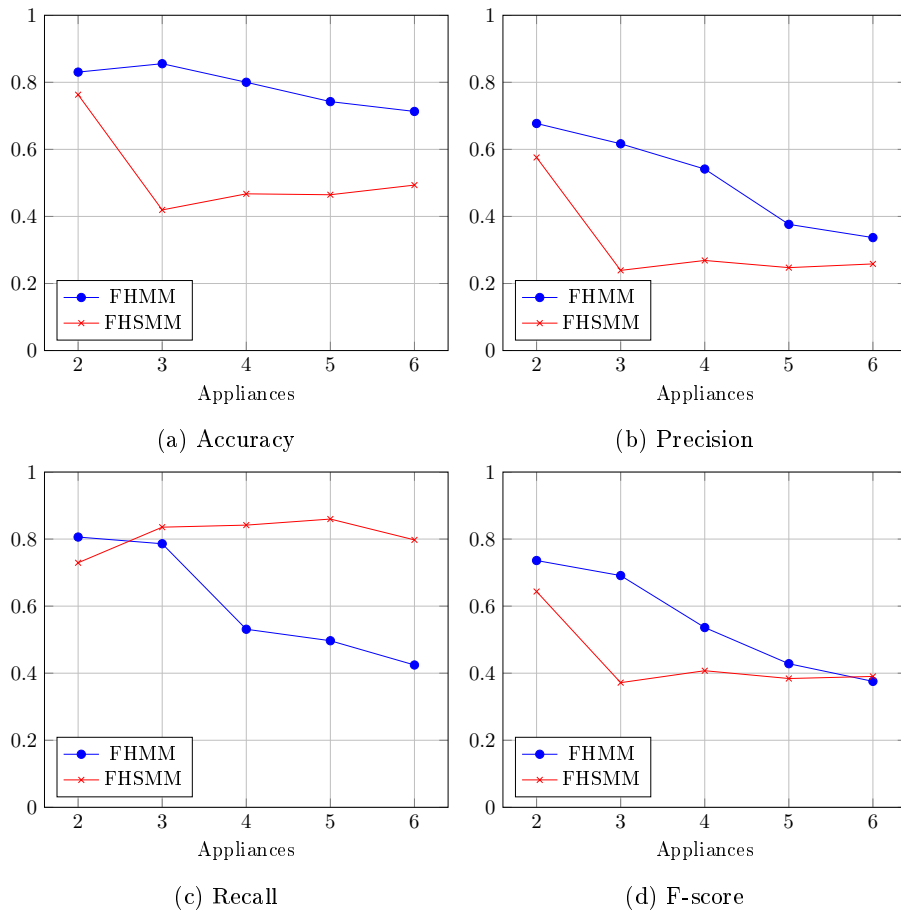


Figure 14: Binary classification results with varying number of appliances. REDD, 100 Gibbs samples.

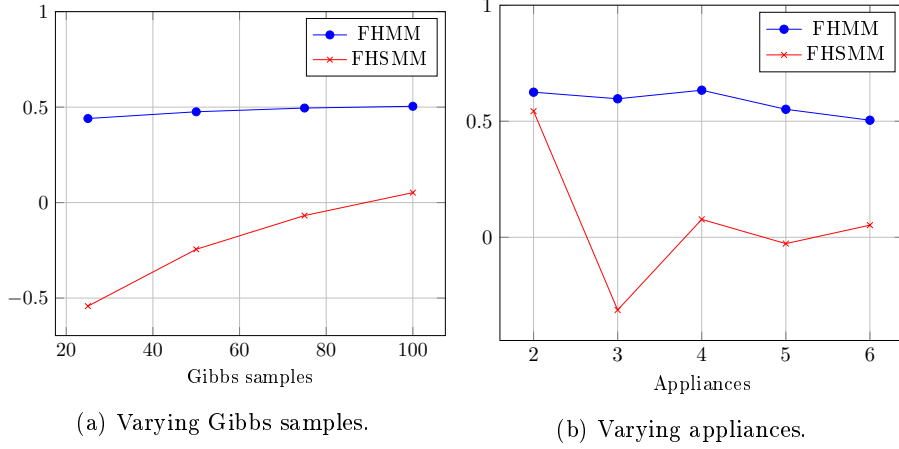


Figure 15: REDD score results.

#### 5.4 Model convergence

To evaluate how well the Gibbs sampling algorithm converged, the log-probability of each subsequent Gibbs sample was calculated. For the FHMM, this was done by calculating the log-probability of all sampled observations, together with the log-probability of all sampled underlying state transitions, by using the values from the fitted model.

For an appliance  $m$ , by letting  $\alpha_k^{(m)}$  represent the transition probability from  $t = k$  to  $t = k + 1$ , and  $\beta_k^{(m)}$  represent the emission probability at time  $t = k$ , the log-probability of the FHMM can be easily represented as

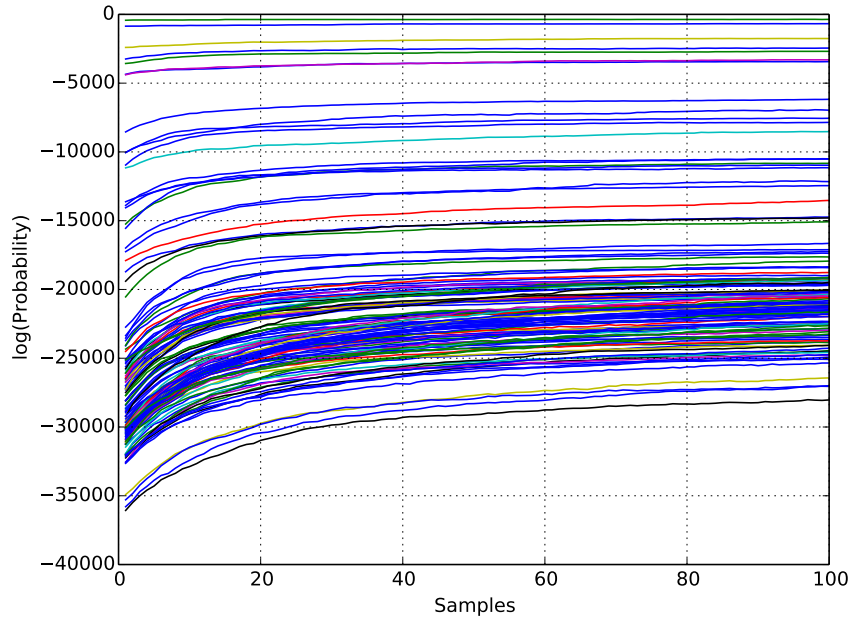
$$\log(P(\text{sample})) = \sum_{m=1}^M \left( \sum_{t=1}^{T-1} \log(\alpha_t^{(m)}) + \sum_{t=1}^T \log(\beta_t^{(m)}) \right), \quad (59)$$

for  $M$  fitted appliances, and sample interval  $[1, T]$ .

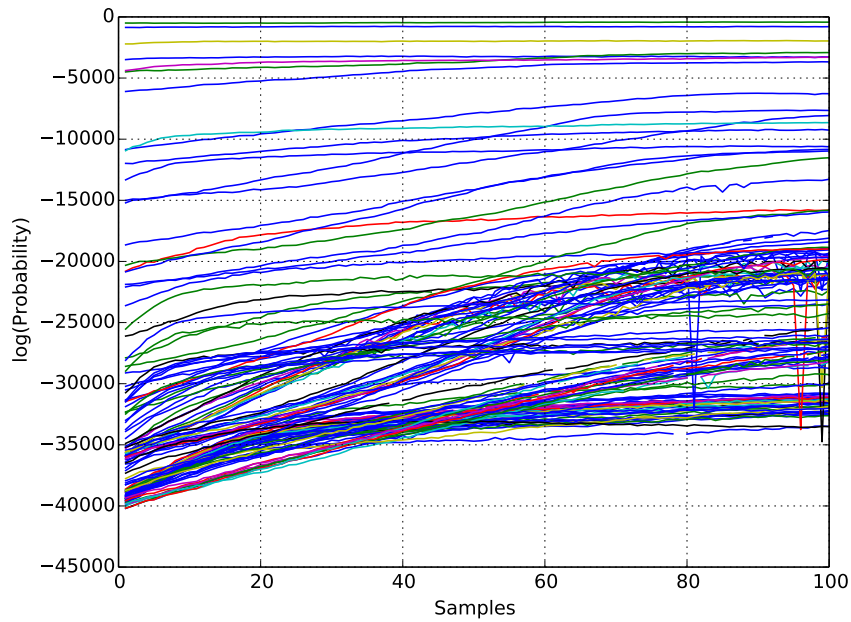
Using the same notation as above, with the difference that  $\alpha_k$  now refers to the underlying state chain of a semi-Markov chain (see section 2.2.4), and introducing  $\gamma_k^{(m)}$  to represent the duration probability of the generalized state at the  $k$ :th step of the underlying state chain, the log-probability of the FHSMM is represented as

$$\log(P(\text{sample})) = \sum_{m=1}^M \left( \sum_{n=1}^{N-1} \log(\alpha_n^{(m)}) + \sum_{t=1}^T \log(\beta_t^{(m)}) + \sum_{n=1}^N \log(\gamma_n^{(m)}) \right), \quad (60)$$

for  $N$  underlying generalized states. The log-probabilities were determined for 100 subsequent Gibbs samples, for all daily data in all households. The results are displayed in Figure 16. Each plotted line in the figure represents the log-probability for a sampled day in one of the six households.



(a) FHMM.



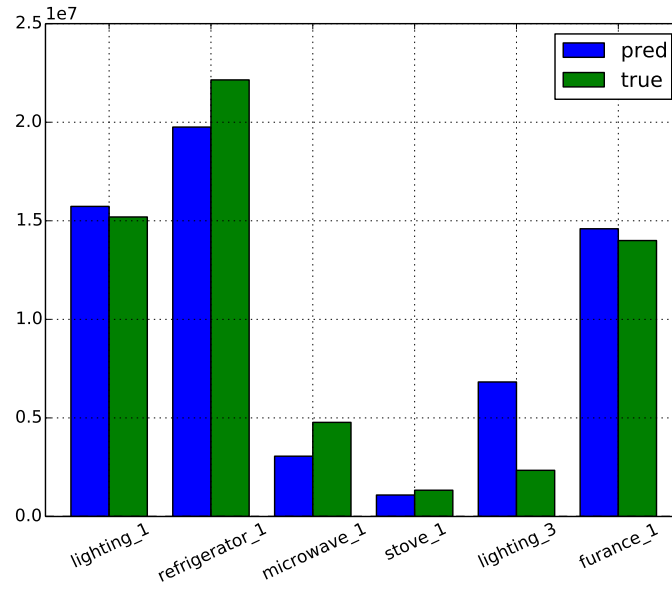
(b) FHSMM.

Figure 16: Log-probabilities for increasing amount of Gibbs samples. Each line represents the log-probability of a household's sampled day.

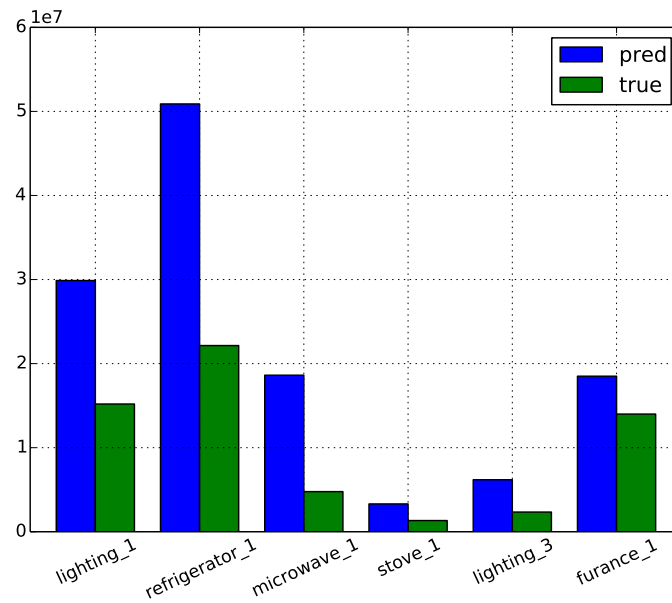
## 5.5 Energy disaggregation results

### 5.5.1 Total energy use comparison

The total per-appliance energy usage, for all days and households in the dataset, was estimated using 100 Gibbs samples. The results are shown in Figures 17-18.

