Enhanced approaches for non-intrusive load disaggregation

Aiad, Misbah M. M.

2018

https://hdl.handle.net/10356/90196

https://doi.org/10.32657/10220/47182

Enhanced Approaches for Non-Intrusive Load Disaggregation

MISBAH M. M. AIAD

School of Electrical and Electronic Engineering

A thesis submitted to Nanyang Technological University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

2018

Acknowledgments

I would like to express my deepest gratitude and thanks to my supervisor, Associate Professor Lee Peng Hin, for his great encouragement, valuable guidance, patience, endless support and motivation throughout the course of my Ph.D. His precious advices have always been inspirational to me, not only in research work, but also for a wide range of personal skills. Without being under his supervision and leadership, it would not have been possible to accomplish this work.

I would like to acknowledge the Singapore International Graduate Award (SINGA) for providing the financial support to carry out my Ph.D.

I would like to express special thanks to my wife for her sincere love and encouragements throughout my Ph.D. candidature.

Table of Contents

List of Figures

List of Tables

Abstract

The modern urban life and increasing demands of energy are calling toward energy conservation and energy efficient strategies. Energy saving and energy management in the residential sectors are of great interests for obvious economic and environmental reasons, with increasing energy consumptions by the consumers. An efficient energy conservation and monitoring program requires some means of monitoring the power consumed by individual appliances within the households. The deployment of smart meters in smart grids in many countries has generated an increase in research interests in the areas of non-intrusive load monitoring (NILM) in recent years. Non-intrusive load monitoring, or load disaggregation, are sets of techniques and methods that decompose the total aggregate consumptions, measured at a single point by smart meters, into the respective appliance-specific consumptions in the household. Studies conducted have shown that information of the energy consumed by individual appliances in the homes can influence the behavior of the household occupants in a way that can achieve noticeable energy savings. There are several challenges in the domain of unsupervised load disaggregation approaches that do not require human intervention for learning or installation of additional measuring instruments for each appliance, apart from the smart meters, allowing a feasible economic adoption of NILM techniques.

In this thesis, a detailed literature review on methods and techniques applied to NILM and common challenges is presented. Enhanced approaches that tackle three essential challenges in the domain of NILM were proposed. Firstly, with the aim to achieve an improved disaggregation accuracy, an unsupervised approach for load disaggregation that embeds the mutual devices interactions information into the factorial hidden Markov model (FHMM) representation of the total aggregate signal was introduced.

The method was further extended with adaptive estimations of the devices main power consumptions effects and their two-way interactions. Secondly, the modeling of continuously varying loads was proposed using a quantized continuous-state hidden Markov model (CS-HMM). A method to estimate the transition matrix that mitigates the both extreme cases of too frequent and never occurred transitions was introduced and the Viterbi algorithm was used to estimate the power consumption profile of the variable loads. Thereafter, the proposed model for the continuously varying loads was integrated with the standard FHMM to produce a hybrid continuous/discrete state HMM, which is capable of modeling and disaggregating energy consumptions from a wider range of home appliances types. Thirdly, to tackle the problem of overlapping clusters that represent devices power consumptions resulting when applying a clustering-based disaggregation, a method to analyze the cohesion of devices' clusters to determine if a cluster should be split into two small clusters was proposed. The analysis of clusters cohesion was investigated based on normality tests performed against two confidence levels.

The proposed approaches and techniques were applied and tested on real houses from the Reference Energy Disaggregation Data Set (REDD). The proposed approaches, in general, enhanced the overall performance and accuracy of disaggregation. The work presented in thesis represents an advancement in the state-of-art in the domain of NILM and contributes toward achieving energy savings in residential homes.

Chapter 1

Introduction

It is an established fact that energy plays a principal part in the contemporary urban life. The global demands of various types of energy has rapidly increased due to the increase of world population and vast developments in industry and transportation sectors. Energy statistics shows that total world energy consumption grew by 1.0% in 2016 to reach 13276.3 million tonnes oil equivalent (toe) [BP P.L.C., 2017]. Consequently, several challenges are augmented particularly regarding utilization of sustainable energy resources alternative to fossil fuel and regarding the arising necessity of mitigating undesired environmental impacts caused due to global energy consumption. Energy saving in residential homes shares in achieving the aforementioned goals as it results in significant reduction in the total consumed raw energy.

In the recent BP energy outlook 2018 report, it is expected in 2020 that approximately 29.12% of worldwide energy consumption is consumed in buildings, of which 45.52% of energy consumption in buildings is in the form of electricity. That is, worldwide electricity consumption in buildings in 2020 is expected to be approximately 13.26% of the total energy consumption. For illustration, Figures 1.1a and 1.1b show the energy consumption by sector and the share of electricity consumption in buildings, respectively [BP P.L.C., 2018].

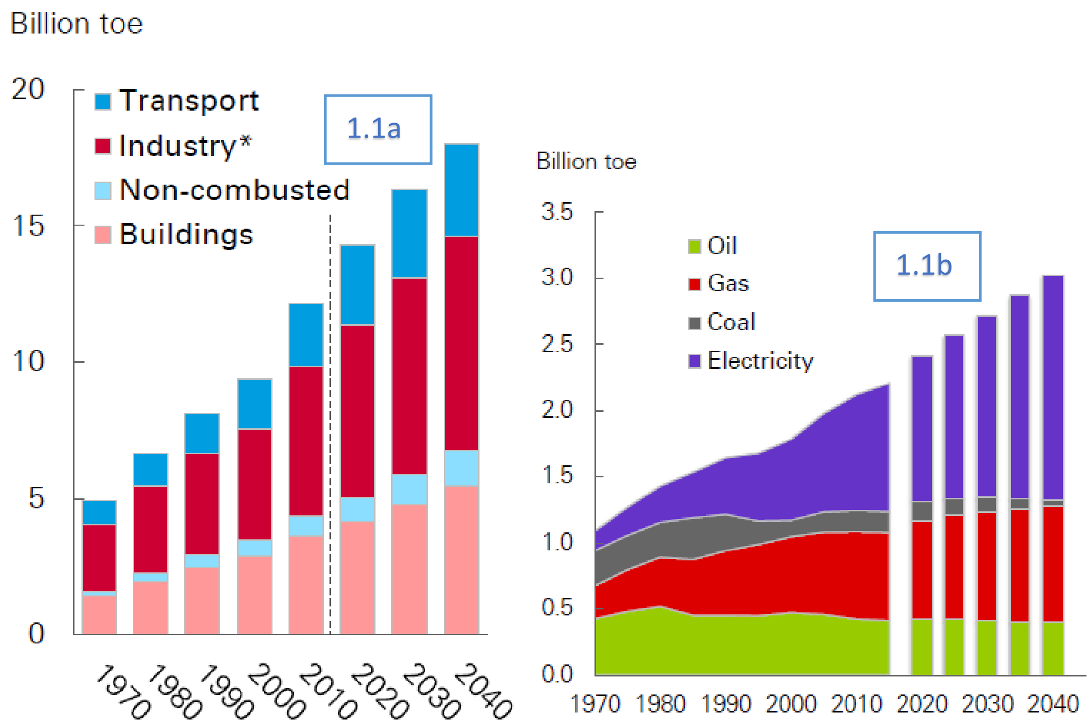

Fig. 1.1a and 1.1b: The energy consumption by sector and the share of electricity consumption in buildings, respectively [BP P.L.C., 2018].

Facts on electricity consumption share in buildings are useful in motivating the research work in the field of non-intrusive load disaggregation, which aims basically to achieve energy saving in households.

1.1 Motivations

Energy saving and management are arising interests in the deployment of the smart grids due to continuous fluctuations in the global raw energy markets (e.g. fossil fuel) and inevitable environment impacts caused due to energy consumption. Savings in electricity consumption in residential homes can be achieved by two principal ways [Bollen, 2011]

1. Using energy-efficient appliances that consume less amounts of electrical energy.
2. Using less appliances, or using existing appliances less often.

Therefore, it is recommended to replace old households' appliances, if possible, with energy-efficient ones. It may not always be possible to reduce the number of appliances in a household. Hence, occupants are advised to use their existing appliances less often to accomplish energy savings.

The current metering infrastructure of electrical energy consumption in majority of countries provides household occupants with their total energy consumption and associated cost at every stipulated time period (e.g. monthly). The occupants are not aware of contribution shares by each individual appliance existing in their homes. Conducted studies showed that significant energy savings between 9% to 20% of current energy consumption can be achieved once the household occupants are aware of detailed information on energy consumption by individual appliances [Kim *et al.*, 2010; Parson *et al.*, 2014]. These studies showed that when occupants are aware of individual consumptions by respective appliances, they are influenced by the breakdown of information in a way to use home appliances less often, especially those appliances with high power ratings.

Non-intrusive load monitoring (NILM) is a set of methods and techniques that aim to decompose the total aggregate measurements by smart meters into the detailed individual consumptions profiles by appliances present in the household [Zeifman and Roth, 2011]. To achieve a promising solution to NILM, it is essential to be able to perform the proposed disaggregation process only using the total aggregate signal measured by smart meters at household-level. Sub-metering or installing additional sensors at appliance-level will apparently incur extra cost, time, privacy and applicability challenges [Zoha *et al.*, 2012; Kim *et al.*, 2010].

Beside targeting energy savings in residential homes, NILM techniques are promising to benefit in broader useful applications such as [Chou and Chang, 2013]

- Energy management systems.

- Demand side management.
- Faults troubleshooting.

1.2 Problem Statement

Load disaggregation objective is to obtain the power consumption profiles of individual appliances merely from the total aggregate signal measured by a smart meter from a single point at a household-level. That is, if there are M appliances present in a household, and the total aggregate measurements $p(t)$ was observed over the time period $t = 1, 2, \dots T$, then

$$p(t) = \sum_{m=1}^{M} p_m(t) \tag{1.1}$$

An approach of NILM aims to disaggregate the total aggregate measurements $p(t)$ to obtain estimates of the individual power consumptions $p_m(t)$ for each appliance m present in the household during the time period of observation $t = 1, 2, \dots T$.

The task of NILM increases in complexity and becomes more challenging based on the following important factors

- The number of active appliances and number of occupants in the household.
- The usage patterns and behaviors of the household occupants.
- Existence of appliances that overlap in their power consumptions.
- Existence of continuously varying loads such as light dimmers, electronic devices and power tools.
- Noisy signals from possible sources that may affect the accuracy of metering.
- Other applicability and scalability concerns.

In brief, the general framework of NILM methodologies consists of three consecutive tasks [Zoha *et al.*, 2012]

1. Data acquisition: measurements of total aggregate consumption and other signals and information, if applicable.
2. Features extraction: extracting useful features (also called appliances signatures or attributes) from the measured signals.
3. Load identification: using a proper method or algorithm to identify individual appliances based on their distinct features and other available information.

Overall, to provide an efficient scalable solution to NILM, a proposed approach for NILM should fulfil a number of requirements such as [Parson, 2014]

1. **Non-intrusive:** the proposed approach should only make use of total aggregate signal measured at household-level. The approach should not require intrusive sub-metering of individual appliances or installing additional sensors to report or record the operation of individual home appliances.
2. **Load disaggregation:** the proposed approach should provide estimates of individual appliances consumptions profiles at same sampling resolution as that of the total aggregate signal.
3. **Low frequency measurements:** the proposed approach should utilize available features in current broadly-adopted low frequency measurement instruments (e.g. smart meters operating at 1 Hz frequency). Though metering at higher frequencies can provide extra useful features (e.g. harmonics and transients), installing additional high frequency metering equipment will incur extra cost and applicability challenges.
4. **Unsupervised learning:** the proposed approach should learn appliances models online from the aggregate signal. Training data (e.g. individual appliances models)

should not be a pre-requirement since this settles an intrusive load disaggregation technique.

1.3 Major Contributions

In this work, three enhanced approaches are proposed to tackle three essential challenges in the domain of NILM which are enhancing the load disaggregation accuracy, modeling and disaggregating varying loads and disaggregating overlapping home appliances. Firstly, mutual interactions between appliances is introduced as a new feature that can be embedded in the factorial hidden Markov model (FHMM) representations of home appliances and the total aggregate signal so as to improve the disaggregation accuracy of individual loads (home appliances). Secondly, a method is proposed to model continuously varying home loads using a quantized continuous-state hidden Markov model (CS-HMM) that estimates the transition matrix in a way to mitigate possible extreme cases of transitions: never-occurred transitions and too-frequent transitions. Thereafter, the proposed CS-HMM is consolidated with the factorial hidden Markov model (FHMM) to produce a hybrid continuous/discrete state FHMM that capable of modeling and disaggregating various types of loads. Thirdly, to disaggregate home appliances with overlapping power consumptions, an approach for clusters splitting is proposed based on inspection of clusters cohesion to split appliances clusters where appliances are expected to be overlapping in their power consumptions.