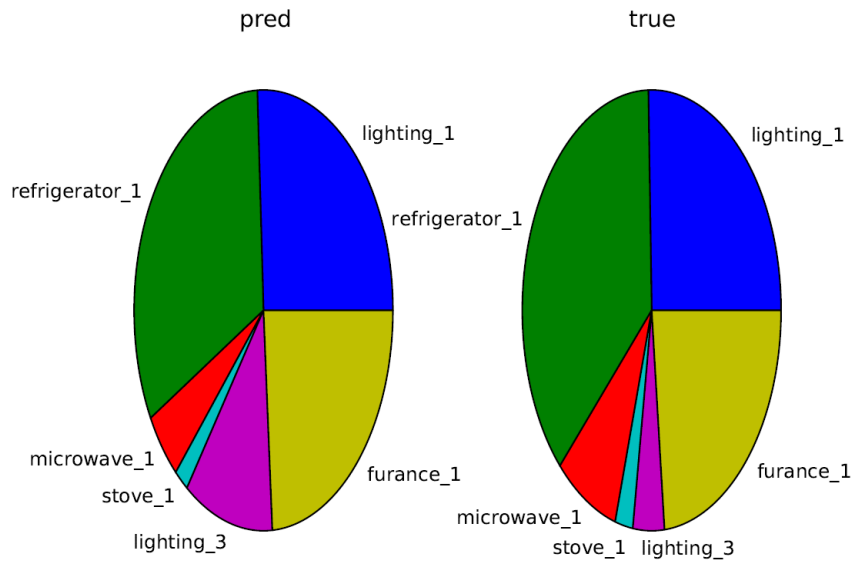


(a) FHMM

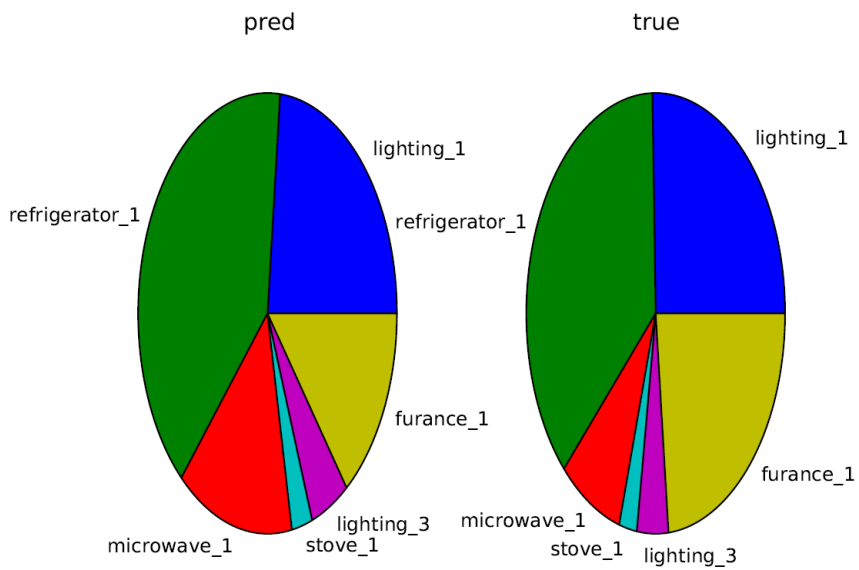


(b) FHSMM

Figure 17: Estimated and true total energy usage per appliance, for all days and households in the dataset.



(a) FHMM

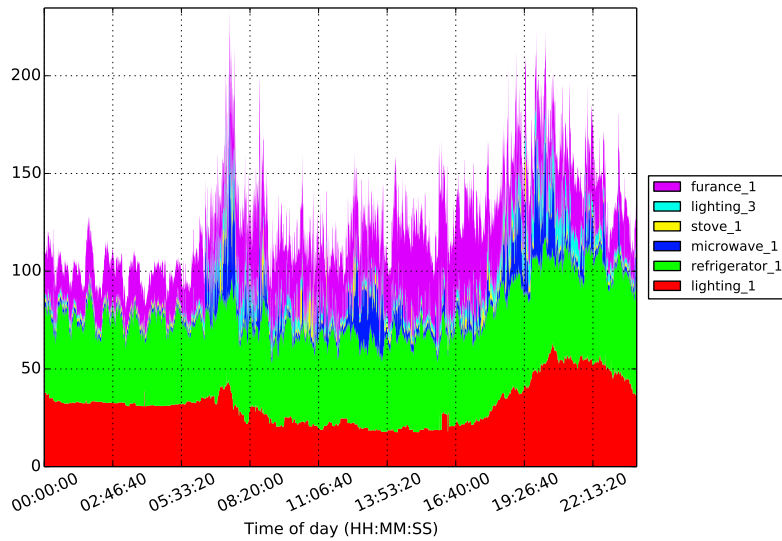


(b) FHSMM

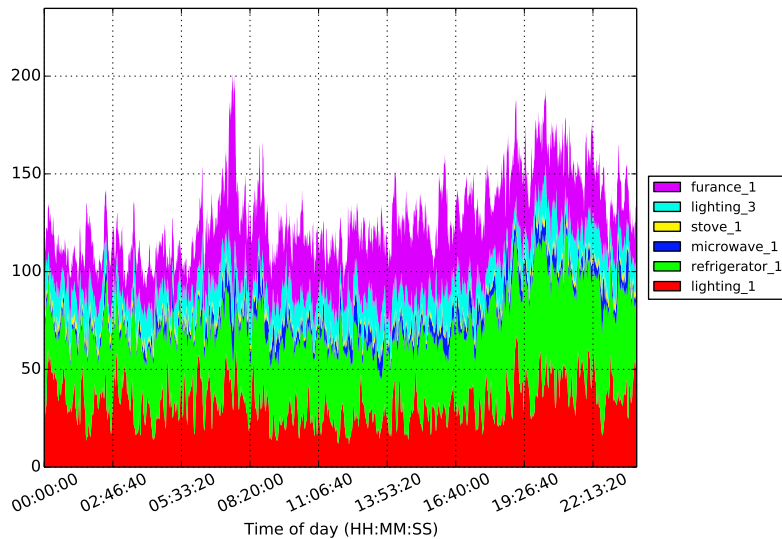
Figure 18: Estimated and true total energy usage per appliance, for all days and households in the dataset.

### 5.5.2 Averaged energy usage time series

The total per-appliance energy usage was estimated using 100 Gibbs samples, for all days and households in the dataset. The mean results at every time of day, for all produced daily data predictions was then computed, in order to produce the dataset average energy time series. The results are displayed in stack plots, which when summed over the vertical axis produce the total averaged energy signal. The results are displayed in Figures 19-20.



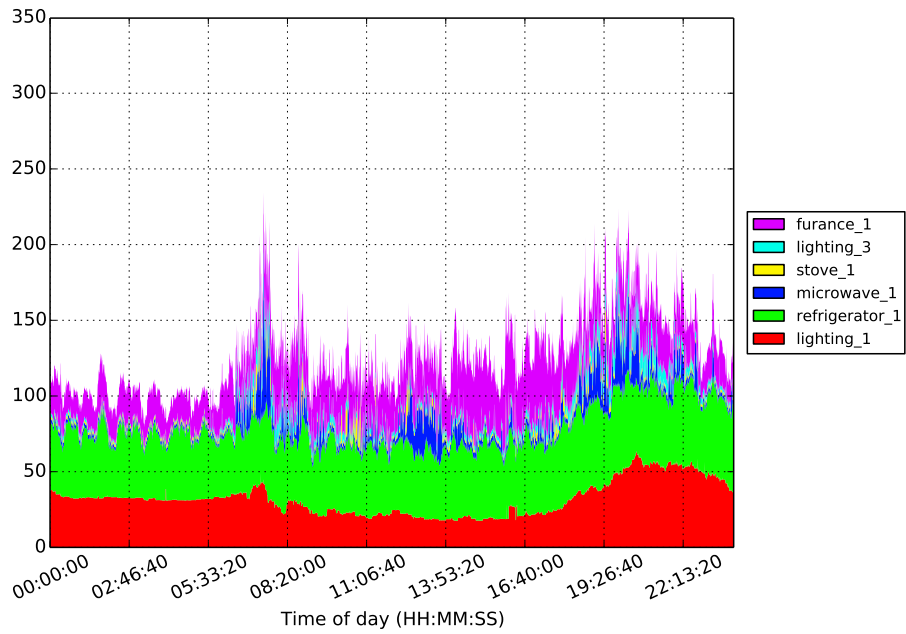
(a) Dataset (true averaged signal).



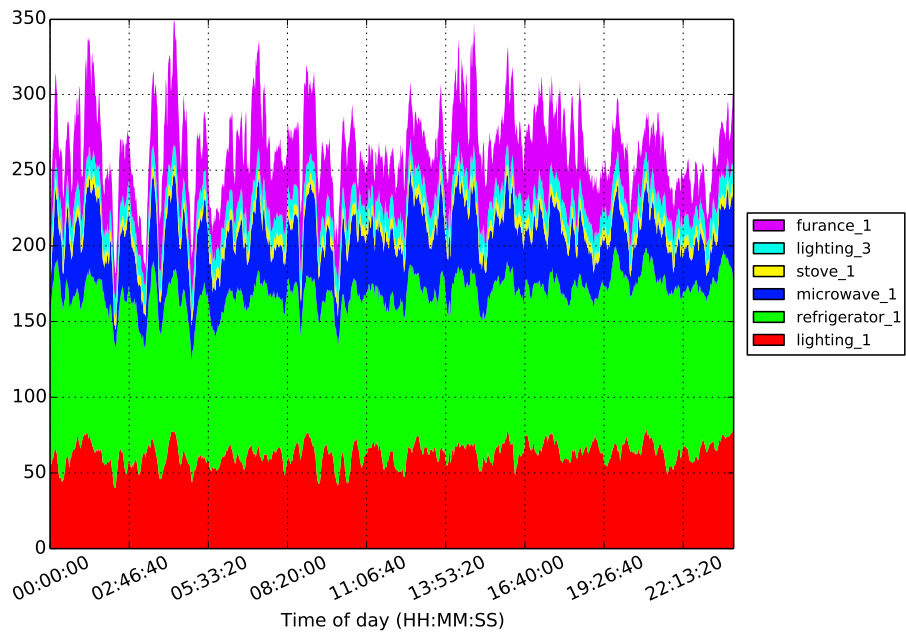
(b) FHMM prediction.

Figure 19: Averaged time series, over all days and households. Displayed as a stack plot, where each energy signal is vertically stacked to produce the total power load.





(a) Dataset (true averaged signal).



(b) FHSMM prediction.

Figure 20: Averaged time series, over all days and households. Displayed as a stack plot, where each appliance signal is vertically stacked to produce the total power load.

### 5.5.3 Model behaviour at a single data point

The FHMM and FHSMM were used to estimate the component signals for the daily data of one single day from one household. This was done using 100 Gibbs samples, with two appliances.

The total per-appliance energy usage estimations are displayed in Figure 21. The appliance wise energy signals are displayed in Figures 22-23. Finally, the daily data stack plot time series are displayed in Figures 24-25.

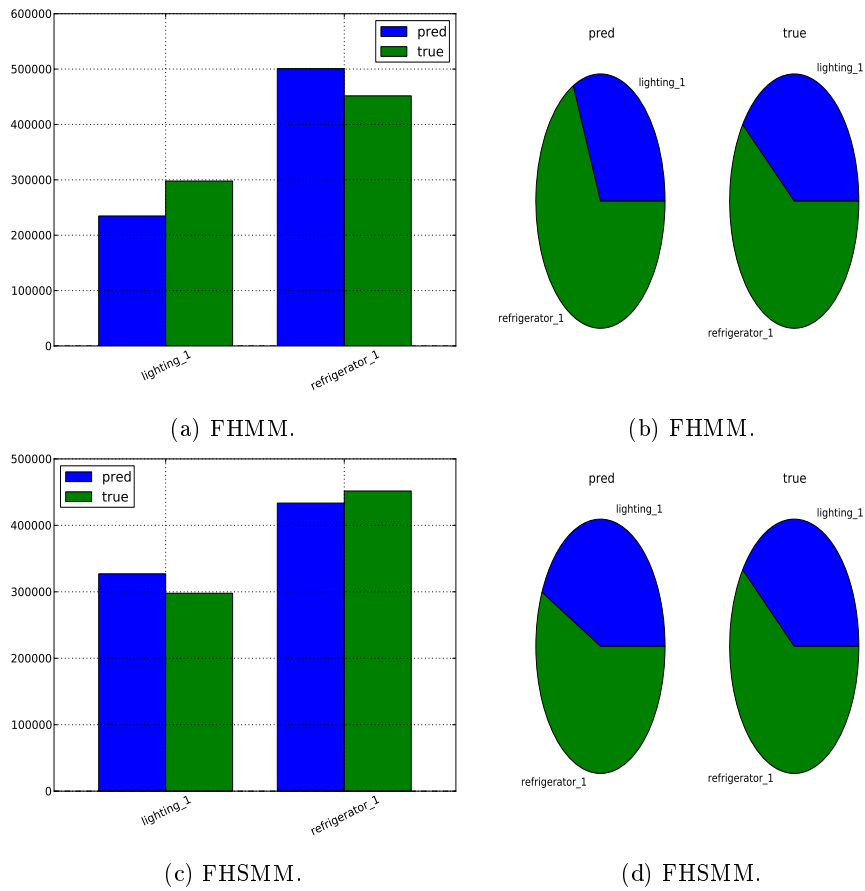
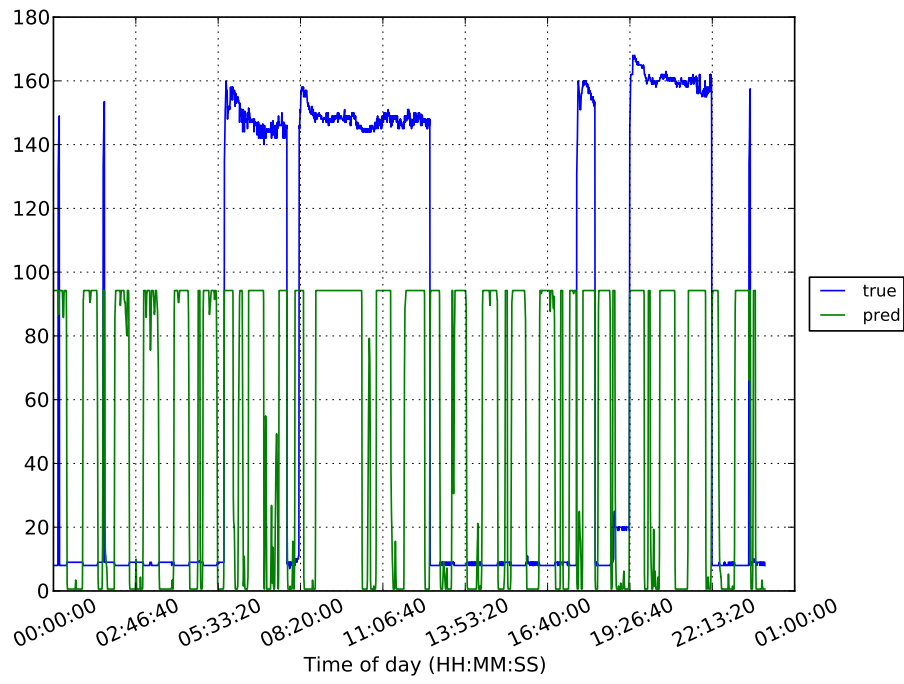
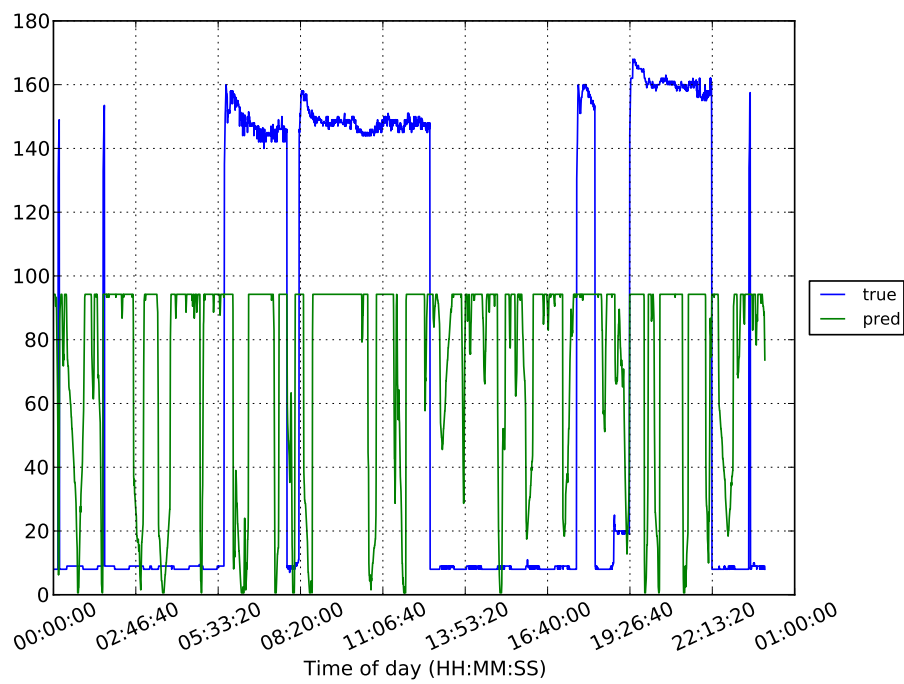


Figure 21: Estimated and true total energy usage per appliance, one household, one day.

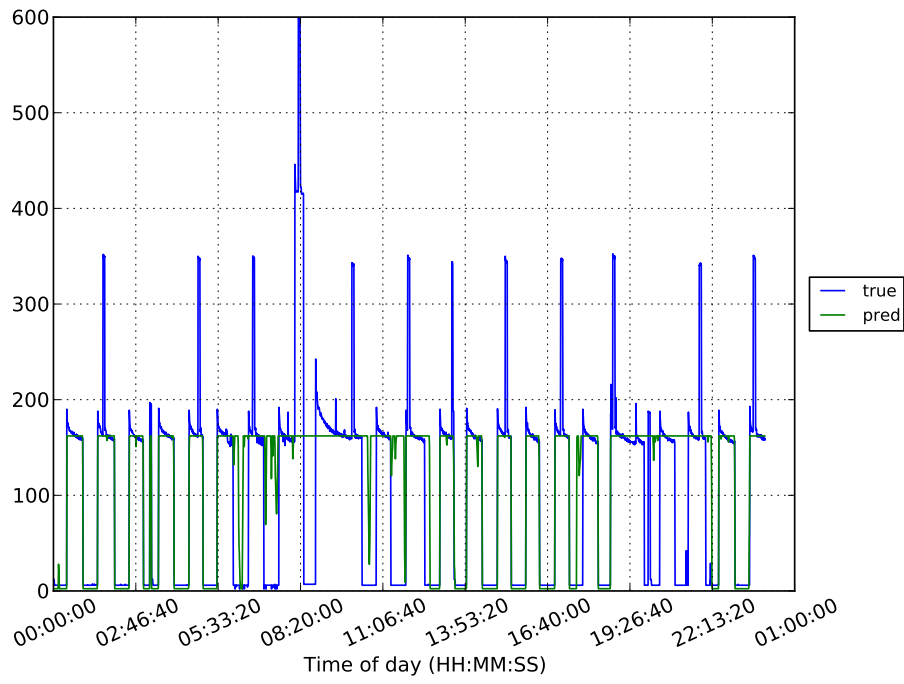


(a) FHMM.

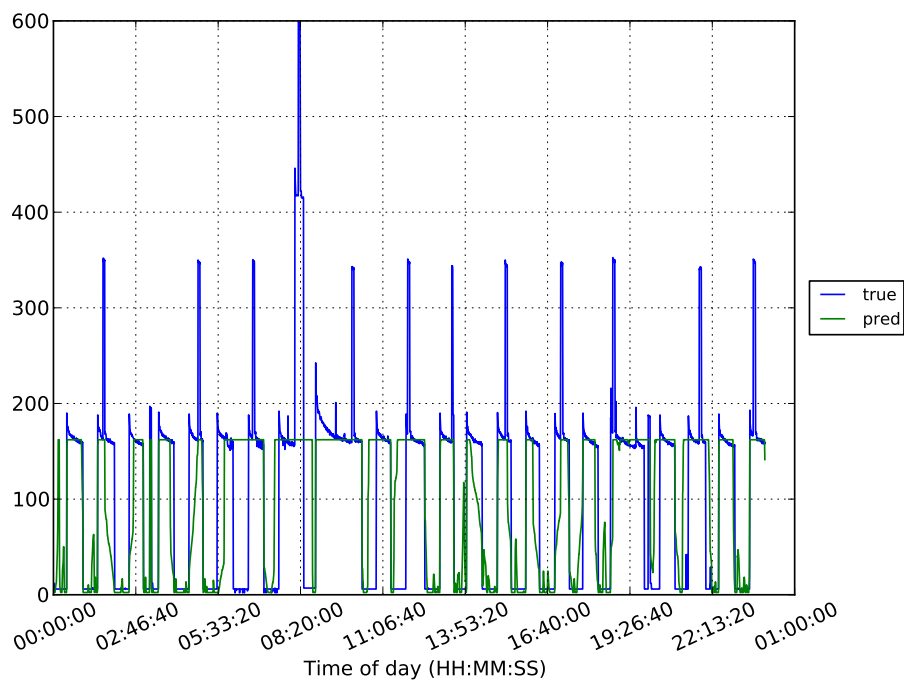


(b) FHSMM.

Figure 22: Time series comparison, lighting 1, one household, one day.

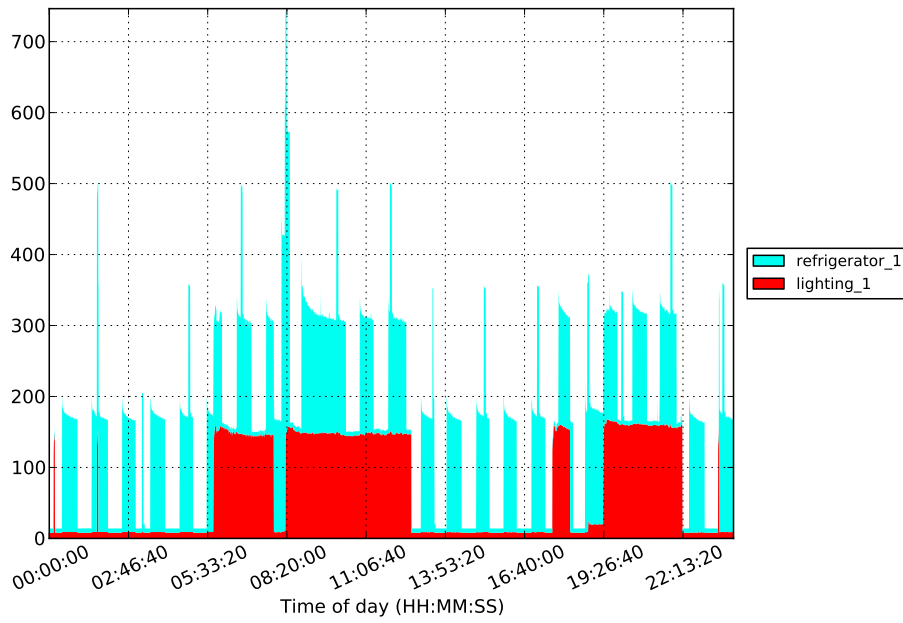


(a) FHMM.