In this work, the low frequency measurements were used for cost reduction, availability and applicability reasons as stated in section 2.2.1. In addition, unsupervised learning approaches were adopted in the learning phase of HMMs and its variants (approaches proposed in Chapter 3 and Chapter 4) for cost reduction, privacy and applicability reasons as stated in section 2.2.2. Figure 1.2 illustrates the adopted research path, which is directed on using unsupervised learning approaches to be applied on low frequency measurements of aggregate consumptions (e.g. sampled at a rate of 1 Hz).

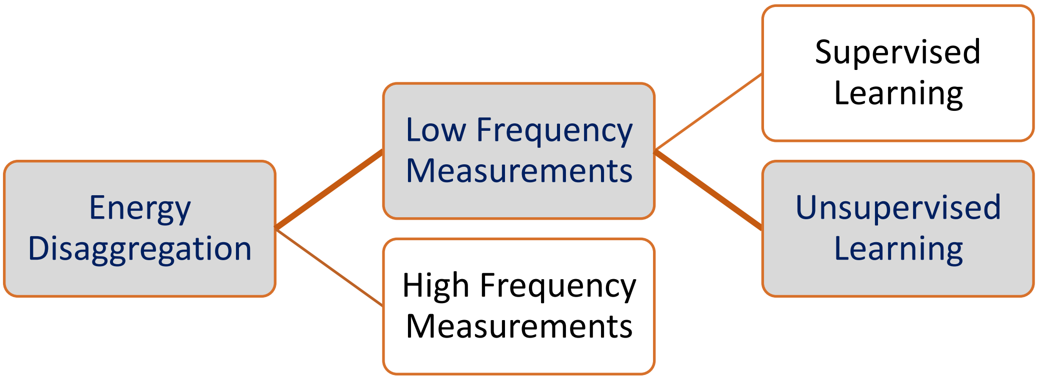

Fig. 1.2: Adopted research path in this thesis (shaded path).

The major contributions of this research, which aim to tackle essential gaps in literature, are articulated in brief in the following sections.

1.3.1 Enhanced Load Disaggregation with Mutual Devices Interactions

The feature of mutual interactions between appliances is introduced and utilized in the FHMM representation of the aggregate total signal in order to achieve improved disaggregation accuracies of individual home appliances. Interactions between appliances is interpreted as the lumped effects of power quality issues such as harmonics and electromagnetic interference. Home devices are prone to interact (to some extent) since they are usually connected in parallel to the same feeder line. A statistical factorial design was adopted to estimate devices main power effects (primary power consumptions) and their possible two-way interactions, which ultimately affect the total aggregate measurements. This information was embedded in the FHMM representation of the aggregate measurements with the purpose of improving the disaggregation accuracies per existing device. An interaction between two appliances is expected to affect the total aggregate signal only in cases when both corresponding appliances are detected to be operating in the ON state.

The above approach was further extended using an adaptive algorithm that updates both devices main power effects and their two-way interactions (where possible) during the disaggregation process. The adaptive method was applied in cases when there are four devices or less operating in the ON state simultaneously. A suitable fractional factorial design was utilized to assist in updating devices models and estimates of two-way interactions based on the number of simultaneously ON devices. An unsupervised approach was applied on low frequency measurements of a real house from the Reference Energy Disaggregation Data set (REDD) [Kolter and Johnson, 2011] to test and validate the proposed approaches.

1.3.2 Modeling of Continuously Varying Loads by a Quantized CS-HMM and a Proposed Hybrid Continuous/Discrete States FHMM

Continuously varying loads (such as light dimmers, power tools and electronic devices) are those having no consistent finite number power consumption levels. Based on their nature of operation, these loads may consume varying amounts of power within their power consumption ranges. Hence, the standard discrete-state HMM (DS-HMM) may be incapable of modeling and disaggregating this type of loads. A quantized CS-HMM is proposed to model and estimate the power consumption of continuously varying loads. A modified method is proposed to estimate the model transition matrix that mitigate two possible extreme cases of transitions

1. Never-visited transitions during the learning phase, which may result in being incapable of detecting these transitions during the testing and validation phase.
2. Too frequent transitions, which may result in a dominant transition probability that is likely to limit the possibility of detecting other transitions in the same row of the transition matrix.

The Viterbi algorithm was used to infer most probable hidden sequence of states, given a sequence of measured observations. The proposed quantized CS-HMM was tested on synthetic data that emulate varying loads and on real devices measurements available from the REDD public data set. In addition, the proposed quantized CS-HMM was combined with the standard FHMM to produce a hybrid CS/DS-FHMM that capable of modeling and disaggregation of various types of discrete loads (FSMs) and continuously varying loads. A framework for both learning and estimation phases was developed so as to model and disaggregate various types of loads that exist simultaneously in a household merely from their total aggregate power measurements.

1.3.3 Disaggregating Overlapping Clusters of Appliances Using a Clusters Splitting Approach

Clustering methods were applied to NILM where the objective is to group home appliances into representative distinct clusters. When two devices or more have close or overlapping power consumptions, it is likely that their corresponding representative clusters be wrongly merged in one cluster when applying some clustering techniques. However, since such merged cluster has originated by data from more than one sources (i.e. two or more home appliances), it is reasonable to observe a degree of non-cohesiveness between its elements. An approach for cluster splitting is proposed based on inspecting the degree of within-cluster cohesion. The cohesion test is carried out using three common normality tests performed against two confidence levels. Since the proposed approach of cohesion test is basically based on normality tests, the following considerations were taken into account

1. Only devices with stable single state or finite state machines (FSMs) were included in the study to maintain good normality fitting’s in their clusters.

2. As some devices show an overshoot starting consumption, the overshoot periods were excluded since they may alter normality patters within the device cluster.
3. A median filter was applied in all cases to remove outliers and noisy signals.

In brief, a splitting of a cluster is carried out by the flowing procedure

1. Apply normality tests on the checked cluster.
2. Conclude the degree of cohesiveness to decide splitting (some threshold values are set to conclude the splitting decision).
3. Once a splitting is decided, split the cluster into two sub-clusters using a proper method like inner expectation maximization (EM).

1.4 Alternate Solutions Other than NILM

There are some alternate methods and techniques other than NILM techniques that may be used or developed to inform households occupants about detailed appliance-level power consumptions. Examples of these methods include [Parson, 2014]

1. **Sub-metering:** where each appliance power consumption is metered or logged individually using a special metering equipment.
2. **Smart appliances:** where some modern appliances are designed with a technology that reports its power consumption information to a central hub in the household.
3. **Appliances surveys:** where surveys on appliances and their use by occupants can be used to conclude estimates of appliances power consumptions.

Overall, these methods face noticeable shortcomings in terms of applicability, availability and estimation accuracy. For example, the sub-metering technique requires installing extra sensors at appliance-level which incurs additional cost, applicability and privacy challenges. Smart appliances are a good option to report appliances consumptions, but this technology is not yet embedded in all appliances types and brands. Surveys on appliances and their usage patterns may provide an estimate of appliances usage. However, these estimates may not be accurate in many cases since people differ widely in their behaviors and usage patterns of different types of home appliances.

1.5 Organization of This Thesis

The remaining parts of this thesis are organized as follows:

Chapter 2 presents a background and detailed literature reviews on related research work done in the domain of NILM. The principal inadequacies in available approaches are demonstrated respectively. It also articulates main useful appliances features used in NILM methods and compares different accuracy metrics used to assess proposed NILM approaches. An overview of HMMs, extensions of HMMs and clustering techniques is presented as these models are enhanced thereafter in this research work.

Chapter 3 presents the first proposed approaches to enhance load disaggregation accuracy using information on appliances mutual interactions. Theoretical parts of the proposed approach are augmented with explanations on how the proposed approaches can mitigate potential shortcomings that often hamper reference methods. In addition, case study applications on real measurements of home appliances from available public data set were used to carry out models learning, testing and validation tasks. Reference methods were also applied on same data for objectives of benchmarking and comparison of methods.

Chapter 4 presents the second proposed approach to model and disaggregate continuously varying loads using a quantized CS-HMM and a framework for the hybrid CS/DS-FHMM to apply on load disaggregation. Theoretical parts of the proposed approach are augmented with explanations on how the proposed approaches can mitigate potential shortcomings that hamper reference methods. In addition, case study applications on synthetic data and real measurements of home appliances from available public data set were used to carry out models learning, testing and validation tasks. Reference methods were also applied on same data for objectives of benchmarking and comparison of methods.

Chapter 5 presents the third proposed approach which aims to disaggregate overlapping clusters of home appliances which result when applying a clustering method to NILM. An approach for clusters splitting is proposed based on investigation of clusters cohesion. Theoretical parts of the proposed approach are augmented in detail. In addition, case study applications on real measurements of home appliances from available public data set were used for testing and validation.

Chapter 6 summarizes conclusions and significant findings concluded from the research work and presents recommendations for potential future work.

Chapter 2

Background and Related Work

This chapter presents background and literature reviews on related research works that targeted the problem of non-intrusive load disaggregation. It highlights the first work and the general framework for an approach of NILM. Commonly utilized appliances features, load identification approaches and accuracy metrics are articulated. In addition, an overview on hidden Markov models (HMMs), extensions of HMMs and basic clustering approaches is presented.

2.1 Intrusive Load Monitoring

Intrusive load monitoring requires sub-metering of individual appliances by installing appliance-dedicated sensors that report or log their individual operation and power consumption. Though this method can provide accurate power consumptions by respective appliances, it apparently incurs additional burdens such as

- Added costs associated with installation of power sensors for each home appliance.
- Concerns about occupant's privacy due to the intrusive nature of this method.
- Extra care needed whenever household occupants add or drop an appliance.

Therefore, intrusive load monitoring is often considered an impractical solution especially when it is supposed to be widely deployed for large scale users [Kim *et al.*, 2010; Parson, 2014].

2.2 Non-Intrusive Load Monitoring

Non-intrusive load monitoring (NILM) utilizes only the total aggregate signal (i.e. measurements by a smart meter) at a household level to extract distinct appliances features in order to deduce the operation and power consumptions of individual appliances present in the household. The concept and first work on NILM was proposed by Hart in 1992 [Hart, 1992]. His research showed the possibility to detect and infer appliances operation from the aggregate signal basically by detecting ON/OFF events (transitions edges) as shown in Figure 2.1 [Zoha *et al.*, 2012].

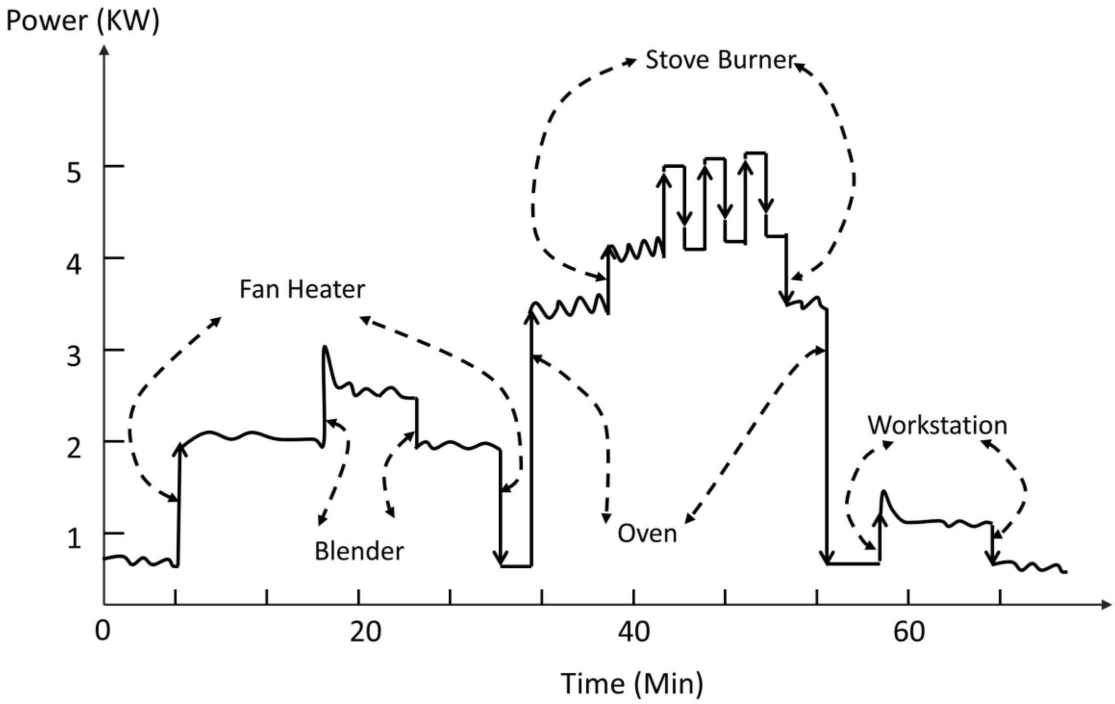

Fig. 2.1: Aggregate signal showing appliances ON/OFF events [Zoha *et al.*, 2012].