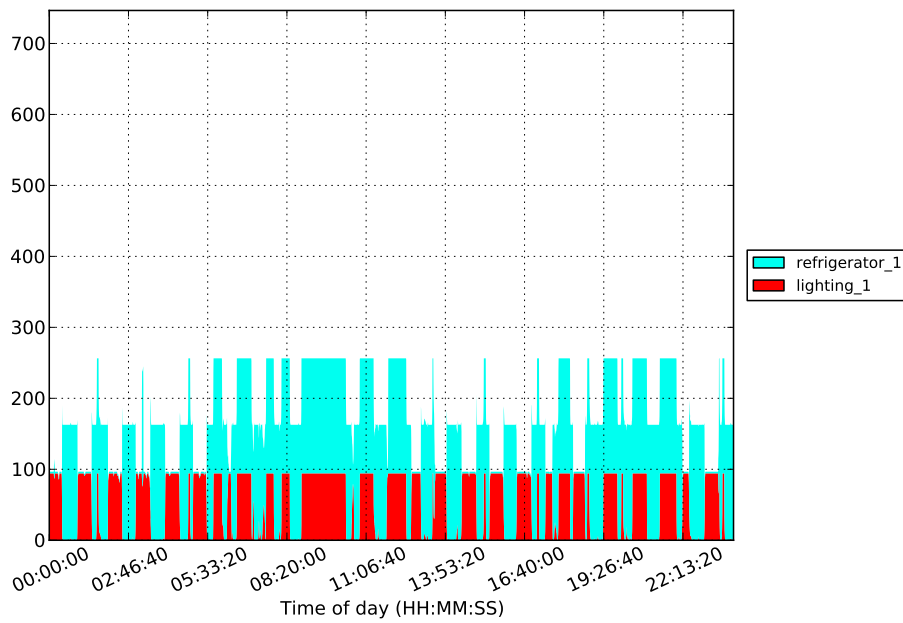


(b) FHSM.

Figure 23: Time series comparison, refrigerator 1, one household, one day.

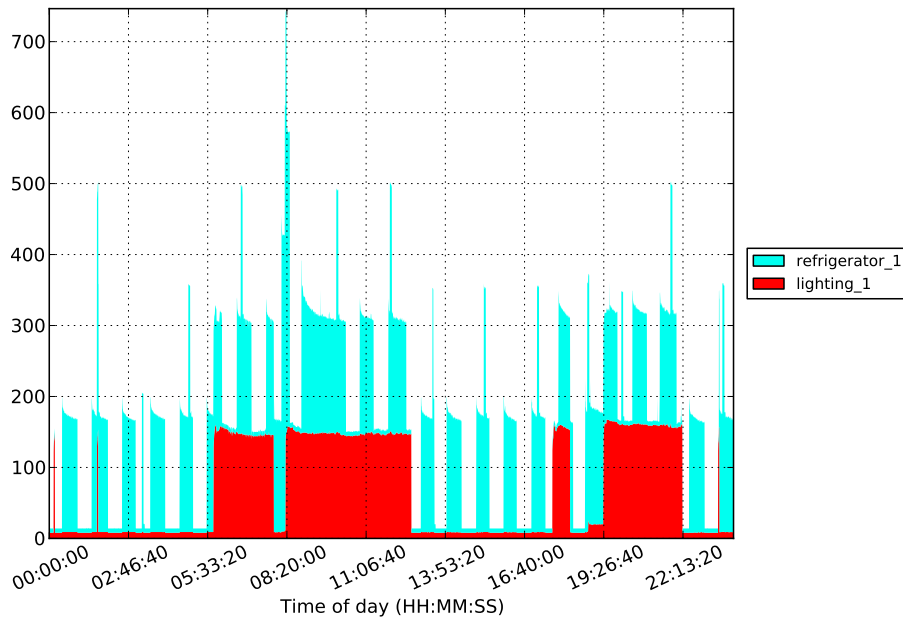


(a) Dataset (true averaged signal).

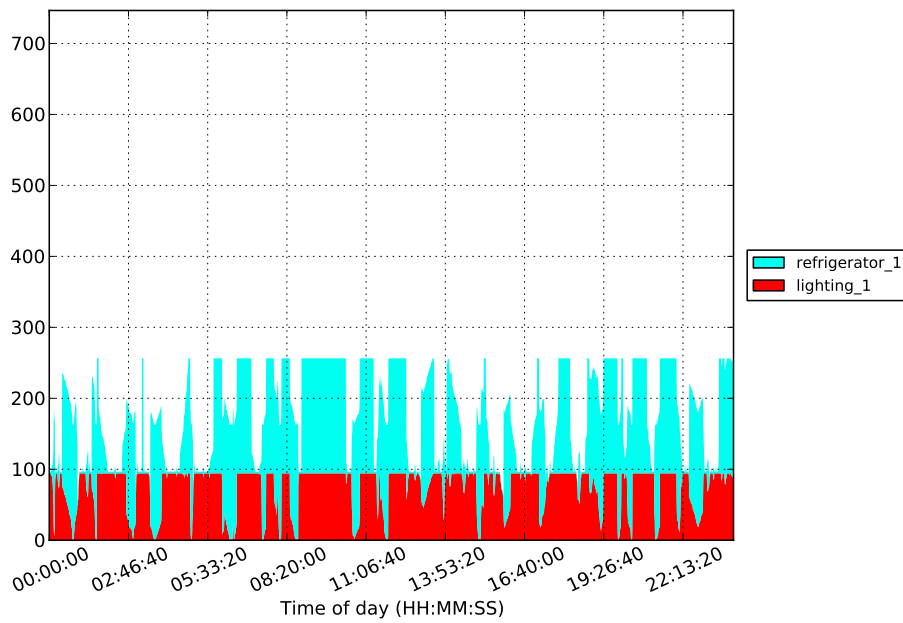


(b) FHMM prediction.

Figure 24: Time series, one household, one day. Displayed as a stack plot, where each appliance signal is vertically stacked to produce the total power load.



(a) Dataset (true averaged signal).



(b) FHSMM prediction.

Figure 25: Time series, one household, one day. Displayed as a stack plot, where each appliance signal is vertically stacked to produce the total power load.

## 6 Conclusions

This section overviews the results from section 5 in sections 6.1-6.5. Sections 6.6 and 6.7 provides the final conclusions, and suggestions for further work in the future.

### 6.1 Binary classifications

The binary classification results in Figure 13, for all families and all dates, displayed little to no increase in score when increasing the amount of Gibbs samples for both models. The FHMM performed better in all scores but the F-score. The hopes were that the amount scores would increase with the amount of Gibbs samples, but the results were somewhat inconclusive.

For an increasing amount of appliances in Figure 14, the disaggregation task would seem more complex, and indeed the FHMM performances exhibited decreasing trends. The FHSMM displayed a big performance drop in accuracy, precision and F-score, when disaggregating  $\{\textit{lighting 1, refrigerator 1, microwave 1}\}$  compared to just disaggregating  $\{\textit{lighting 1, refrigerator 1}\}$ , but on the other hand increased its recall. Since recall was the only score to increase, this implies from the definitions in section 5.2 that more states were categorized as *ON*. One could possibly infer that the FHSMM had specific problems with categorizing the microwave's signal. Overall, the binary classification results would discourage the choice of FHSMM as a replacement for the base model.

### 6.2 REDD score analysis

Increasing the amount of Gibbs samples seemed favourable to the FHSMM, when it came to improving the total energy correctly assigned. Also, similarly to what was observed in section 6.1, the score was found to drop suddenly when adding the microwave to the aggregated power signal. Again, in this analysis the FHMM showed superior results to the FHSMM.

### 6.3 Convergence analysis

The log-probability of both models during the sampling process exhibited increasing trends, as would be expected, over the first 100 Gibbs samples. Three major differences can be seen in Figure 16.

1. The FHSMM generally starts off with with lower log-probabilities.
2. Some household days converge much faster than with the FHMM.
3. The FHSMM convergence lines are more "shaky", suggesting that each new sample is more prone to greater differences from the previous sample.

The first point can be attributed as an effect of the uniform distribution used to start the Gibbs sampler, see section 3.2.3.

It remains unclear whether or not the FHSMM would have performed better or worse than the FHMM if more than 100 Gibbs samples had been generated. The results for less than 100 Gibbs samples would suggest the FHMM as a better model choice, seeing as its bulk of obtained log-probabilities was generally less negative than the FHSMM.

## 6.4 Total energy disaggregation results

The energy disaggregation results were the most conclusive, in regard to the purpose of this thesis. These results displayed the estimated signals in relationship to the true signals. The resulting comparison provided a more intuitive feel for how well the models performed.

The number of correctly predicted states as in section 6.1, although interesting, is not necessarily linked to the rejection or acceptance of the model in a commercial setting. From this view, it is much more important to correctly classify the largest amounts of actual energy usage. The bar plots in Figure 17, show that the FHMM performed better than the FHSMM on the whole dataset, from an energy usage point of view. The estimations of energy usage for the most power-hungry appliance, the refrigerator, were more than twice the true value.

This last result was also observed in the averaged time series in Figures 19-20. In these figures, the FHSMM estimated the average power signal to often more than twice of its actual value.

## 6.5 Comparison of model behaviour

For the estimations of energy signals in a single day, it could be seen that both models performed poorly at exact disaggregation of lighting, but much better when disaggregating the refrigerator. For lighting, it was clear that the models' predicted *ON*-state emissions were much lower than the true *ON*-state emissions for this particular day.

One possible reason for this is that the emissions for lighting 1 are not clustered around two states *ON* and *OFF*, but more, as can be seen in Figure 38 in appendix D. As there is limited information about the REDD, it was not possible to know what kind of appliances were measured, and as such it is possible that lighting 1 is in fact a collection of several appliances (lamps and other light bulbs). In such a case, our model is clearly limited as it is constructed to handle appliances with no more than two states.

Another possible explanation is that since the estimate is the expected value of all Gibbs samples, see section 3.2.4, it could be that this result is an average including many individual "bad" sample estimates.

## 6.6 Accepting or rejecting the implemented model

It became clear from the experiments performed that the FHMM was the more able model for disaggregating the chosen appliances. However, the results remained unsatisfactory, in that they seemed to great extent be heavily dependent on the chosen appliances and preprocessing of the data.

Because the FHSMM discarded data at the edges of the daily data, section 4.2, the FHMM was able to use a larger training set than the FHSMM, something which should then have given the FHMM an advantage in its predictions.

Further, it has not been clarified if the Gibbs sampling method was carried out long enough for it to converge toward the equilibrium distribution. Handbooks on Gibbs sampling suggest tens of thousands of Gibbs samples [3]; amounts that were unfortunately not feasible given the current algorithm. Also,



a burn-in of 10 samples was used, together with a step size of 1, given the analysis in section 6.3, experiments with larger burn-in could have been more optimal for the FHSMM.

In the end, the FHSMM was rejected as a result of the performed experiments.

### 6.7 Further work

Section 6.6 touches on many unanswered questions by the implemented algorithm.

- The existence of appliances with more than two states, makes it important to enhance the model's abilities to handle multiple states.
- The Gibbs sampling method should be tested with parameter variations, such as varying burn-in and step size, as well as larger sampling amounts.
- Evidently, more and larger data sets would also be useful to shed light on the possible shortcomings and benefits of the model.
- Because this thesis has dealt with explicit state durations, an interesting future extension would be to evaluate other probability distributions than the gamma distribution to model the state durations.
- The work of Johnson and Willsky on Bayesian Nonparametric Hidden Semi-Markov Models [8] has been presented as successful in the implementation of state durations. Future work would center on researching their applied algorithm.

## A Proofs and calculations

### A.1 Conditional dependence in some graphical models

#### A.1.1 Chain structure

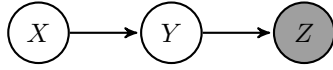


Figure 26: Chain structure  $X \rightarrow Y \rightarrow Z$ . The shaded node represents the observed variable.