Hart's first work of load disaggregation consisted of two subsequent steps. First, detecting events when a device switches ON/OFF and extracting useful features (e.g. real and reactive power consumptions). Thereafter, these distinct features (devices signatures or attributes) were grouped in clusters that represent different appliances. The work showed that high power consuming devices were easier to be classified as being placed far from each other. Conversely, there were overlapping regions of low power consuming devices, which encouraged further research on efficient methods for NILM.

Hart [Hart, 1992] named his first work in NILM as non-intrusive appliance load monitoring (NIALM). Throughout literature, some other alternate terms were used to describe the problem of NILM such as energy disaggregation and load disaggregation.

Progression in the domain of NILM was usually by targeting improvements in either or both of the two following paths [Zeifman and Roth, 2011]

- Extracting extra useful features from the total aggregate signal that can lead to distinguish between household appliances.
- Enhancing the identification methods which would lead to improved performance by the proposed NILM approaches.

To simplify research work on NILM, home appliances are usually categorized into four main categories as follows [Zeifman and Roth, 2011]

- Type I appliances, or always ON appliances, which are always active and operating in the ON state. These devices never switch to the OFF state and, hence, they represent the baseload of the household. Smoke detectors and home internet routers are typical examples of this appliances category.
- Type II appliances, or ON/OFF appliances, which can be either OFF or operating at only one possible ON state. These appliances consume power consistently around a specific level. Lights and toaster represents examples of these category of appliances.
- Type III appliances, or multi-state appliances, which are also called finite state machines (FSMs), are appliances that can operate at a finite number of possible states. These appliances are expected to consume different amounts of power

according to the state of operation. Examples of this category include ceiling fans and washing machines.

- Type IV appliances, or continuously varying loads, which may consume power at infinite number of states during their operation. Light dimmers, power tools and electronic devices are typical examples of this appliances category.

2.2.1 Appliances Features for Load Disaggregation

Extracting distinct features (also known as attributes or signatures) of appliances is an essential driving factor to achieve accurate load disaggregation. In general, a feature is valuable given that

- The feature is measurable from the total aggregate signal when the respective appliance is active (ON).
- It can be used to distinguish between two or more of individual appliances.
- It remains consistent and does not change with time or other factors.

The type and number of extractable features from the total aggregate signal in the domain of NILM depends mainly on the frequency of the measurement equipment. In addition, some extra non-electric features from ambient facts can be utilized to assist in the load disaggregation process.

2.2.1.1 Low Frequency Measurements Features

Common installed smart meters (or to be installed with the deployment of smart grids) measure a household's total aggregate consumptions at low frequencies such as 1 Hz in USA [Kolter and Johnson, 2011] and 0.1 Hz in UK [Parson, 2014]. Therefore, using available smart meters measurements is reliable and cost effective in the domain of NILM since it does not require installations of additional metering equipment. By means of low frequency equipment, NILM approaches can only use relatively macroscopic features of appliances [Zeifman and Roth, 2011].

The first work on NILM [Hart, 1992] utilized changes in both real and reactive power draws from the aggregate consumption signal. These changes were used to construct a two-dimensional signatures space of real and reactive power ($\Delta P - \Delta Q$ plane), thereby to group appliances in distinct representative clusters.

Albicki and Cole [Albicki and Cole, 1998] have extended the previous mentioned NILM method with the use of *edges* and *slopes* as appliances extra features. These two new features are defined, respectively, as the initial upward spike in power and the slower changing variation that occurs during the turn ON events. The method is specifically appropriate for appliances characterized by significant spikes in power draw such as heat pumps, dishwashers, and refrigerators.

Baranski and Voss [Baranski and Voss, 2003] proposed a *histogram-based* approach that analyzes occurrences of power transitions. Only frequent events (ON/OFF transitions) were considered for analysis. Thereafter, they applied an optimization method to search for best matches between events and home appliances.

2.2.1.2 High Frequency Measurements Features

Measuring the household total aggregate consumptions at high frequencies provides data that is rich in information and fine features such as harmonics, transient power profile, electromagnetic interference (EMI), etc. However, in most cases, high frequency measurements require installation of additional metering equipment which means added cost and time disadvantages [Zoha *et al.*, 2012].

Analysis of harmonics obtained by means of Fourier transform (FT) helped in distinguishing between appliances such as computers and incandescent light bulbs

[Zeifman and Roth, 2011]. Spectral envelope of harmonics, which is a vector of coefficients of harmonics, was found to be a characteristic feature for continuously varying loads [Lee *et al*., 2005; Wichakool *et al*., 2009]. Nonetheless, detecting the presence of a varying load is not sufficient for NILM since estimation of the power consumption is the crucial goal of a NILM approach [Zeifman and Roth, 2011]. Akbar and Khan [Akbar and Khan, 2007] considered a method to analyze harmonics based on short-time fast Fourier transform (STFFT) of transient signals. The method was developed to identify some non-linear devices (e.g. power electronics appliances) which usually produce rich harmonics content. In comparison with Fourier transform, the wavelet transform was used to characterize the transient behavior of the loads. Using wavelet transform, Chang [Chang, 2012] showed that the transient response time and the transient energy features are better than steady-state features for the task of loads disaggregation. However, the study considered selected appliances with distinct turn-on attributes [Zoha *et al.*, 2012].

The geometrical shape of the current-voltage (also called the $I - V$ trajectory) was used to separate appliances based on their categories [Lee *et al.*, 2004; Lam *et al.*, 2007]. Lam *et al.* [Lam *et al.*, 2007] utilized the $I - V$ trajectory to decompose appliances into eight groups with high accuracy, providing further sub-divisions within each group.

Patel *et al.* [Patel *et al.*, 2007] proposed to monitor electric noise in a socket for transient signals and then to use its fast Fourier transform (FFT) as a feature. A training phase was required for each appliance and their possible combinations. Results in terms of detection accuracy were found comparable to those of other methods [Zeifman and Roth, 2011]. However, computational expense of capturing and analyzing transient noise, the necessity for training for each household device and the dependence of the obtained signatures on household wiring were some obvious shortcomings of using the noise FFT as a feature [Gupta *et al.*, 2010; Zeifman and Roth, 2011].

Gupta *et. al.* [Gupta *et al.*, 2010] proposed an approach showing that appliances equipped with switch mode power supply (SMPS) can be characterized by analyzing the steady-state voltage noise generated due to their operation, i.e., electromagnetic interference (EMI). They utilized Fourier features of the EMI signals in the 36-500 kHz range. The signatures were found to be distinct for different appliances. While this feature can easily distinguish between motor-based appliances (as they generate voltage noise), it remains sensitive to wiring architecture, EMI signatures overlap with each other, and not all appliances are equipped with SMPS technology [Zoha *et al.*, 2012].

2.2.1.3 Non-Traditional Appliance Features

Beside appliance electrical features, some non-traditional features were proposed to assist in the task of NILM. Examples of these features include [Kim *et al.*, 2010]

- The duration of use of an appliance, especially for appliances with repeatable operation pattern (operation cycle) such as refrigerators and washing machines. This feature is less likely to be helpful for devices fully dependent on occupants' behavior such as lights, TV, electronics, etc. However, if home occupants show some usage pattern of an appliance, this may be learnt over time by a NILM approach.
- The time and day of which the appliance was used, which depends on the usage pattern and behavior of home occupants. For example, some occupants tend to use specific appliances in the evening time or on weekends.
- The operational relation (inter-dependency) between two or more appliances. For example, switching ON some entertainment devices may indicate the probability to switch ON the LCD screen.

- The ambient weather (e.g. temperature) information. For example, the air conditioner is likely to be used when there is a high ambient temperature.

2.2.2 Approaches for Load Identification

After collecting useful appliances features, an approach for load identification is required to detect and identify the operation of the corresponding appliances. In addition, the main objective of NILM is to estimate the individual appliances power consumptions as accurate as possible merely by using the total aggregate signal. Since the main goal of NILM is to provide estimations of power consumed by individual appliances, it becomes essential to achieve accurate estimations by a proposed NILM method. Hence, the impact of NILM and potential energy saving become achievable, given the estimation of individual appliances consumptions is as accurate as possible. The problem of NILM falls under the general blind-source separation class of problems. A common example in this field is the speakers' decomposition (or who-spoke-when) problem; where a single audio recording of a meeting is given, and we wish to infer the number of speakers present, when they speak, and some characteristics governing their speech patterns [Johnson and Willsky, 2013]. According to the method of appliances models learning, a NILM approach can generally fall in either *supervised* or *unsupervised* learning category.

2.2.2.1 Supervised Learning Approaches

In *supervised* learning approaches, there is a learning phase in which individual appliances models are built and other features are extracted from appliance-level consumption measurements. To acquire appliance-level measurements, additional metering equipment are usually required, which obviously incur extra installation cost and efforts. Records of appliances (their consumption profiles or models, and other extractable features) are hence recorded in a database for use of appliances matching in the disaggregation process. Thereafter, the NILM approach performs the process of load disaggregation on the total aggregate signal with the goal of detecting the operation of individual appliances (e.g. when an appliance switches ON/OFF). To apply a *supervised* learning approach, two broad methodologies were used to carry out the load disaggregation process: optimization and pattern recognition [Zoha *et al.*, 2012].

The core task of optimization methods in the context of NILM is to search for an optimal combination of known appliances (which were recorded in a database during the learning phase) that would minimize the error between the observed and the estimated total aggregate consumption. Various optimization methods such as integer programming and genetic algorithm were used to search for optimal solutions for NILM [Baranski and Voss, 2003; Liang *et al.*, 2010; Baranski and Voss, 2004; Suzuki *et al.*, 2008]. However, these methods increase in complexity and computational time as the number of existing appliances increase. In addition, it becomes challenging to detect appliances that were added to the household after the appliances models were learnt and their database was recorded.

On the other hand, pattern recognition methods construct distinct patterns or models for individual appliances so as to be used in the disaggregation process by detecting the most likely appliance patterns from the total aggregate signal that can be matched to recorded patterns of appliances. Development of patterns recognition techniques are hot topics in the scope of machine learning and other related fields in the broad scope of artificial intelligence. Marchiori *et al.* [Marchiori *et al.*, 2011] used a Bayesian approach to infer the state of home appliances using the power features and state-switching information. The temporal information along with power transitions were found effective in appliance detection [Zeifman and Roth, 2011; Lin *et al.*, 2010]. Hence, hidden Markov models (HMMs) and their extensions were used to model and disaggregate the operation of home appliances [Kim *et al.*, 2010; Kolter and Jaakkola, 2010]. Artificial neural networks (ANN) has also shown good performance for the task of load disaggregation [Ruzzelli *et al.*, 2010]. Support vector machines (SVM) methodology was also used with utilization of features such as harmonics and other low frequency signatures and showed good performance in appliances classification [Srinivasan *et al.*, 2006; Kato *et al.*, 2009; Lin *et al.*, 2010]. Lai *et al.* [Lai *et al.*, 2012] proposed to utilize a gaussian mixture model (GMM) to describe the distribution of currents waveforms and then to use a SVM to classify individual loads.

2.2.2.2 Unsupervised Learning Approaches

Since *supervised* learning approaches often require extra equipment and labor costs to acquire appliance-level measurements, models and features, research direction toward *unsupervised* learning and inference methods gained increasing attention recently [Kim *et al.*, 2010; Parson, 2014; Zoha *et al.*, 2012]. An *unsupervised* learning approach does not require measurements or models of individual appliances in prior. Alternatively, appliances models and other features are learnt online from the total aggregate signal or using some historical measurements of the total aggregate signal of the same household. This online modeling process is usually carried out by applying a set of probabilistic analysis and classification methods [Kim *et al.*, 2010; Kolter and Jaakkola, 2010].

In brief, the basic advantages of *unsupervised* learning approaches compared to *supervised* learning approaches are

- Reduction of overall NILM system deployment expense, as there is no need to install extra metering equipment or features extraction at appliance-level in prior.
- When a detected appliance in the disaggregation process found to be not matching to any previously known appliances, it can be considered as an added new appliance to

the household. Hence, its model and features should be added as a new record of the known appliances database. Thereafter, it becomes possible to be detected once it turns in operation.

However, appliances modeling and feature extraction from the total aggregate signal introduce some noteworthy challenges such as

- Analysis methods should be capable of distinguish between models that belongs to different appliances. Therefore, the applied statistical analysis or machine learning techniques may need to be equipped with some rules to differentiate between individual appliances and their models/features.
- Since the *unsupervised* learning approaches results in unlabeled groups or clusters (i.e. appliances), a manual labeling step is required by a field expert in order to label the NILM outcomes with the respective home appliances [Parson, 2014].