Given a *chain-structured* graphical model [15]  $X \rightarrow Y \rightarrow Z$  as displayed in Figure 26, the joint probability distribution is, (using the chain rule for probabilities),

$$P(X, Y, Z) = P(X)P(Y|X)P(Z|Y), \quad (61)$$

or likewise,

$$P(X, Y, Z) = P(Z)P(Y|Z)P(X|Y), \quad (62)$$

which illustrates that indirect connections also imply a dependence between nodes, if the in between lying nodes are not observed, and as such if  $Y \not\perp Z$  and  $X \not\perp Y$ , then

$$X \not\perp Z. \quad (63)$$

#### A.1.2 V-structure

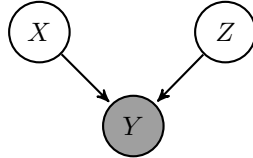


Figure 27: V-structure  $X \rightarrow Y \leftarrow Z$ . The shaded node represents the variable conditioned upon.

Given a *v-structured* graphical model [15]  $X \rightarrow Y \leftarrow Z$  as displayed in Figure 27, the joint probability distribution is

$$P(X, Y, Z) = P(X)P(Z)P(Y|X, Z). \quad (64)$$

The conditional probability distribution of  $X$  and  $Z$  given  $Y$  calculates as,

$$P(X, Z|Y) = \frac{P(X)P(Z)P(Y|X, Y)}{P(Y)}. \quad (65)$$

Thus,

$$X \perp Z|Y. \quad (66)$$

## A.2 Factorizing the conditional probability distribution

### A.2.1 FHMM

We want to find the probability distribution of the state variable  $X_t^{(m)}$  in a FHMM, for an appliance  $m$  at time  $t$ , given its Markov blanket. This translates as the conditional probability distribution

$$P(X_t^{(m)} | X_{t-1}^{(m)}, X_{t+1}^{(m)}, X_t^{(1)}, \dots, X_t^{(m-1)}, X_t^{(m+1)}, \dots, X_t^{(M)}, \bar{Y}_t), \quad (67)$$

and with  $\mathbf{V} = X_t^{(1)}, \dots, X_t^{(m-1)}, X_t^{(m+1)}, \dots, X_t^{(M)}$ ,

$$P(X_t^{(m)} | X_{t-1}^{(m)}, X_{t+1}^{(m)}, \mathbf{V}, \bar{Y}_t). \quad (68)$$

Further, we will denote the previous state variable  $X_{t-1}^{(m)}$  as  $\alpha$ , and the next state variable  $X_{t+1}^{(m)}$  as  $\beta$ . Finally we will denote  $X_t^{(m)}$  simply as  $X$ . Expression (68) simplifies as

$$P(X | \alpha, \beta, \mathbf{V}, \bar{Y}_t), \quad (69)$$

and the Markov blanket is displayed in Figure 28 below.

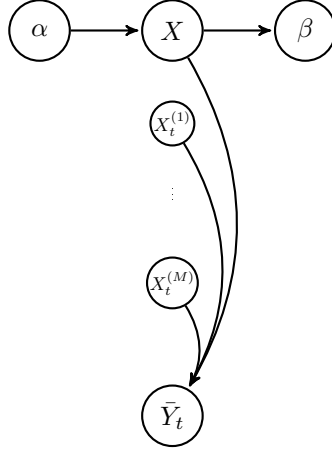


Figure 28: Dependence structure.

Using Bayes' theorem, we can see that

$$P(X | \alpha, \beta, \mathbf{V}, \bar{Y}_t) = \frac{P(X)P(\alpha, \beta, \mathbf{V}, \bar{Y}_t | X)}{P(\alpha, \beta, \mathbf{V}, \bar{Y}_t)} \quad (70)$$

$$\propto P(X)P(\alpha, \beta, \mathbf{V}, \bar{Y}_t | X) \quad (71)$$

and because  $\beta \perp \alpha, \mathbf{V}, \bar{Y}_t | X$ , (from  $\beta$ 's Markov blanket),

$$= P(X)P(\beta|X)P(\alpha, \mathbf{V}, \bar{Y}_t|X) \quad (72)$$

and following the same reasoning for  $\alpha$ ,

$$= P(X)P(\beta|X)P(\alpha|X)P(\mathbf{V}, \bar{Y}_t|X) \quad (73)$$

(chain rule),

$$= P(X)P(\beta|X)P(\alpha|X)P(\bar{Y}_t|X, \mathbf{V})P(\mathbf{V}|X) \quad (74)$$

and from the definition of FHMMs, section 2.2.3,  $X \perp \mathbf{V}$ ,

$$= P(X)P(\beta|X)P(\alpha|X)P(\bar{Y}_t|X, \mathbf{V})P(\mathbf{V}) \quad (75)$$

$$\propto P(X)P(\beta|X)P(\alpha|X)P(\bar{Y}_t|X, \mathbf{V}) \quad (76)$$

(Bayes' theorem),

$$= \frac{P(\alpha)P(X|\alpha)}{P(X)}P(X)P(\beta|X)P(\bar{Y}_t|X, \mathbf{V}) \quad (77)$$

$$\propto P(X|\alpha)P(\beta|X)P(\bar{Y}_t|X, \mathbf{V}). \quad (78)$$

Hence, the original expression has been factorized into the following components:

1.  $P(X|\alpha)$ , a transition probability from state variable  $\alpha$  to state  $X$ ,
2.  $P(\beta|X)$ , a transition probability from state  $X$  to state variable  $\beta$ ,
3.  $P(\bar{Y}_t|X, \mathbf{V})$ , a probability of observing  $\bar{Y}_t$ , given all states at time  $t$ ,
4. a constant term, for different  $X$ , other parameters unchanged.

### A.2.2 FHSMM

We want to evaluate the conditional probability of the state  $Q_t^{(m)}$ , for appliance  $m$  at time  $t$ , given its preceding state block  $A$ , its following state block  $B$ , the total emission  $\bar{Y}_t$ , and all other appliance's state blocks occurring at time  $t$ , referred to as  $\mathbf{V}$ . We will denote  $Q_t^{(m)}$  simply as  $Q$ . The simplified expression is displayed below, as

$$P(Q|A, B, \mathbf{V}, \bar{Y}_t), \quad (79)$$

Using Bayes' theorem, we can see that

$$P(Q|A, B, \mathbf{V}, \bar{Y}_t) = \frac{P(Q)P(A, B, \mathbf{V}, \bar{Y}_t|Q)}{P(A, B, \mathbf{V}, \bar{Y}_t)} \quad (80)$$

$$\propto P(Q)P(A, B, \mathbf{V}, \bar{Y}_t|Q) \quad (81)$$

and because  $A, B \perp \mathbf{V}, \bar{Y}_t|Q$ , (from  $A$  and  $B$ 's Markov blankets),

$$= P(Q)P(A, B|Q)P(\mathbf{V}, \bar{Y}_t|Q) \quad (82)$$

(chain rule),

$$= P(Q)P(A, B|Q)P(\bar{Y}_t|Q, \mathbf{V})P(\mathbf{V}|Q) \quad (83)$$

and from the definition of FHMMs, section 2.2.3,  $Q \perp \mathbf{V}$ ,

$$= P(Q)P(A, B|Q)P(\bar{Y}_t|Q, \mathbf{V})P(\mathbf{V}) \quad (84)$$

$$\propto P(Q)P(A, B|Q)P(\bar{Y}_t|Q, \mathbf{V}) \quad (85)$$

(Bayes' theorem),

$$= \frac{P(A, B)P(Q|A, B)}{P(Q)}P(Q)P(\bar{Y}_t|Q, \mathbf{V}) \quad (86)$$

$$\propto P(Q|A, B)P(\bar{Y}_t|Q, \mathbf{V}). \quad (87)$$

Hence, the original expression has been factorized into the following components:

1.  $P(Q|A, B)$ , the conditional probability of the state  $Q_t^{(m)}$  given the previous and next state blocks,
2.  $P(\bar{Y}_t|Q, \mathbf{V})$ , the conditional probability of observing  $\bar{Y}_t$ , given  $Q$  and the state blocks for all other appliances at time  $t$ ,
3. a constant term, for different  $Q$ , other parameters unchanged.

## B Data set

The Reference Energy Disaggregation Data set (REDD) is a freely available data set containing detailed power usage information from several homes, which is aimed at furthering research on energy disaggregation [11]. The dataset comprises measurements from six households, measured over several weeks. The measurements used from this dataset were resampled to a regular frequency of 20 seconds.

The appliances were ranked according to power usage, as a percentage of the whole data set. The top ten appliances are displayed in Table 3.

Label	Rel. power usage
Refrigerator 1	3.82 %
Furnance 1	1.72 %
Dishwasher 1	0.75 %
Microwave 1	0.70 %
Air conditioning 2	0.58 %
Air conditioning 3	0.57 %
Washer dryer 3	0.55 %
Washer dryer 1	0.49 %
Electric heat 1	0.40 %
Washer dryer 2	0.37 %

Table 3: Top ten appliances, ranked according to power usage, as a percentage of all power measurements. The avid reader will spot the misspelling of "furnace", which was kept throughout this report for consistency with the used dataset.

### B.1 Resampling REDD

The measurements in the REDD dataset were found to have varying frequency, (with the lowest frequency documented as about 1 Hz). As the model required the fitting to be done with regularly measured data, all data from REDD was preprocessed accordingly. First, the dataset was up-sampled to regular frequencies (2 Hz) with a forward fill method to remove gaps in the data. In the second step, the data was down-sampled to the low frequency interval of our choice. As down-sampling method, the median of each segment was chosen as the sample.

## C ML distribution tables

Table 4: REDD *ON*-durations: Estimated parameters for the exponential ( $\lambda$ ) and gamma ( $k, \theta$ ) distributions, and LLR. Also displayed is the number of samples used for the estimation of these parameters.

<i>ON</i> -duration distributions, REDD					
Label	Samples	$\lambda$	$k$	$\theta$	LLR
bathroom gfi 1	11438	1.060	0.843	1.119	-4.08E+02
dishwaser 1	6095	1.216	1.212	0.679	7.19E+02
electric heat 1	5123	1.127	1.154	0.769	3.17E+02
kitchen outlets 2	4430	0.284	0.292	12.075	3.95E+04
microwave 1	3848	0.574	0.602	2.896	3.27E+03
stove 1	3537	0.572	0.510	3.431	3.10E+03
bathroom gfi 2	2743	1.079	0.479	1.934	-4.92E+02
smoke alarms 1	2465	0.329	0.283	10.735	1.46E+04
mains 2	2259	0.117	0.561	15.210	1.53E+05
refrigerator 1	2164	0.059	3.301	5.139	6.08E+05
kitchen outlets 1	1535	0.815	0.439	2.795	-1.53E+02
outlets unknown 2	1454	0.626	0.432	3.697	6.77E+02
washer dryer 1	1386	0.375	0.373	7.159	5.61E+03
mains 1	1361	0.096	0.699	14.889	1.40E+05
lighting 1	1180	0.156	0.496	12.951	4.33E+04
outlets unknown 1	991	0.225	0.374	11.899	1.57E+04
lighting 3	854	0.182	0.152	36.211	2.20E+04
furance 1	656	0.100	0.219	45.550	6.17E+04
disposal 1	394	0.605	0.392	4.223	2.39E+02
lighting 4	378	0.080	0.279	44.810	5.70E+04
electronics 1	296	1.333	0.638	1.176	-2.27E-01
washer dryer 3	250	0.448	4.782	0.466	7.08E+02
air conditioning 1	203	0.294	0.525	6.468	1.64E+03
kitchen outlets 3	184	0.414	1.478	1.634	5.61E+02
air conditioning 3	169	0.549	0.253	7.182	1.68E+02
air conditioning 2	162	0.435	0.393	5.851	4.12E+02
washer dryer 2	131	0.305	0.076	43.301	9.18E+02
lighting 5	131	0.033	0.556	54.948	1.21E+05
lighting 2	101	0.008	1.387	88.601	1.52E+06
subpanel 2	87	0.044	2.086	10.791	4.35E+04
outlets unknown 3	61	0.108	0.078	119.341	4.92E+03
kitchen outlets 4	48	0.576	3.520	0.493	5.04E+01
oven 1	47	0.415	1.682	1.434	1.50E+02
oven 2	47	0.415	1.682	1.434	1.50E+02
outdoor outlets 1	45	1.007	1.629	0.609	4.80E+00
miscellaeneous 1	7	0.102	4202.500	0.002	6.54E+02
electric heat 2	2	0.029	2.185	15.559	2.30E+03
subpanel 1	0	0.000	0.000	0.000	0.00E+00

Table 5: REDD *OFF*-durations: Estimated parameters for the exponential ( $\lambda$ ) and gamma ( $k, \theta$ ) distributions, and LLR. Also displayed is the number of samples used for the estimation of these parameters.