There are several *unsupervised*-fashioned methods and frameworks that were applied in the domain of load disaggregation. Parson *et al.* [Parson *et al.*, 2014] proposed an approach in which a one-time *supervised* learning process with already labeled data set was used to create general probabilistic models of appliances. Thereafter during the load disaggregation process, these general models can be tuned to previously unseen households in an *unsupervised* manner. The tuning process resulted in specific models of devices that were existing in the tested households. Goncalves *et al.* [Goncalves *et al.*, 2011] applied an *unsupervised* blind source separation technique to obtain appliance-level consumptions from the aggregate data. They utilized both genetic K-means and agglomerative clustering with features like $\Delta P - \Delta Q$ to cluster appliances. Shao *et al.* [Shao *et al.*, 2012] proposed a motif mining method to perform load disaggregation in an *unsupervised* fashion. Their method was mainly applicable to appliances with distinctive repeatable events. Kamoto *et al*. [Kamoto *et al*., 2017] presented a new approach based on competitive agglomeration (CA) which incorporates the good qualities of both hierarchical and partitional clustering aiming to carry out energy disaggregation to discover appliances without prior information about the number of appliances.

Hidden Markov models (HMMs) and their extensions were also used in *unsupervised* manners to achieve the task of load disaggregation. Kim *et al.* [Kim *et al.*, 2010] applied four extensions of the HMMs to the problem of NILM. They utilized some non-traditional features, such as the duration of use and time of use, in order to enhance the performance of the proposed models. Kolter and Jaakkola [Kolter and Jaakkola, 2010] presented a new approximate inference method based on FHMM. They applied two complementary models, the additive and the difference FHMM. The additive FHMM captures well the total aggregate output signal while the difference FHMM encodes the signal differences between subsequent power levels (when a device switches ON or OFF). Henao *et al.* [Henao *et al.*, 2018] targeted to the disaggregation of the electric space heater by applying a robust disaggregation approach based on the difference factorial hidden Markov model (DFHMM) that uses their common prior knowledge. Johnson and Willsky [Johnson and Willsky, 2013] introduced a hierarchical Dirichlet process hidden semi-Markov model (HDP-HSMM) where state durations are explicitly defined. The methods introduced also provide new techniques for sampling inference in the finite Bayesian HSMM. The proposed model was applied to synthetic data and to selected home appliances from available public data set. Lu *et al.* [Lu *et al.*, 2017] proposed an event-based detection algorithm that apply a simplified Viterbi algorithm, which considers fewer state transitions each time than the traditional Viterbi.

2.2.3 Public Data sets for Load Disaggregation

Research on load disaggregation requires measurements of both total aggregate signal at household-level and of measurements at appliance-level. Measurements at appliance-level (or sub-metering) are useful to build appliances models in *supervised* learning approaches and they are also important for validation and judging the accuracy of outcomes of the proposed NILM approach. Hence, some data sets of detailed measurements of real houses were published and made available for public to facilitate and expedite research work on load disaggregation. In addition, public data sets can be used for benchmarking purposes, i.e., to compare the performance of two or more approaches by testing on the same data.

One of the largest and commonly used public data sets is the Reference Energy Disaggregation Data set (REDD) [Kolter and Johnson, 2011]. The REDD includes both low and high frequency measurements of six real houses for an approximate period of two weeks. While REDD satisfies most research needs, it was noted that the summation of individual consumptions is not always equal to total household consumption at the same time instant. This indicates existence of some unmeasured or excluded devices. Moreover, the devices are not grouped to be connected to either mains 1 or mains 2 phase circuits (the main feeder lines). In this work, the REDD data set was used to apply and test the proposed approaches.

A number of other data sets were made available for public to test their NILM approaches, but they differ from many aspects such as: number of monitored houses and appliances, frequency of metering, etc. For illustrations, Table 2.1 from [Bonfigli *et al.*, 2015] shows a brief comparison between a number of available data sets collected and prepared to test NILM approaches.

Table 2.1: Comparison of available data sets for NILM [Bonfigli *et al.*, 2015].

Data Set	Location	Appliance Sample Resolution	Aggregate Sample Resolution	Reference
REDD	USA	3 seconds	1 second and 15 kHz	[Kolter and Johnson, 2011]
BLUED	USA	State transitions label	12 kHz	[Anderson *et* al., 2012]
UMass Smart	USA	1 second	1 second	[Barker *et* al., 2012]
AMPds	Canada	1 minute	1 minute	[Makonin *et al.*, 2013]
BERDS	USA	20 seconds	20 seconds	[Maasoumy *et* al., 2014]

For more information and comparisons of available public data sets, see e.g. [Parson, 2014; Bonfigli *et al.*, 2015].

2.2.4 Evaluation of Load Disaggregation Approaches

Though load disaggregation approaches have the same goal of obtaining accurate estimations of individual appliances consumptions, there were different accuracy metrics used in literature to assess the performance of an approach. The following accuracy metrics were commonly used in research to evaluate a proposed approach

1. The accuracy of power estimated of individual appliances compared to their actual power consumptions over the tested time period. This accuracy metric for each home device, Acc_m, can be given as follows [Kolter and Johnson, 2011; Johnson and Willsky, 2013]

$$Acc_m = 1 - \frac{\sum_{t=1}^{T} |\hat{p}_m(t) - p_m(t)|}{2\sum_{t=1}^{T} p_m(t)} \tag{2.1}$$

where $p_m(t)$ is the true power consumed at time t by the device m, $\hat{p}_m(t)$ is the estimated power consumption at time t by the device m and T is the time length of the sequence used for the disaggregation process. This accuracy metric can also be given, as a unified accuracy across all monitored devices in a household, as follows

$$Acc = 1 - \frac{\sum_{t=1}^{T} \sum_{m=1}^{M} |\hat{p}_m(t) - p_m(t)|}{2\sum_{t=1}^{T} p(t)} \tag{2.2}$$

where $p(t)$ is the total aggregate power at time t, which is consumed by the existing M home devices. Since this accuracy metric is based on the accuracy of power estimation of individual device, which is the principal goal of load disaggregation, it will be used to assess the proposed approaches in Chapter 3 and Chapter 4 of this work. The division by 2 in equations (2.1) and (2.2) comes from the fact that using the absolute value will double count the errors [Kolter and Johnson, 2011; Johnson and Willsky, 2013].

2. Using *Precision*, *Recall* and their extensions. These metrics are used often to assess the accuracy of assignments of events or energy to individual appliances. In this metric, detected events are categorized into four groups based on reality and concluded outcomes as shown in Table 2.2. In brief, *TP* and *TN* are the counts of true assignments and true non-assignments of events with respect to a specific appliance. The *FP* indicates the occurrences of wrong assignments of events to a particular appliance. This should not be confused with the *FN*, which indicates the occurrences of wrong non-assignments of events to that particular appliance.

Table 2.2: Actual versus concluded events.

Test conclusion	Real event occurred	Real event did not occur
Detected	True Positive (*TP*)	False Positive (*FP*)
Not detected	False Negative (*FN*)	True Negative (*TN*)

Precision is a measure that describe the relevance ratio of the detected elements (percentage of truly detected events for an appliance from all detected events) and it is given by

$$Precision = \frac{TP}{TP + FP} \tag{2.3}$$

Recall is a measure of how many of relevant elements (events actually occurred by an appliance) were detected and it is given by

$$Recall = \frac{TP}{TP + FN} \tag{2.4}$$

F-Measure is a commonly used measure that combines *Precision* and *Recall* as the harmonic mean and it is given by

$$F = 2(\frac{Precision \times Recall}{Preciosion + Recall}) \tag{2.5}$$

For more details about *Precision*, *Recall*, *F-measure* and some other related metrics of assignments accuracy, refer e.g. to [Powers, 2011].

To encompass the impact of appliances modeling (i.e. building power consumptions profiles by the approach) in these accuracy metrics, Kim *et al.* [Kim *et al.*, 2010] suggested to divide the TP category into two sub-categories: Accurate True Positive (ATP) and Inaccurate True Positive (ITA), which were defined as follows

$$TP\begin{cases} ATP, & where\ \dfrac{|\hat{p}-p|}{p} \le \rho \\ ITP, & where\ \dfrac{|\hat{p}-p|}{p} > \rho \end{cases} \tag{2.6}$$

$$Precision = \frac{ATP}{ATP + ITP + FP} \tag{2.7}$$

$$Recall = \frac{ATP}{ATP + ITP + FN} \tag{2.8}$$

where $\hat{p}$ refers to the estimated power consumption of a specific device and p is its actual power consumption. Usually, power consumptions fluctuate within 20% of the average power, thus a ρ threshold can be taken as 0.2 [Kim *et al.*, 2010].

Parson [Parson, 2014] used receiver operating characteristic (ROC) curves to trade-off between True Positive Rate (TPR), which is same as the *Recall* measure, and its complementary: False Positive Rate (FPR).

3. Liang *et al.* [Liang *et al.*, 2010] differentiated between the detection and classification accuracy as follows: if detected events include all *true* and *false* detections, then detection accuracy is the percentage number of those *classified* correctly from detected events. Classification accuracy is the percentage of number of those *classified* correctly from *truly* detected events. Overall accuracy is the percentage number of those classified correctly to events that actually occurred.

2.3 Hidden Markov Models and Their Extensions

Hidden Markov models (HMMs) and their extensions are useful stochastic models that were applied on various scientific problems. As they were used in research work on load disaggregation (including this work), HMMs and their common extensions (or variants) are presented here in brief. In their basic structure, HMMs model a sequence of observations (or emissions) that is evolving due to another hidden (or latent) sequence of states [Bishop, 2006]. The hidden sequence of states cannot be observed directly but may be inferred from the sequence of observations. In brief, the ingredients of a typical HMM (denoted as λ) are as follows

$$\lambda = \{\pi_i, A, b_j\} \tag{2.9}$$

where π_i represents the likelihood of the initial state of the model (i.e. the hidden state s_t at time $t = 0$), the transition matrix A provides the probabilities of which the model to switch from a state i at time $t - 1$ to state j at time t, thus each element in in the transition matrix A, $a_{i,j}$, gives the following transition probability

$$a_{i,j} = P(s_t = j | s_{t-1} = i) \tag{2.10}$$

The observation likelihood b_j give the likelihood of an observed item x_t at time t. Thus, $b_j(x_t)$ is the likelihood to observe the item x_t at time t for some hidden state j. While observations likelihood may be discrete or continuous distribution, the above configuration is valid for cases of discrete or finite number of hidden states. Once the HMM has been set up (i.e. the model parameters have been leant), there are two (among others) basic and useful tasks that can be carried out using the modeled HMM

- Inferring the maximum likely sequence of hidden states, given a sequence of observations. This task can be done using the dynamic optimization by Viterbi algorithm [Bishop, 2006; Parson, 2014]. The Viterbi algorithm is a dynamic optimization technique that aims to select the best possible hidden sequence of states that provides maximum likelihood sequence, give the sequence of observations for a given modeled HMM. Various examples of applications in this domain include the wide applications in speech and patterns recognition, bioinformatics application, energy disaggregation, etc.

- Prediction of possible future observations, given current sequence of observations and states. This task can be done, for example, by the forward algorithm [Zucchini *et al.*, 2016]. Various examples of applications in this domain include computational finance, time series analysis and other forecasting problems.

2.3.1 Factorial Hidden Markov Models and Other Extensions

Factorial HMMs (FHMMs) [Ghahramani and Jordan, 1997] are an extension to the standard HMM with belief that there are several hidden states chains evolving independently and in parallel producing the measurable sequence of observations as shown in Figure 2.2.

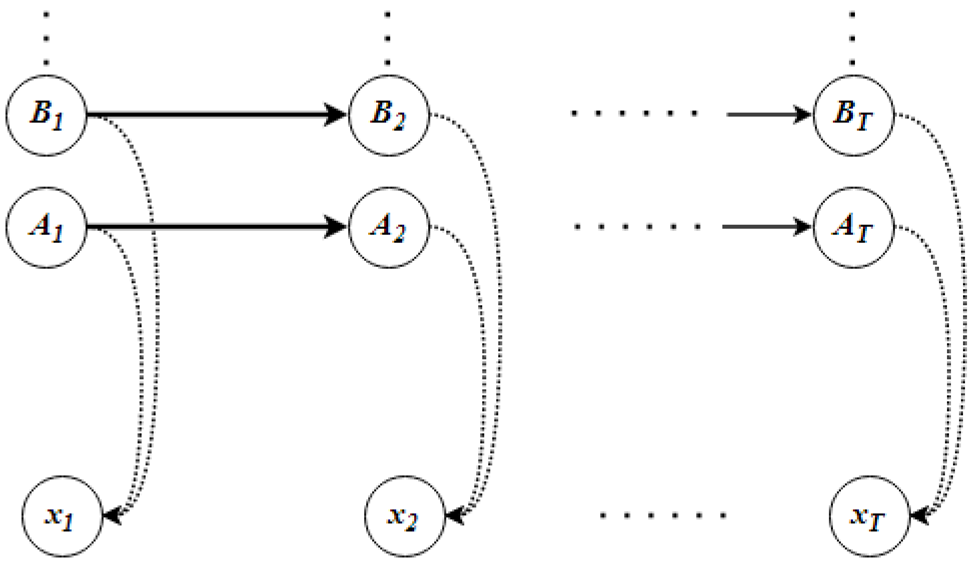

Fig. 2.2: The basic structure of FHMM.

Figure 2.2 shows the basic structure of FHMM where a sequence of observed emissions $X = \{x_1, x_2, \dots x_T\}$ are evolving over corresponding time $t = 1, 2, \dots T$ due to several independent parallel hidden state chains or factors denoted as $A, B, \dots$ etc. The hidden states chains evolve independently from each other. For example, the hidden state chain A in Figure 2.2 would have its own transition matrix that consists of elements $a_{i,j} = P(A_t = j | A_{t-1} = i)$ to give likelihoods that states of hidden chain A switch from state i

at time $t-1$ to state j at time t. The observed output (emissions chain) is usually a joint discrete or continuous function of all hidden states [Kolter and Jaakkola, 2012]. That is,

$$X = f_{A,B,\dots}(A, B, \dots) \tag{2.11}$$

The configuration of FHMMs gained increasing interest in the domain of load disaggregation because it is suitable to model existing individual home appliances as Markov models evolving concurrently resulting in the observed total aggregate power as the summation of individual consumptions of individual home devices.

In addition to the FHMM, there are a number of developed extensions of HMM that may be considered according to the case study and the available information in prior. Examples of these extensions of the HMM include the following

- Hidden semi-Markov model (HSMM) [Yu, 2010; Johnson and Willsky, 2013], where the hidden states are supposed to occupy some time durations that could be defined by a specific probability distribution. In the context of NILM, this model could be useful if durations of appliances operation are known to have repeatable patterns or fit into some probability distribution.
- Conditional FHMM (CFHMM), where the transition probabilities are not constant but are conditioned on the extra features. Although the CFHMM is similar to a coupled hidden Markov model (CHMM), CFHMMs are more generic as they consider the dependencies between hidden state sequences and the additional input sequences [Kim *et al.*, 2010].
- Conditional Factorial HSMM (CFHSMM), which combines both the CFHMM and the HSMM [Kim *et al.*, 2010].

- Input-Output HMM, where additional information that might have impacted upon the hidden variables' states can be integrated into the model at a little extra computational cost [Parson, 2014].

2.3.2 Continuous-States Hidden Markov Models

The structure of the common discrete-states HMM (DS-HMM) consists of a sequence of hidden finite number of states (e.g. N number of states). Thus, it is reasonable to model transitions probabilities by a $N \times N$ matrix; the transitions matrix. A continuous-state HMM (CS-HMM) is sought to model the cases of infinite number or continuous space of hidden states (denoted as Z in the settings of continuous space of hidden states). The essential challenge of complexity arises when transitions between states are difficult to be modeled by a known joint continuous likelihood function such that [Ainsleigh, 2001]

$$P(s_t = z_t | s_{t-1} = z_{t-1}) = f_{Z,Z}(z_t, z_{t-1}) \tag{2.12}$$

where $f_{Z,Z}$ is a joint continuous function that represents likelihoods of transitions in the continuous space from state z_{t-1} at time $t-1$ to state z_t at time t. In case the continuous transitions probability likelihood is unavailable (which is common in real-life applications), approximating methods such as nonparametric statistics or quantizing the continuous states range techniques may be employed.

2.4 Bayesian Nonparametric Models and Dirichlet Process

Nonparametric models are used to model selection and adaptation of the studied problem where the sizes of models can grow progressively with data size. This is opposed to parametric models, which use a fixed number of parameters [Orbanz and Teh, 2011]. The Dirichlet process (DP) [Teh, 2011] is a stochastic process used in Bayesian nonparametric models of data, particularly in DP mixture models (which is also known as infinite mixture models). It is a distribution over distributions which means that each draw from a DP is itself a distribution.

Teh *et al.* [Teh *et al.*, 2005] proposed a hierarchical DP approach where they assumed that the number of mixture components is unknown a priori and is to be inferred from the data. Thus, they considered a set of DPs, one for each group, where the well-known clustering property of the DP provides a nonparametric prior for the number of mixture components within each group.

Johnson and Willsky [Johnson and Willsky, 2013] introduced the hierarchical Dirichlet process hidden semi-Markov model (HDP-HSMM) where state durations are explicitly defined. The methods introduced also provide new techniques for sampling inference in the finite Bayesian HSMM. These models can be used as tools in constructing and performing inference in larger hierarchical models. They applied their model on synthetic data and selected home appliances from the REDD public data set.

These methods are related to the CS-HMM or clustering techniques in the sense that the possible number of states (or clusters) is unknown in prior and grow progressively according to the available data (often, when a data stream or time series of data is the main supplier to build these clusters/states). Though the number of states (or clusters) is possibly infinite theoretically, the number of detected states (or constructed clusters) are usually countable and determined mainly by the available data.

2.5 Common Clustering Approaches

Clustering is a broad concept in the domain of artificial intelligence in which heterogenous data is grouped into distinct clusters based on significant features or attributes [Jain *et al.*, 1999]. Clustering is a reasonable approach when applied to data that is believed to have originated from different sources and later merged in a data stream or time series. In the domain of NILM, for example, consumptions of individual home appliances are lumped together in the form of the total aggregate signal that is usually metered by the household smart meter. In general, a clustering method is supposed to result in clusters consisting of items that are alike by some means and differ from those in other clusters.

Clustering methods were used in NILM techniques, but they were usually efficient in identifying non-overlapping and high power consuming devices [Zeifman and Roth, 2011]. In [Farinaccio and Zmeureanu, 1999], the clustering approach was extended by filtering and smoothing mechanisms to deal with power variations. Nonetheless, this method requires excessive training and it was applicable mainly to high power loads [Zoha *et al.*, 2012].

2.5.1 Overview of Basic Clustering Methods

Clustering methods aim to group items in the data stream or time series measurements into a distinct number of clusters. However, the performance of a clustering method depends on many factors such as: number of originating sources that have merged in the data stream, number and nature of extractable features, the dimensions of features, etc. Since clustering approaches were used in the domain of load disaggregation, an overview of basic clustering methods such as nearest neighbor (NN), K-means, fuzzy K-means and expectation maximization (EM) is presented in brief.

The nearest neighbor (NN) clustering is a method to assign unlabeled items to clusters according to their distance from items that has been previously assigned to clusters. An iterative procedure of NN clustering it assigns each unlabeled item to the cluster of its nearest labeled neighbor item, provided the distance to that labeled neighbor is below a threshold. The process continues until all items are labeled or no additional labeling occur [Jain *et al.*, 1999].

The K-means clustering approach assign items to clusters based on their distance from the clusters centers. The K-means method firstly initialize items randomly to clusters, so each cluster has a center in the features space. Then, the K-means method iteratively run the following two steps

- Update clusters centers using all items assigned to the corresponding clusters.
- Assign items to clusters whose center has the minimum distance from each investigated item.

The above steps are repeated till some termination criteria met. The termination criteria could be, for example, a maximum number of iterations, relative change in clusters centers, or other enhancement measures [Jain *et al.*, 1999]. It is noticeable that, since it is a distance-based assignment method, using K-means method for clustering is reasonable in cases where no information is available in prior about the distribution fitting of the targeted clusters. A basic shortcoming of K-means method is that it requires to know the number of distinct clusters prior to execution of the clustering process.

The fuzzy K-means approach is similar to the K-means approach, except that each item is not strictly assigned to one cluster in the fuzzy K-means approach. Instead, a membership function is used to present how likely an item belongs to each cluster.

The expectation maximization (EM) is an iterative clustering method that assigns items to clusters based on their probability likelihoods (not based on a distance measure). The EM is also commonly used for parameters estimation from data. If items in one cluster are believed to follow some probability likelihood (e.g. normal distribution), then using the EM clustering method is an efficient way to perform the clustering task [Jain *et al.*, 1999; Bishop, 2006]. Analogous to the K-means method, the EM method iteratively runs the following two steps

- Update likelihood functions using all items assigned to the corresponding clusters.
- Assign items to clusters based on achieving maximum likelihood of each item to be belonging to the investigated cluster.

The above steps are iterated till some termination criteria met such as a maximum number of iterations or relative change in some enhancement measures [Jain *et al.*, 1999]. It is noticeable that, since it is a likelihood-based assignment method, using EM method for clustering is reasonable in cases where information is available in prior about the distribution fitting of the targeted clusters. Same as in K-means clustering, a basic shortcoming of the EM method is that it requires to know the number of distinct clusters prior to execution of the clustering process.

2.5.2 Clusters Splitting and Merging

In real world applications, clustering may not be a straight forward process. Many challenges may arise such as overlapping between features from different clusters, difficulties in extracting significant features, limitations on number of available features, noisy measurements environment, etc. Therefore, the following two cases may occur

- Two or more clusters that represent different sources has merged in one cluster. In such cases, cluster splitting could be reasonable to retrieve the right clusters.
- A cluster has been retrieved as two or more clusters due to some dissimilarity measures or dispersion among items. In such cases, clusters merging could be reasonable to retrieve the right cluster.

Unfortunately, to decide splitting or merging on clusters is a problematic decision that highly depend on the available information and the case under study. Wagstaff *et al.* [Wagstaff *et al.*, 2001] proposed a *constraint-based* clustering which modifies the K-means method considering a set of constraints to be satisfied. Lu and Leen [Lu and Leen, 2007] used a mean field approximation to produce the *Penalized Probabilistic Clustering* (*PPC*) to handle increasing complexity in large data sets. Instead of using item-to-item constraints as in *PPC*, the *Class-Level Penalized Probabilistic Clustering* (*CPPC*) was proposed by defining cluster-to-cluster constraints [Preston *et al.*, 2010]. These constraints provide probabilities about how likely two clusters should be merged, split or kept unchanged. As the number of clusters is often significantly less than the total number of items to be classified, *CPPC* provides a noticeable reduction in complexity over *PPC* [Preston *et al.*, 2010].

Hierarchical divisive clustering is a top-to-bottom clustering that repeatedly partition a present cluster into two smaller clusters till reaching some stopping criteria. Common approaches in considering such splitting decisions are size priority, cluster cohesion tests and dissimilarity between the possible sub-clusters [Ding and He, 2002]. It is notable that size priority method to split clusters requires a prior knowledge about the size (e.g. total number of items) of the clusters. On the other hand, cluster cohesion tests require to adopt a cluster cohesion metric (e.g. a particular pattern or a distribution fitting). Dissimilarity between clusters could be indicated by how far the clusters are in terms of some defined features in the features space.

2.6 Summary

This chapter presented an in-depth literature overview on related research works in the field of non-intrusive load disaggregation. Significant extractable appliances' features and load identification approaches investigated in previous works were presented. In addition, a brief outlook was given about HMMs, extensions of HMMs and clustering techniques, as these models and methods were further enhanced in the proposed approaches in this thesis. To articulate the presented clustering methods, Table 2.3 compares between DP, K-means and EM methods.

Table 2.3: Comparison between DP, K-means and EM methods.

Clustering Method	Mechanism	Number of Distinct Clusters
DP	Nonparametric methods	Variable, grows with data
K-means	Distance-based	Fixed, must be known in prior
EM	Likelihood-based	Fixed, must be known in prior

Chapter 3

Enhanced Load Disaggregation Using Information on Appliances Interactions

This chapter presents the first proposed approach that aims to enhance the overall performance of non-intrusive load disaggregation. To improve the load disaggregation accuracy, two-way interactions between appliances is introduced as a new feature that can be embedded in the factorial hidden Markov model (FHMM) representations of home appliances and total aggregate signal. In addition, this Chapter presents in-depth interpretations of mutual devices interactions from power quality aspects. Then, a statistical factorial design is adopted to model devices and estimate their primary power consumptions and possible two-way interactions. Thereafter, information on devices interaction is embedded in the FHMM representation of the total aggregate signal so as to carry out the task of non-intrusive load disaggregation.

Since home appliances are often connected in parallel to the same main feeder line, it is possible that the operation of one appliance may affect the power consumption of other appliances to some extent due to several issues of power quality. Figure 3.1 shows a typical connection of two devices connected in parallel to the main feeder line in a typical residential household.

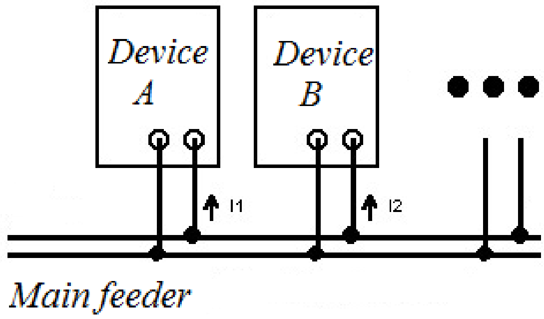

Fig. 3.1: Parallel connection of devices in residential households.

The connection of home devices as shown in Figure 3.1 allows some power quality issues to arise such as harmonics and electromagnetic interference (EMI) which may lead to inter-devices interactions (two-way interactions) that can affect their individual power consumptions.