<i>OFF</i> -duration distributions, REDD					
Label	Samples	$\lambda$	$k$	$\theta$	LLR
bathroom gfi 1	11401	1.265	1.005	0.787	1.10E+03
dishwasher 1	6015	0.798	0.191	6.564	-2.61E+03
electric heat 1	5104	1.307	1.064	0.719	7.08E+02
kitchen outlets 2	4389	0.211	0.205	23.119	8.07E+04
microwave 1	3809	0.284	0.233	15.081	3.39E+04
stove 1	3544	0.556	0.410	4.389	3.40E+03
bathroom gfi 2	2728	0.292	0.315	10.891	2.25E+04
smoke alarms 1	2468	1.355	0.759	0.972	1.71E+02
mains 2	2275	0.106	0.607	15.492	1.89E+05
refrigerator 1	2145	0.033	4.119	7.333	1.94E+06
kitchen outlets 1	1503	0.267	0.166	22.551	1.56E+04
outlets unknown 2	1413	0.156	0.251	25.534	5.15E+04
washer dryer 1	1372	1.027	0.993	0.981	-4.64E-01
mains 1	1358	0.053	0.880	21.497	4.77E+05
lighting 1	1189	0.084	0.352	33.644	1.61E+05
outlets unknown 1	950	0.082	0.233	52.319	1.36E+05
lighting 3	808	0.082	0.116	105.631	1.16E+05
furance 1	661	0.101	0.182	54.217	6.07E+04
disposal 1	367	0.047	0.204	104.351	1.64E+05
lighting 4	349	0.016	0.431	148.395	1.42E+06
electronics 1	297	0.509	0.577	3.407	4.36E+02
washer dryer 3	240	1.047	1.134	0.843	9.10E+00
air conditioning 1	177	0.061	0.219	74.937	4.64E+04
air conditioning 3	162	0.127	0.238	32.999	9.17E+03
kitchen outlets 3	146	0.020	0.311	159.022	3.57E+05
air conditioning 2	140	0.056	0.179	99.870	4.40E+04
washer dryer 2	120	0.444	0.146	15.458	2.53E+02
lighting 5	114	0.013	0.580	135.900	7.08E+05
subpanel 2	90	0.037	3.816	7.134	6.60E+04
lighting 2	72	0.004	1.194	228.774	5.37E+06
outlets unknown 3	55	0.395	0.062	41.167	1.36E+02
oven 1	39	0.237	3.133	1.345	5.50E+02
oven 2	39	0.237	3.133	1.345	5.50E+02
outdoor outlets 1	39	0.929	0.178	6.039	-2.15E+01
kitchen outlets 4	33	0.013	0.328	228.538	1.85E+05
miscellaaneous 1	13	0.067	61.708	0.242	2.84E+03
subpanel 1	0	0.000	0.000	0.000	0.00E+00
electric heat 2	0	0.000	0.000	0.000	0.00E+00



## D Emission distributions

This section displays the emission distributions of the appliances, as discussed in section 3.2.1. Table 6 is a summary of the empirical distribution counts. Figures 29-34 show the positive emission distributions as histograms, with the threshold value  $\theta$  displayed as a dotted line, where appropriate. Bin size is 10 W, and all figures exclude the first bin (0-10 W).

Emission counts					
Label	Household	$\theta$ (W)	Count: $y^+ > \theta$	Count: $y^+ \leq \theta$	Count: $y^+ > \theta$
microwave 1	1	9.80	2088	154618	156706
stove 1	1	10.47	62	3093	3155
refrigerator 1	1	26.24	37574	119128	156702
lighting 1	1	22.87	86506	56959	143465
lighting 3	1	6.89	36709	119451	15610
furance 1	3	4.97	1519	191510	193029

Table 6: Count of categorized emissions, with respect to each appliance's threshold value  $\theta$ , for the selection of appliances from REDD.