An experimental factorial design is adopted aiming to capture appliances main power factor effects (primary power consumptions) and their mutual two-way interactions (if possible). Statistical design of experiments refers to the process of planning the experiment so that the appropriate data that can be analyzed by statistical methods will be collected, resulting in some valid and objective conclusions. The statistical approach to experimental design is necessary if there is an objective to draw significant conclusions from the data [Montgomery, 2001]. In the context of NILM, operation of appliances can be considered as the main factor behind the transitions and fluctuations in the total aggregate measurements. Hence, it is reasonable to plan the experimental design considering the home appliances to be the main factors that affect the observed values of total aggregate measurements collected by the household smart meter. Besides, noisy signals and other possible disturbances can be considered as existing but uncontrollable factors as explained in section 3.2.

3.1 Interpretation of Appliances Interactions from Power Quality Perspective

In the ideal case of home devices operation, each appliance is expected to operate as an isolated entity without affecting or being affected by the operation of other home appliances. However, due to the increasing number of electronic devices and low power quality in the design or operation of some devices, the flow of different interference components such as harmonics and electromagnetic interference currents between home appliances is possible. The design or manufacturing of a device is considered to have low power quality issues when its operation does not satisfy the desired case in its operation. This includes emissions of harmonics, interference currents, being affected by the operation of other devices, etc. In general, electricity networks, components, and home appliances are becoming more sophisticated in terms of their functionalities and the way they interact with other equipment connected to the same network [Bhattacharyya *et al.*, 2007].

Gil-de-Castro *et al.* [Gil-de-Castro *et al.*, 2014a] studied harmonic interactions between different devices (e.g. LED lamp, microwave oven and TV) and how different combinations of these devices could affect the charging process of an electric vehicle (EV). It was found that the presence of a neighboring device could have a strong influence on the harmonics emitted by the EV. In addition, the EV was found to be affected by different emissions (amplitudes and frequencies) coming from different nearby devices. They concluded that to understand the current flowing through an electronic device, the interactions between connected devices must be understood.

In a different study, Gil-de-Castro *et al.* [Gil-de-Castro *et al.*, 2014b] investigated harmonic emission of different domestic equipment combined with different types of lighting. Their study showed that the total emissions produced by equipment depended on the lamps that were connected in the background.

Pavas *et al.* [Pavas *et al.*, 2012] carried out a statistical analysis of power quality disturbances propagation by means of the method of disturbances interactions. Their study was performed on a large scale to estimate possible interactions between two buildings. The measurements were done at the medium/low voltage level substation that feeds the two buildings. The presence of disturbances was found to be responding to possible interactions of the two buildings circuits and all connected devices.

3.2 Interpretation of Appliances Interactions from a Statistical Perspective

An experimental factorial design is adopted to represent the statistical model of home appliances and being able to estimate their possible two-way interactions. Figure 3.2 shows a typical experimental factorial model where an output is generally driven by a set of controllable factors and uncontrollable factors (e.g. noise signals, poor power quality factors) [Montgomery, 2001; Montgomery and Runger, 2011].

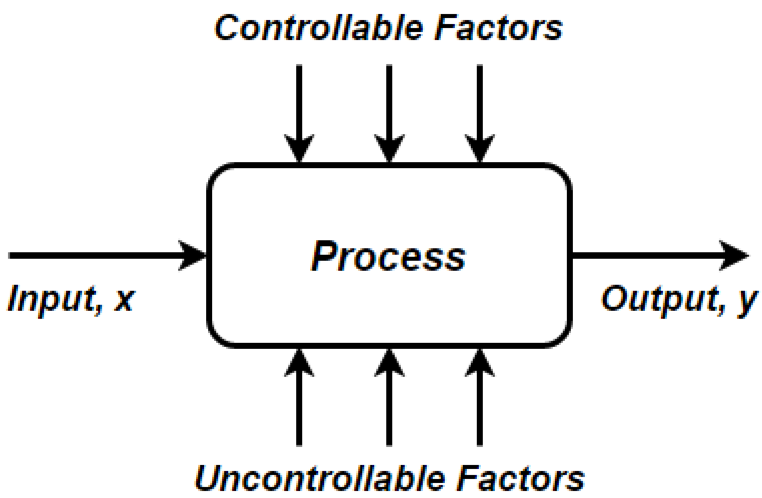

Fig. 3.2: Experimental factorial model.

In the context of the problem of load disaggregation, the input represents the main feeder line, home appliances are represented by the controllable factors which are controlled by the house occupants, and the output represents the total aggregate consumptions that is usually measured by a smart meter.

A model of 2^M factorial design was considered, where there exist M controllable factors (i.e. home appliances) and each factor has only two possible states (ON/OFF). Multi-state appliances or finite state machines (FSM) can also be modeled using this representation as they can be encoded, as several two-state appliances [Kim *et al.*, 2010] (by means of some defined constraints/rules to assure that a device occupies one state at a time). To estimate possible mutual interactions between appliances, the case of two devices (say D_A and D_B) is presented for simplicity. The used notation of the output response of the factorial design model (y) at different states of the two devices D_A and D_B is as shown in Table 3.1.

Table 3.1: Output response (y) at different devices D_A and D_B states.

Device D_A state	Device D_B state	Output response (y)
OFF	OFF	i
ON	OFF	a
OFF	ON	b
ON	ON	ab

In the context of NILM, the output response (y) represents the aggregate power consumption at different states combinations of existing home devices. It is often noted that even when all devices are OFF, there is a still a baseload power consumption due to always ON devices, which is denoted by the value of i in Table 3.1. A device is assumed to interact with other devices in the same household if its power consumption is not the same in cases of different combinations of the states of other devices. To apply on the case of two devices D_A and D_B, the devices main power effect (primary power consumption), denoted as A and B, and their two-way (or second order) interaction, denoted as AB, can be estimated from the following [Montgomery, 2001; Montgomery and Runger, 2011]

$$A = (ab + a - i - b)/2 \tag{3.1}$$

$$B = (ab + b - i - a)/2 \tag{3.2}$$

$$AB = (ab + i - a - b)/2 \tag{3.3}$$

For illustration of the above, if one devices D_A is ON, then the output response should be $y = a$, where the actual effect (main power effect) of device D_A is $A = a - i$ because it is still possible to have a baseload consumption before device D_A switches to ON.

Assuming the case of no devices interactions, when device D_B switches to ON beside device D_A, then the output response becomes $y = ab = a + B = a + b - i$, which will result in a value of $AB = 0$ when applied to equation (3.3).

Theoretically, an interaction AB may hold a positive or negative value and it interprets the difference of power consumptions when two devices are ON simultaneously and when considering their summation of individual power consumptions at the ON states. Thus, to be able to estimate an interaction between two devices, it requires to monitor them operating together (i.e. both are ON) and also when each of them is the only ON device. To benefit from equation (3.3) in the domain of load disaggregation, it can be re-written as

$$ab - a = b - i + 2AB \tag{3.4}$$

The term $ab - a$ in equation (3.4) represents a transition upward (denoted as $|tx|$) in the total aggregate signal when a device D_A is ON at time $t - 1$ then both devices D_A and D_B are ON at time t. It is noted that the term $b - i$ in equation (3.4) represents an estimate for the main power effect (primary power consumption) of the device D_B (i.e. $b - i = B$). Therefore, it can be concluded that

$$|tx| = ab - a \approx B + 2AB \tag{3.5}$$

The importance of equation (3.5) comes from the fact that when a device switches ON, not only its main power effect will be reflected on the total aggregate signal, but also its interactions effects (e.g. two-way interactions) with already ON devices. Conversely, when a device switches OFF, not only its main power effect will be suppressed from the total aggregate signal, but also its interactions effects with other ON devices. Hence, information on devices interaction are helpful in identifying home devices when a transition is detected in the total aggregate signal. Thereafter, information on devices interactions were embedded in the proposed FHMM representation of home devices and the total aggregate measurements to enhance the performance of load disaggregation.

3.3 The Sparsity of Effects Principle

In experimental factorial models, factors are normally expected to interact with each other at possible levels of interactions (two-way, three-way, etc.) which basically depend on the number of controllable factors. However, the principle of sparsity states that an experimental factorial model is usually dominated by main factors effects and low order interactions (e.g. two-way interactions). High order interactions often hold negligible effects on the output response [Montgomery and Runger, 2011]. Sometimes, this phenomenon is referred to as the hierarchical ordering principle which can be useful in the screening of significant factors from the insignificant ones [Montgomery, 2001].

Following the principle of sparsity of effects, only main factors effects of home devices and their two-way interactions, where possible, will be included in the proposed FHMM to tackle the problem of non-intrusive load disaggregation.

3.4 Models of Power Consumptions by Individual Appliances

It is common that each appliance has a rated power consumption that is supposed to be consumed during its operation. Finite state machines (FSMs) are also supposed to consume power at some rated levels for each of its possible states of operation. However, in the practical case, appliances may consume inconsistent amounts of power during their cycle of operation. Barker [Barker, 2014], who developed model-driven analytics of energy measurements, found that the profile of power consumptions by home appliances may start at overshot high values then decrease gradually, or, for other appliances, may start at some value then increase gradually during their operation.

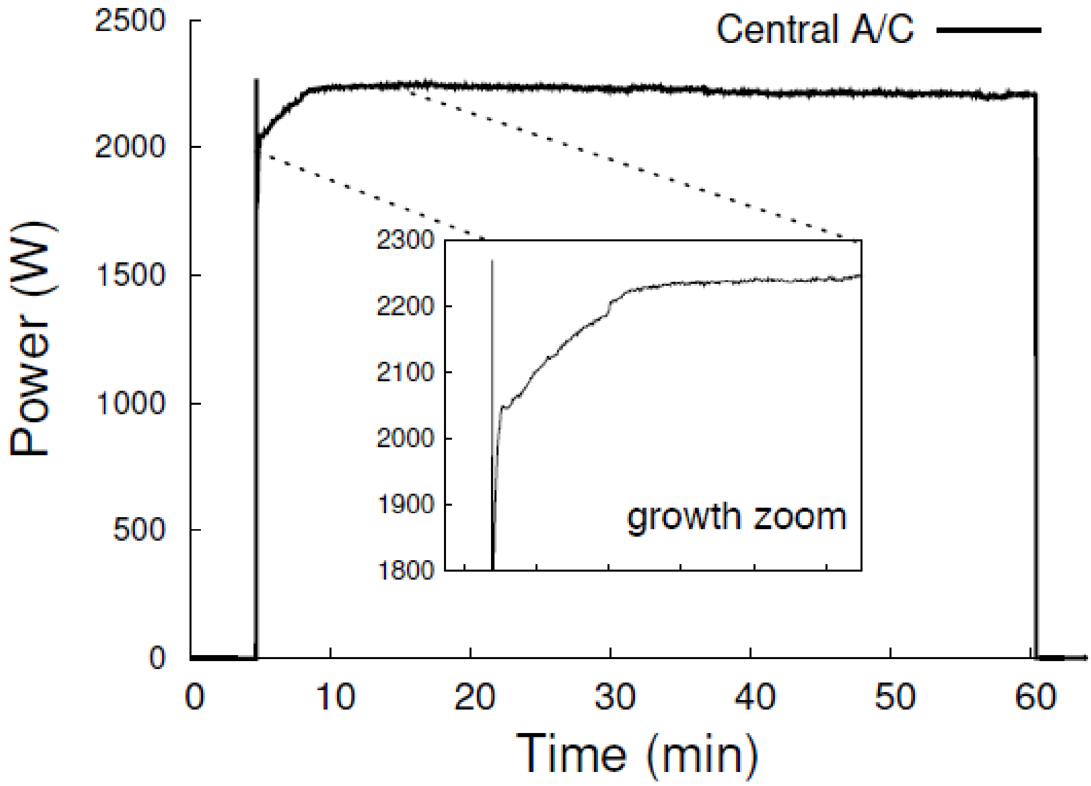

Fig. 3.3: Consumption pattern by a central air conditioner [Barker, 2014].

For illustration, Figure 3.3 shows a consumption pattern by a central air conditioner [Barker, 2014] in which the power consumption starts at some value then increase gradually during its operation. On contrary, Figure 3.4 shows a consumption pattern by a refrigerator from the REDD public data set [Kolter and Johnson, 2011], in which the power consumption starts at some value then decrease gradually during its operation.