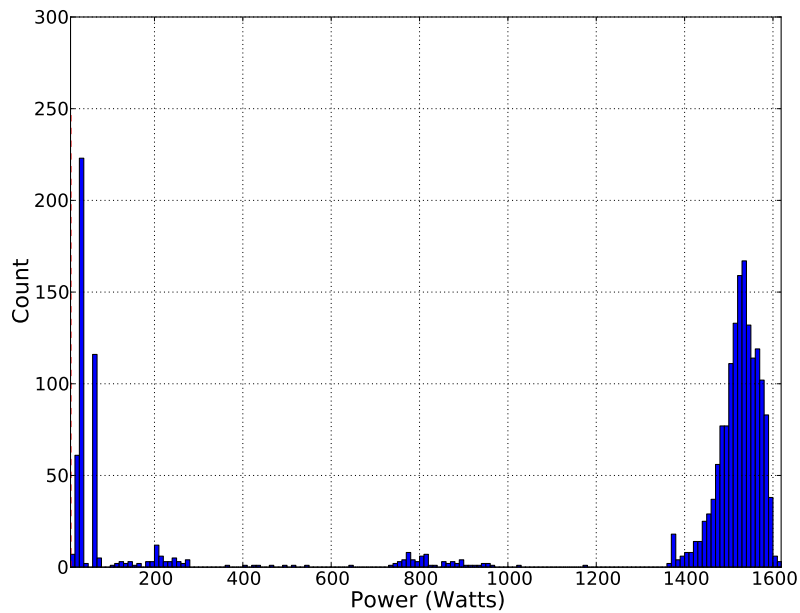


Figure 29: Microwave 1

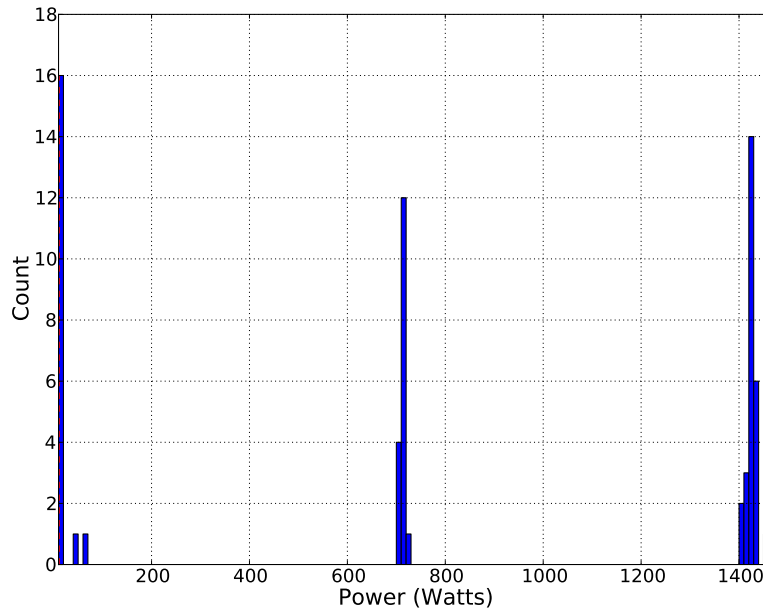


Figure 30: Stove 1

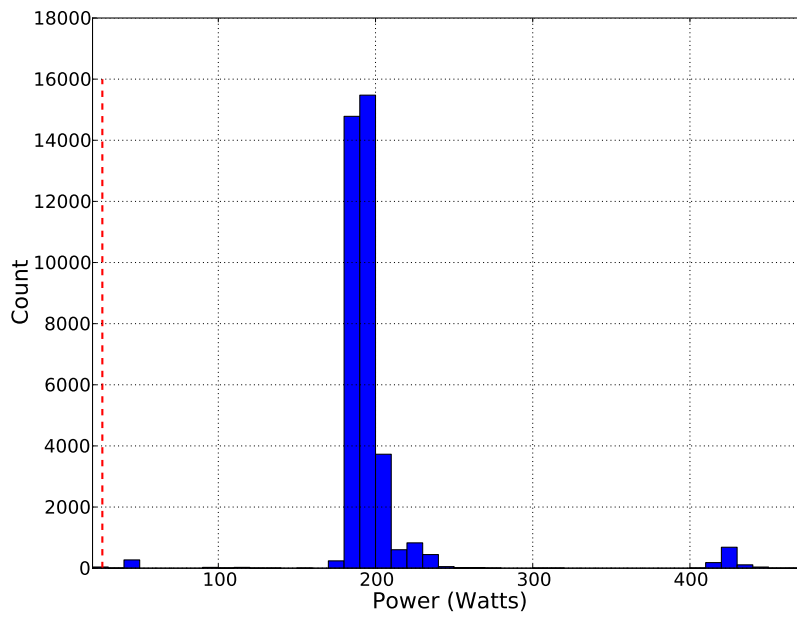


Figure 31: Refrigerator 1

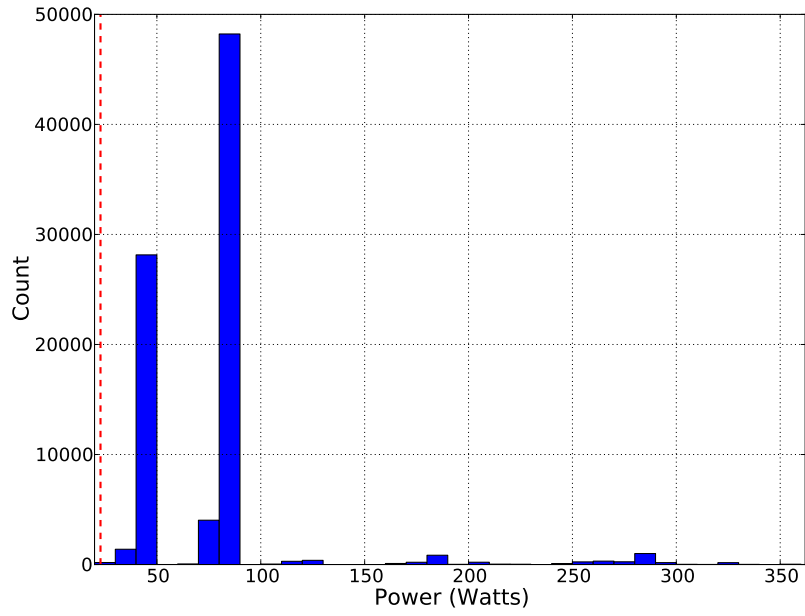


Figure 32: Lighting 1

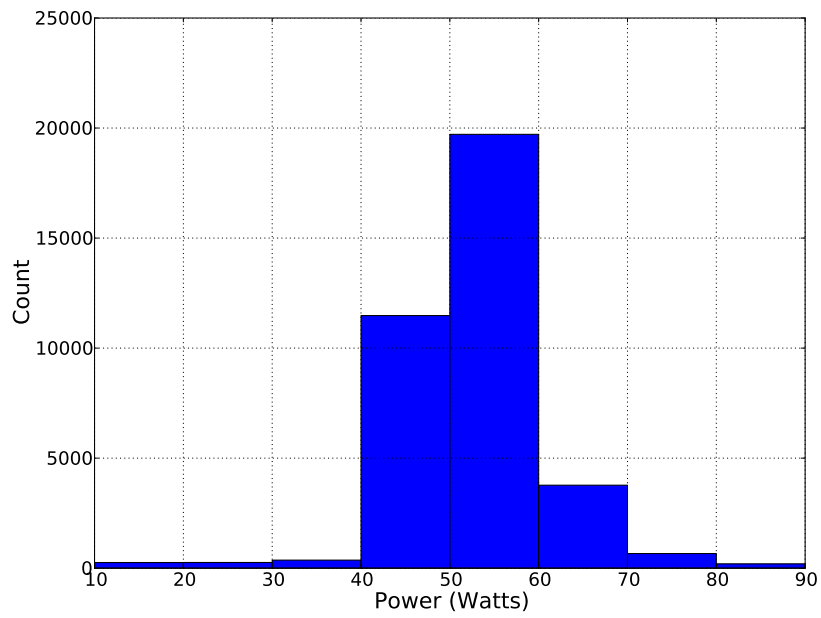


Figure 33: Lighting 3

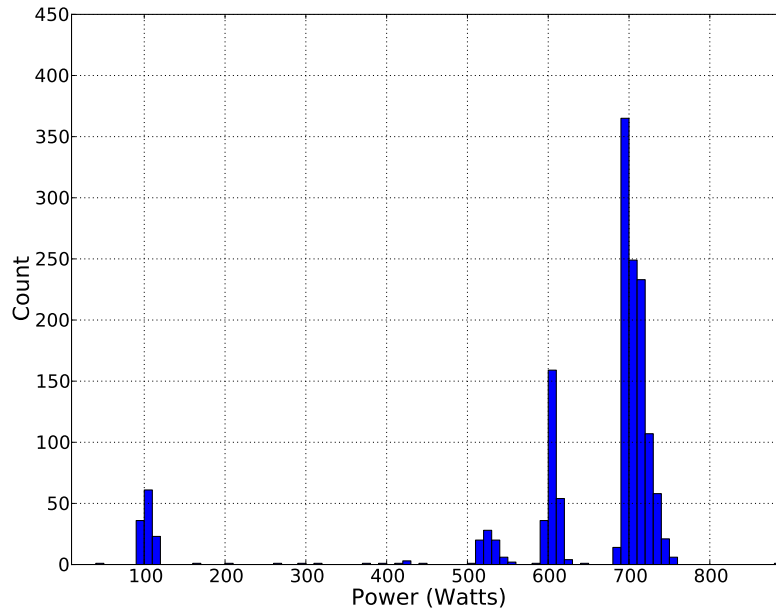


Figure 34: Furance 1

## E Duration distributions

This appendix displays the duration distributions of *ON* and *OFF*-states in the appliances from Table 1. The distributions are cut of at the 95th percentile, and the red curve represents the ML gamma density function.

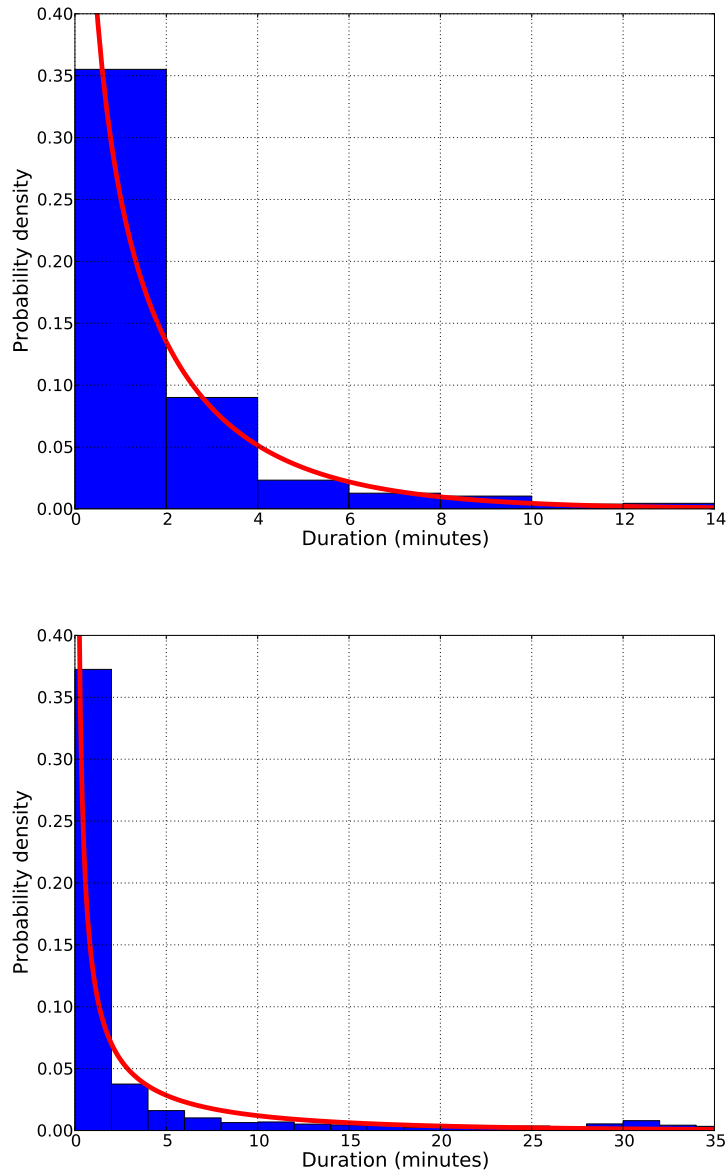


Figure 35: Microwave 1. Above: *ON*-durations. Below: *OFF*-durations.

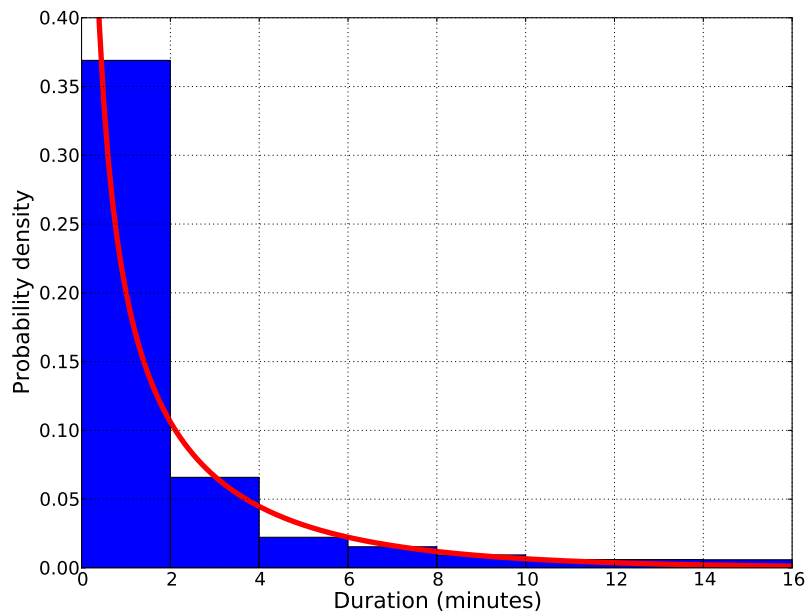
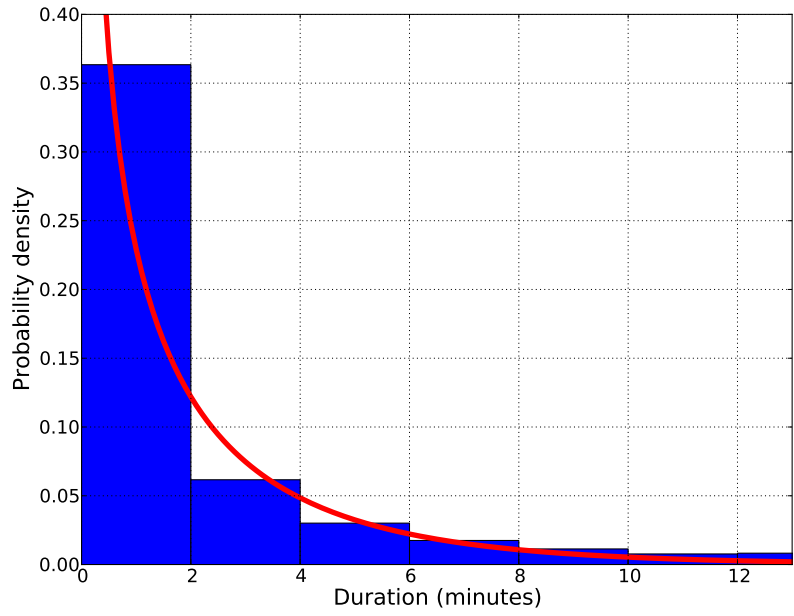


Figure 36: Stove 1. Above: *ON*-durations. Below: *OFF*-durations.

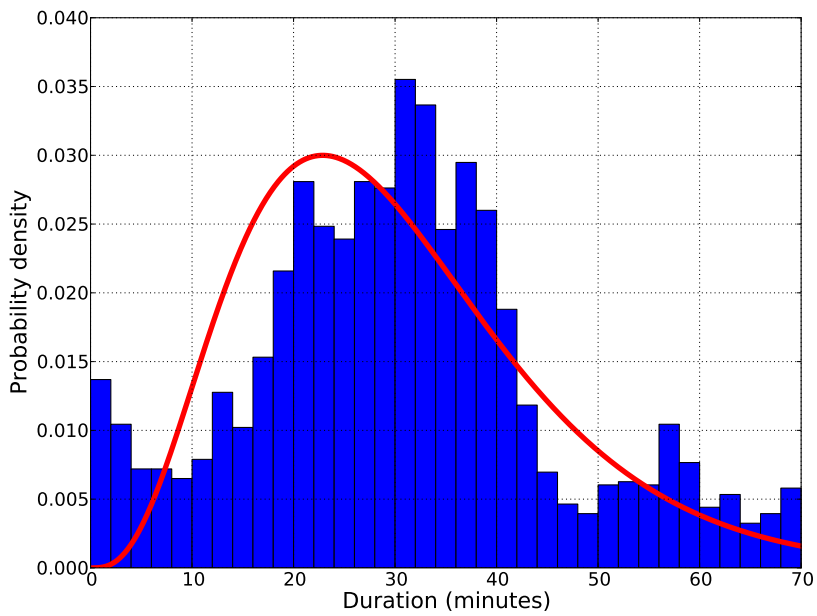
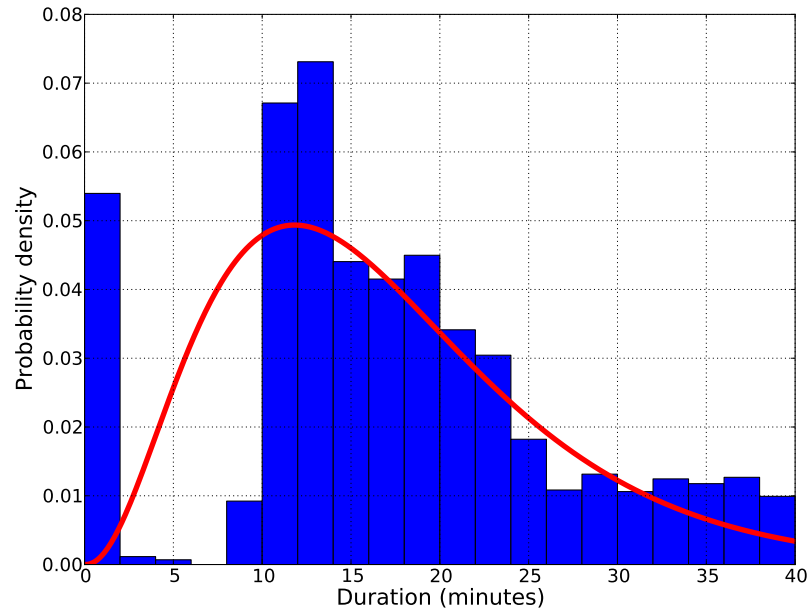


Figure 37: Refrigerator 1. Above: *ON*-durations. Below: *OFF*-durations.