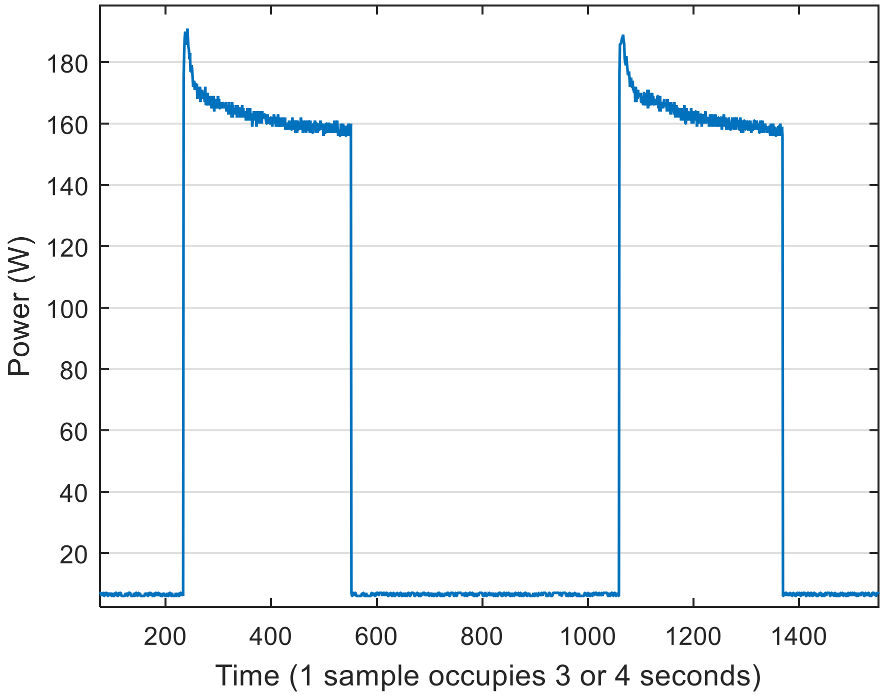

Fig. 3.4: Consumption pattern by a refrigerator [Kolter and Johnson, 2011].

From the above, it is noticeable that ON and OFF transitions may not be close to each other for some devices. Therefore, a model that consists of three power levels is considered to model the power consumption profile of each home appliance. The initial power consumption is used to represent the starting power consumption usually within few seconds duration after the device is switched ON (e.g. for five seconds duration). The final power consumption is used to represent the power consumption usually with few seconds before the devices is switched OFF (e.g. for five seconds duration). Overall, the average power consumption is used to represent the average (or mean) level of power consumed during the entire period of operation of a device.

To carry out the load disaggregation process, a transition upward in the total aggregate signal should be compared to initial power consumptions of known devices in order to identify the switched-on device. Conversely, a transition downward in the total aggregate signal should be compared to final power consumptions of known devices in order to identify the switched-off device. In all cases, the average power consumption of an appliance can be used as a reasonable estimate of its actual power consumption during its operation period.

3.5 Embedding Two-Way Interactions of Device in the FHMM

Information of mutual devices interactions (were possible to be estimated) are then embedded in the FHMM representation of home devices and the total aggregate signal. Figure 3.5 shows the proposed FHMM including all possible two-way interactions of home appliances.

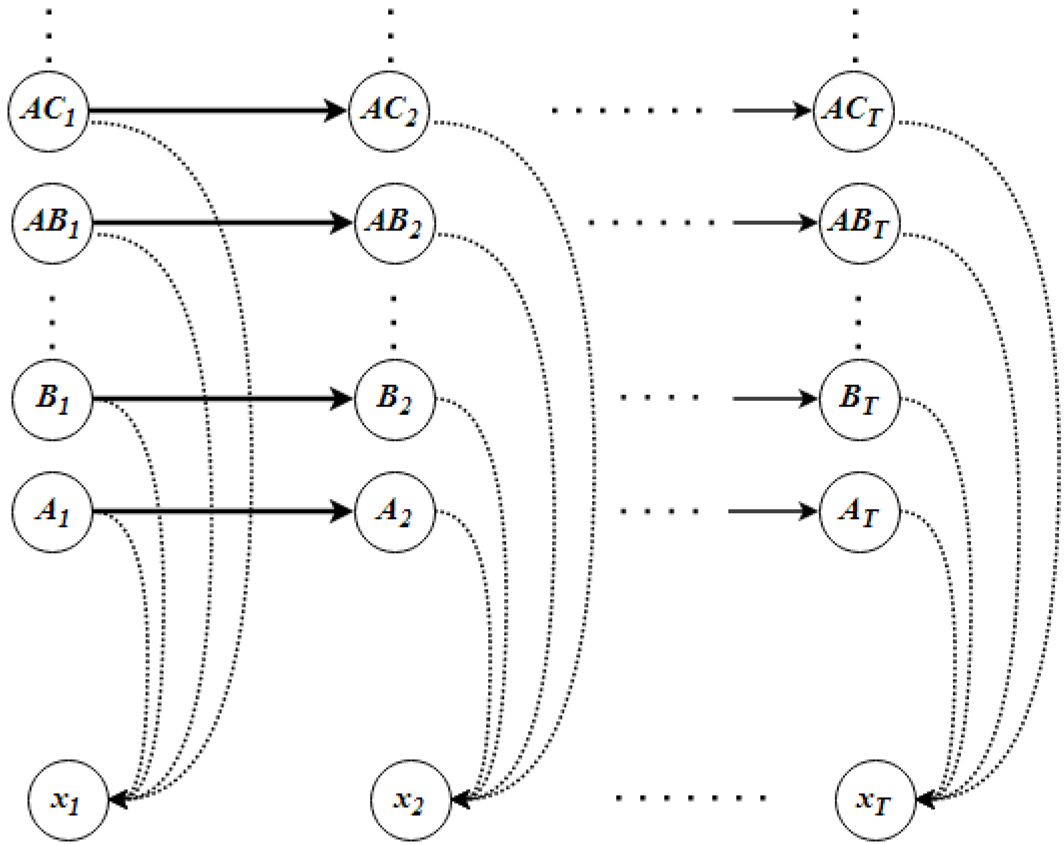

Fig. 3.5: FHMM with two-way interactions chains.

An interaction hidden chain is considered only when both of corresponding devices are in the ON state. For example, the two-way interaction AB is considered (i.e. to be affecting the X observations) only when both devices D_A and D_B are in the ON state. The observed sequence of measurements $X = \{x_1, x_2, \ldots x_T\}$ at a time instant t can be seen as the additive outcome of individual power consumptions of the ON devices and their mutual two-way interactions at that time instant.

To perform the task of load disaggregation, hidden states of devices could be inferred from the sequence of observed measurements using the Viterbi algorithm. However, a transition in the aggregate signal should consider the mutual interaction between the switched devices and all previously ON devices. To articulate this concept, consider a transition upward in the total aggregate signal, then it is reasonable to look for the best matching device D_X that satisfy the following

$$D_X \sim \underset{D_X}{\arg min} \left| \left(X_i + 2\sum_{Y} XY \right) - |tx| \right| \tag{3.6}$$

Likewise, when a transition downward in the total aggregate signa is detected, then it is reasonable to look for the best matching device D_X that satisfy the following

$$D_X \sim \underset{D_X}{\arg min} \left| \left(X_f + 2\sum_{Y} XY \right) - |tx| \right| \tag{3.7}$$

where X_i is the initial power level of the possible device D_X, X_f is the final power level of the possible device D_X, Y represents the set of currently ON devices just before the transition occurs, XY represents the two-way interaction pairs of devices D_X and D_Y and $|tx|$ is the absolute value of the transition caused by the switched device D_X (ON/OFF). Equations (3.6) and (3.7) are concluded from equation (3.5) and the fact that when a device is switched ON, not only its main power effect will be added to the total aggregate signal, but also its two-way interactions with currently ON devices. Conversely, when a device is switched OFF, not only its main power effect will be suppressed from the total aggregate signal, but also its two-way interactions effects with other ON devices.

It is notable that estimating both X_i and X_f should be over some period of time (periods of 5 samples were used in this research) as a transition is detected in order to avoid misleading abrupt spikes (often happen for one or two samples). Therefore, information on devices interaction are helpful in identifying home devices when a transition is detected in the total aggregate signal. It is noteworthy that identifying of appliances from equations (3.6) and (3.7) is not based only on the closest one that matches the detected transition in the aggregate signal, but it represents a measure that can be used via Viterbi algorithm to identify the most probable home device that has caused the transition in the total aggregate signal. Thereby, the proposed method provides an enhanced method for detection of appliances. In addition, estimation of power consumption by individual appliances could be enhanced by considering the possible changes of an appliances power consumption due to the operation of other existing home appliances.

3.6 Adaptive Estimation of Appliances Main Power Effects and Two-Way Interactions

It is possible to estimate appliances models and their mutual interactions (where possible) from a portion of the aggregate signal, then keeping these models fixed when performing the load disaggregation on subsequent portions of the total aggregate measurements. Hence, such fixed-model technique will not benefit from the extra information that is extractable during the disaggregation process. Therefore, applying an adaptive method that keep updating of original models, whenever the corresponding appliance is detected, could improve the appliances models' profiles and thereby the accuracy of detection and power estimation of individual home appliances.

In the context of load disaggregation, the smart meter measurements are usually of long sequences and they represent an infinite data stream in the real-life case. Therefore, it is reasonable to use sequential mean updating based on the current mean value and the new sample that belongs to the same population (i.e. the same device). That is, the new mean as a function of the current mean can be given by [Bishop, 2006]

$$\mu_{k+1} = \mu_k + \frac{1}{N}(x - \mu_k) \tag{3.8}$$

where μ_{k+1} is the new updated mean value (i.e. of an appliance power consumption) to be used in the subsequent process of load disaggregation, μ_k is the current mean value and N is the new total number of observed samples, including the new sample x used to model the average of a device power consumption. It is interesting to note that the shift (amount of deviation) from the current mean to the new mean value depends on how far the value of the new sample x is from the current mean value. In addition, since N is always increasing during the disaggregation process, the deviations from the current mean to the new mean caused by new samples gradually become smaller.

Though many appliances often exist in a standard household, only few of them are usually operating (ON) simultaneously [Kim *et al.*, 2010]. To update devices models and possible two-way interactions, an adaptive method is proposed according to the number of simultaneously ON devices. An adaptive approach was developed for cases when there are four devices or less operating (ON) simultaneously. Figure 3.6 shows the developed criteria to update estimates devices models and their two-way interactions (where possible) according to the number of simultaneously ON devices. The cases when there are five or more ON devices were neglected for two reasons: increasing complexity and the case is less frequent than other cases in ordinary situations of typical residential households.

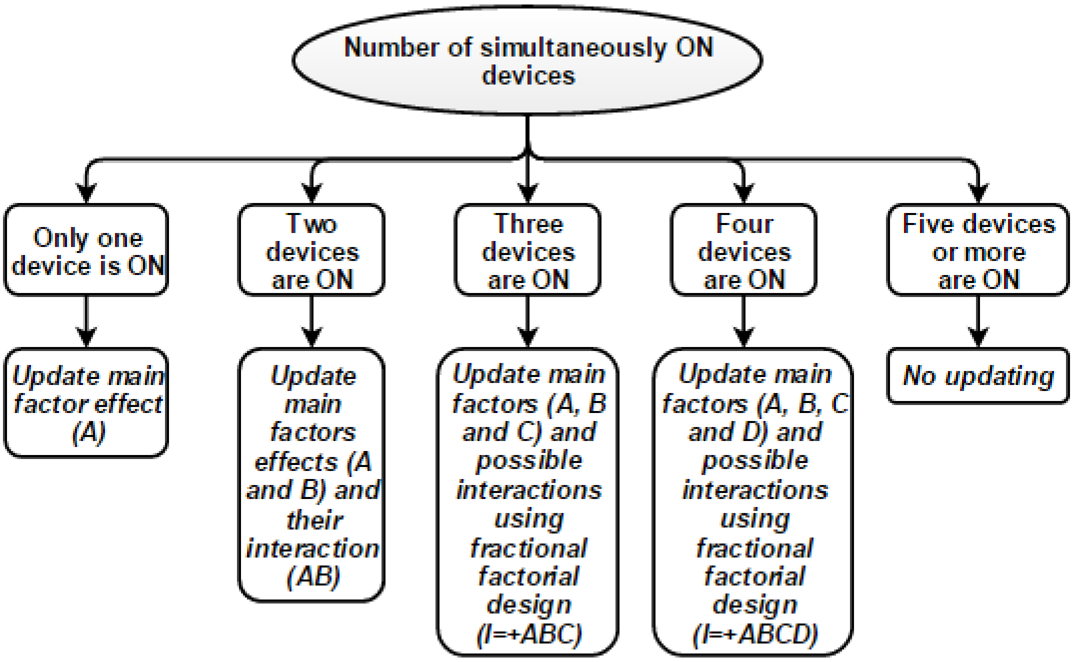

Fig. 3.6: Criteria of adaptive estimations based on the number of ON devices

When there is no active (ON) device in the household, the total aggregate measurements represent the household baseload (denoted as i). To articulate the adaptive estimation techniques, the following is considered for each corresponding case of simultaneously ON home devices.

In the first case, when there is only one ON device (denoted as device D_A), the total aggregate measurements can be considered as the observation of a. Hence, obtaining a new sample to update the model (i.e. average power consumption) of the device D_A is straightforward as follows

$$A' = a - i \tag{3.9}$$

where A' is a new sample to be used to update the estimate of the main power effect (denotated as A) of the corresponding device D_A. Thereafter, the updating process is directly handled by equation (3.8) to obtain a new estimate of the device model using the old estimate and the new obtained sample. It is notable that the number of samples

N in equation (3.8) should be increased by one whenever a new sample is available and used to update the estimate of the device main power effect.