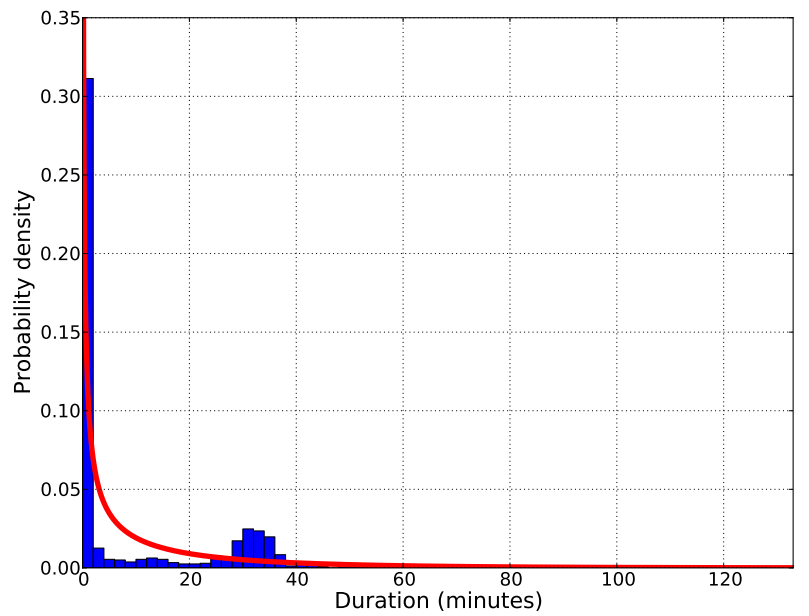
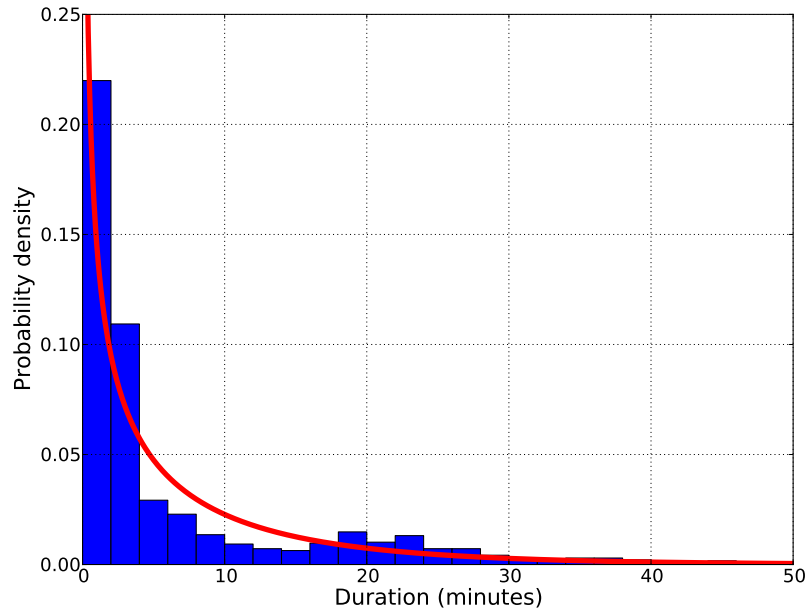


Figure 38: Lighting 1. Above: *ON*-durations. Below: *OFF*-durations.



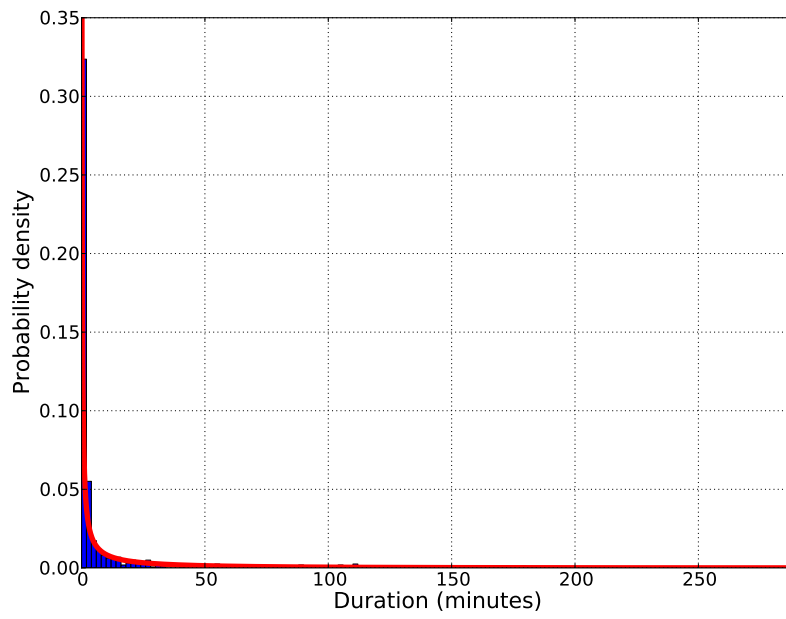
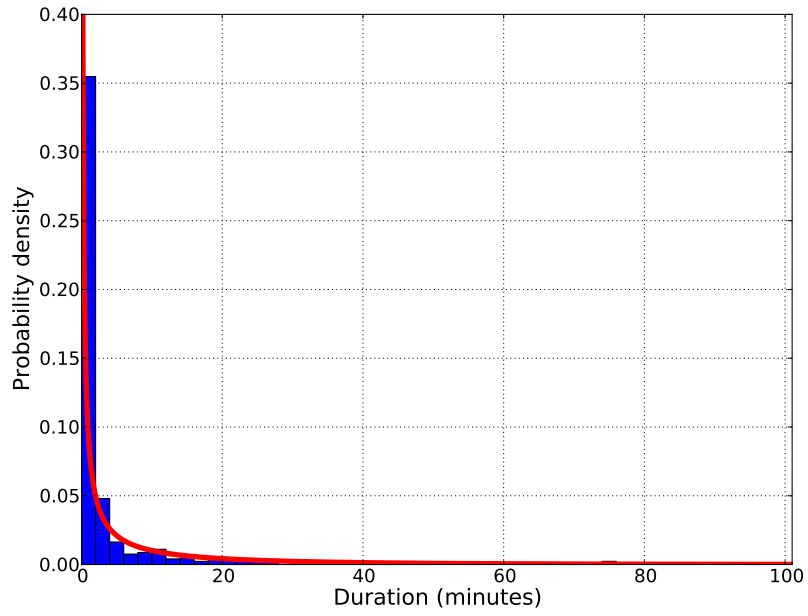


Figure 39: Lighting 3. Above: *ON*-durations. Below: *OFF*-durations.

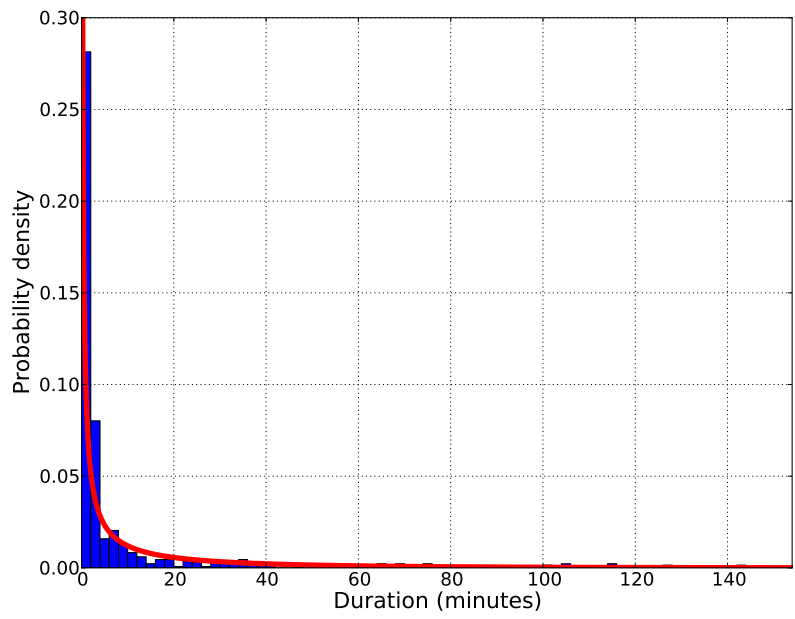
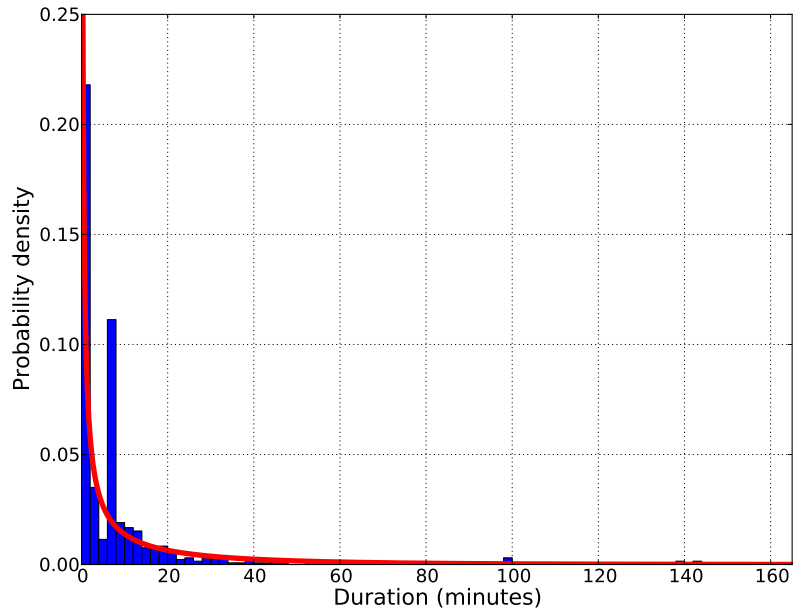


Figure 40: Furance 1. Above: *ON*-durations. Below: *OFF*-durations.

## References

- [1] Comission of the European Communities, Action Plan for Energy Efficiency: Realising the Potential, COM(2006)545 final, 19.10.2006.
- [2] Jon Froehlich, Eric Larson, Sidhant Gupta, Gabe Cohn, Matthew S. Reynolds, Shwetak N. Patel. Disaggregated End-Use Energy Sensing for the Smart Grid. *IEEE Pervasive Computing, Special Issue on Smart Energy Systems*, 2011.
- [3] Steve Brooks, Andrew Gelman, Galin Jones and Xiao-Li Meng. *Handbooks of Markov Chain Monte Carlo, Handbooks of Modern Statistical Methods*. Chapman & Hall/CRC, 2011.
- [4] Jan Enger and Jan Grandell. *Markovprocesser och köteori. Compendium*, KTH: Division of Mathematical Statistics, 2003.
- [5] Zoubin Ghahramani and Michael I. Jordan. Factorial Hidden Markov Models. *Machine Learning*, 29(2), pages 245-273, 1997.
- [6] Wally R. Gilks, Sylvia Richardson and David J. Spiegelhalter. *Markov Chain Monte Carlo in Practice, Interdisciplinary Statistics*. Chapman & Hall/CRC, 1996.
- [7] Allan Gut. *An Intermediate Course in Probability*, Second edition. Springer, 2009.
- [8] Matthew J. Johnson and Alan S. Willsky. Bayesian Nonparametric Hidden Semi-Markov Models. *Journal of Machine Learning Research*, 14, pages 673-701, 2013.
- [9] Hyungsul Kim, Manish Marwah, Martin Arlitt, Geoff Lyon and Jiawei Han. *Unsupervised Disaggregation of Low Frequency Power Measurements*. Technical report, HP Labs Tech. Report, 2010.
- [10] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.
- [11] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In *Proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*, 2011.
- [12] J. Zico Kolter and Tommi Jaakkola. Approximate Interference in Additive Factorial HMMs with Application to Energy Disaggregation. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2012.
- [13] Christopher Laughman, Kwanduk Lee, Robert Cox, Steven Shaw, Steven Leeb, Les Norford, and Peter Armstrong. *Power Signature Analysis*. *IEEE Power and Energy Magazine*, Vol. 1 No. 2, pages 56-63, 2003.
- [14] Kevin P. Murphy. *Hidden semi-Markov models (HSMMs)*. Technical Report, November 2002. URL <http://www.cs.ubc.ca/~murphyk/Papers/segment.pdf> (last accessed October 2013).

- [15] Kevin P. Murphy. Machine learning: A Probabilistic Perspective. MIT Press, 2012.
- [16] Lawrence R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, pages 257-286, 1989.
- [17] Felix L. Rios and Timo Koski. Non-Intrusive Appliance Load Monitoring (NIALM) - Proof of Concept. Unpublished technical report.
- [18] Ahmed Zoha, Alexander Gluhak, Muhammad Ali Imran and Stharshan Rajasegarar. Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey. Sensors (Basel), 12(12), pages 16838-16866, 2012.