In the second case, when there are only two ON devices (denoted as D_A and D_B), it is possible to update both of their main power effects (primary power consumptions, denoted as A and B) and their two-way interaction (denoted as AB). In this case, the total aggregate measurements can be considered as the ab observation. Thus, new samples of main power effects and two-way interactions can be obtained directly using equations (3.1), (3.2) and (3.3). Thereafter, the estimates of these factors can be updated using equation (3.8) exactly in the same logic as articulated previously in the case of only one ON device.

In the third case, when there are only three ON devices (denoted as D_A, D_B and D_C), it is possible to update their main power effects (primary power consumptions, denoted as A, B and C). In some cases, it could be possible to update some or all of their two-way interactions (denoted as AB, AC and BC). In this case, the total aggregate measurements can be considered as the abc observation. To obtain new samples for either main power effects or two-way interactions, the half fractional factorial design (denoted as 2^{3-1}) is proposed to be used with the defining relation of $I = +ABC$. For illustrations, Table 3.2 shows a typical 2^3 full factorial design with upper half is the fractional factorial design (half fraction) with the defining relation of $I = +ABC$ [Montgomery, 2001; Montgomery and Runger, 2011]. In brief, using a half-fractional factorial design is proposed when it could be difficult to obtain all combinations of observations. Using a half-fractional factorial design (half of observations only) would incur additional complexity cost, which is the aliasing between a number of factors. When two factors are aliases, it means that it would be impossible to separate their individual estimates, but the sum of their effects, however, could be estimated.

Table 3.2: Plus (ON) and Minus (OFF) signs for the 2^3 Factorial Design.

Combination	**Factorial Effect**						
	$\boldsymbol{A}$	$\boldsymbol{B}$	$\boldsymbol{C}$	$\boldsymbol{AB}$	$\boldsymbol{AC}$	$\boldsymbol{BC}$	$\boldsymbol{ABC}$
$\boldsymbol{a}$	+	-	-	-	-	+	+
$\boldsymbol{b}$	-	+	-	-	+	-	+
$\boldsymbol{c}$	-	-	+	+	-	-	+
$\boldsymbol{abc}$	+	+	+	+	+	+	+
$\boldsymbol{ab}$	+	+	-	+	-	-	-
$\boldsymbol{ac}$	+	-	+	-	+	-	-
$\boldsymbol{bc}$	-	+	+	-	-	+	-
$\boldsymbol{i}$	-	-	-	+	+	+	-

Since lower half of the factorial design in Table 3.2 consists of combinations that are more unlikely to be observed than combinations in the upper half, the upper half, represented by the defining relation $I = +ABC$ (upper half has ABC column always with positive signs) can be used to obtain samples to update main power effects and two-way interaction (if possible). However, it becomes difficult to compute samples of individual factors from the half fractional design due to the confounded factors (sometimes called as aliasing between different factors). Nonetheless, an estimate of the summation of the aliased factors (confounded together) can be obtained from the following

$$L_A = \frac{1}{2}(abc + a - b - c) = A + BC \tag{3.10}$$

$$L_B = \frac{1}{2}(abc + b - a - c) = B + AC \tag{3.11}$$

$$L_C = \frac{1}{2}(abc + c - b - a) = C + AB \tag{3.12}$$

For example, equation (3.10) shows the confounding case between the main factor A and the interaction BC, which means that it is not possible to observe (i.e. have an estimate) of each of them individually, but it is possible to obtain a measurement of their summation (which is denoted as L_A). To illustrate the updating process, the case of L_A as appears in equation (3.10) is considered, but the same logic applies for cases of L_B in equation (3.11) and L_C in equation (3.12). For the case of L_A, there are two possible scenarios based on the availability of the interaction AB in prior. The first scenario is when the interaction BC is unknow in prior (i.e. $BC = 0$). Hence, the value of L_A can be used as a new sample to update the main power effect of the factor A exactly using the same method as applied in the case of only one ON device. The second scenario happens when there is a previous estimate of the interaction BC. Hence, the previous estimates of both A and BC can be updated using their known prior *ratios* as follows

$$A' = \tilde{A}\left[\frac{L_A}{\tilde{A} + \widetilde{BC}}\right] \tag{3.13}$$

$$BC' = \widetilde{BC}\left[\frac{L_A}{\tilde{A} + \widetilde{BC}}\right] \tag{3.14}$$

where L_A represents the new measurement of $A + BC$ as a confounded value, $\tilde{A}$ and $\widetilde{BC}$ are the previous averaged values for the main factor effect A and the two-way interaction BC, respectively. Thereafter, new samples (A' and BC') can be used to update the corresponding main power factor and the two-way interaction using equation (3.8).

In the fourth case, when there are only four ON devices (denoted as D_A, D_B , $D_C$and D_D), it is possible to update their main power effects (primary power consumptions, denoted as A, B, C and D). In some cases, it could be possible to update some or all of their two-way interactions (denoted as AB, AC, AD, BC, BD and CD). In this case, the total aggregate measurements can be considered as the $abcd$ observation. To obtain new samples for either main power effects or two-way interactions, the half fractional factorial design (denoted as 2^{4-1}) is proposed to be used with the defining relation of $I = +ABCD$. For illustration, Table 3.3 shows a portion of the 2^{4-1} fractional factorial design with the defining relation of $I = +ABC$ [Montgomery, 2001; Montgomery and Runger, 2011]. Since three-way interactions are negligible with accordance of the principle of sparsity, they were not included in the fractional factorial design shown in Table 3.3.

Table 3.3: A portion of the 2^{4-1} fractional factorial design.

Response	Factorial Effect										
	A	B	AB	C	AC	BC	D	AD	BD	CD	$ABCD$
i	−	−	+	−	+	+	−	+	+	+	+
ab	+	+	+	−	−	−	−	−	−	+	+
ac	+	−	−	+	+	−	−	−	+	−	+
bc	−	+	−	+	−	+	−	+	−	−	+
ad	+	−	−	−	−	+	+	+	−	−	+
bd	−	+	−	−	+	−	+	−	+	−	+
cd	−	−	+	+	−	−	+	−	−	+	+
$abcd$	+	+	+	+	+	+	+	+	+	+	+

To update the main power effects, a new sample for corresponding factors can be obtained from Table 3.3 as follows

$$A' = \frac{1}{4}(abcd + ab + ac + ad - bc - bd - cd - i) \tag{3.15}$$

$$B' = \frac{1}{4}(abcd + ab + bc + bd - ac - ad - cd - i) \tag{3.16}$$

$$C' = \frac{1}{4}(abcd + ac + bc + cd - ab - ad - bd - i) \tag{3.17}$$

$$D' = \frac{1}{4}(abcd + ad + bd + cd - ab - ac - bc - i) \tag{3.18}$$

Thereafter, new samples of main power effects obtained from equations (3.15) to (3.18) can be used to update the corresponding device main power effect using equation (3.8) exactly in the same logic as done in the previous cases. It is notable that though main factors are confounding with three-way interactions, three-way interactions are neglected in accordance with the principle of sparsity. It is notable also that some two-way responses (observations of ab, ac, ad, bc, bd and cd) may not be available in prior to updating of main power effects. In this case, an estimate of the two-way response can be obtained as in the following example for the response of ab

$$ab \approx A + B + i \tag{3.19}$$

In case a two-way response is not available in prior (e.g. ab is unknown), the corresponding two-way interaction will usually be unavailable (e.g. $AB = 0$).

To update the estimates of the two-way interactions in this configuration, it is important to handle the aliasing between each pair of two-way interactions and its alias (the pair that it confounds with). From Table 3.3, the following linear combinations can be concluded to show the confounding between the available two-way interactions

$$L_{AB} = \frac{1}{4}(abcd + i + ab + cd - ac - bc - ad - bd) = AB + CD \quad (3.20)$$

$$L_{AC} = \frac{1}{4}(abcd + i + +ac + bd - ab - bc - ad - cd) = AC + BD \quad (3.21)$$

$$L_{AD} = \frac{1}{4}(abcd + i + ad + bc - ab - ac - bd - cd) = AD + BC \quad (3.22)$$

For illustration, equation (3.20) articulates that it is impossible to obtain an estimate of either interaction AB or CD individually. Instead, the estimate L_{AB} represents their summation (i.e. confounded together as $AB + CD$). Hence, it is only possible to update the estimates of interactions (e.g. AB and CD) only if a previous estimate is known from another measurements or estimation. Otherwise, it would be impossible to update the estimates of the two-way interaction. In case both confounding two-way interactions (e.g. AB and CD) are known from some previous estimation, they can be updated by obtaining new samples from the following

$$AB' = \widetilde{AB}\left[\frac{L_{AB}}{\widetilde{AB} + \widetilde{CD}}\right] \quad (3.23)$$

$$CD' = \widetilde{CD}\left[\frac{L_{AB}}{\widetilde{AB} + \widetilde{CD}}\right] \quad (3.24)$$

where L_{AB} observation represents the new measurement of $AB + CD$, $\widetilde{AB}$ and $\widetilde{CD}$ are the current known, or previously estimated, averaged values for the interactions AB and CD, respectively. Thereafter, the new samples AB' and CD' can be used to update the corresponding two-way interaction in the same logic as explained earlier.

In the fifth case, when there are five or more ON devices simultaneously, no updating's were performed due to two main reasons

1. Having five or more ON devices simultaneously becomes more unlikely to happen in households as shown in [Kim *et al.,* 2010] and as found from the results of this research work.
2. The increasing complexity of the factorial design in this configuration.

3.7 Procedure to Perform the Load Disaggregation

To carry out the load disaggregation task, the following summarized procedure was applied merely on the total aggregate measurements

- Step 1: using a portion of the total aggregate signal (e.g. two days of low frequency measurements), models of devices were built, and their two-way interactions were estimated (if possible).
- Step 2: the process of load disaggregation was performed on the remaining part of the total aggregate signal.
- Step 3: at every sample processed for disaggregation in step 2, the main power factors effects and two-way interactions of ON devices were adaptively estimated and updated using the criteria explained in section 3.6.
- Step 4: accuracy of disaggregation can be evaluated by comparing the resulting power profiles estimates to the actual individual consumptions of devices as presented in subsection 2.2.4.

3.8 Case Study Application on Using Appliances Interactions to Enhance Load Disaggregation

The approach proposed to utilize appliances two-way interactions (where possible) was applied on real home measurements from the REDD public data set as explained earlier in this Chapter. House 2 from the REDD was selected to test the proposed approach since it includes nine appliances only, which made it easier to extract information on mutual interactions between appliances. The nine devices are the microwave, refrigerator, outlet 1, outlet 2, washer dryer, lights, stove, dishwasher and disposal. These devices are fed by two feeder lines, which are labeled as mains1 and mains 2 circuits. As a pre-processing of smart meter measurements, a 9-ample median filter was applied in order to remove noisy signals and outliers which may mislead the learning, modeling or disaggregation processes. The median filter sorts the processed samples in an ascending or descending order and then return the mid-point value as a result; thereby removing noisy spikes or values that are inconsistent with its neighboring samples.

3.8.1 Extracting Interaction Information and Performing Load Disaggregation

The learning of appliances models and extracting possible two-way interactions of appliances were done using a portion of two days of measurements by applying methods explained in this Chapter. Firstly, the base load of each feeder line should be estimated from the minimum level of total aggregate power consumptions. The baseload estimations represent the response (i) in the proposed factorial design and were found as shown in Table 3.4.

Table 3.4: Baseload for mains 1 and mains 2 feeder lines.

Circuit	Mains 1	Mains 2
Base load (W)	15.7	22.8

Although positive baseload values mean that there are some always ON appliances, it can be considered as the response i in the factorial design since, at baseload consumptions, all other home appliances (to be detected later when they are switched ON) are still in the OFF state.

In the learning phase using two days of aggregate data, the models of appliances were built using the three power levels as described in section 3.4. Table 3.5 shows that detected appliances (or detected states in appliances) with corresponding initial, average and final power consumptions of these appliances (or states of appliances). As examples from Table 3.5: D_B refers to the refrigerator, D_C refers to a microwave, both D_I and D_J are states a washer dryer, etc. The matching process of these states/appliances symbols to the actual home appliance was done *manually*, since an unsupervised adopted approach results in unlabeled group of appliances.

Table 3.5: Detected appliances/states with three levels of power consumptions.

State/Appliance Symbol	Initial Power (W)	Average Power (W)	Final Power (W)
$\boldsymbol{D_A}$	22.8	22.8	22.8
$\boldsymbol{D_B}$	277.5	237.6	255.0
$\boldsymbol{D_C}$	1973.2	1934.7	1945.7
$\boldsymbol{D_D}$	230.3	199.0	223.5
$\boldsymbol{D_E}$	66.0	42.0	64.7
$\boldsymbol{D_F}$	102.8	78.3	95.2
$\boldsymbol{D_G}$	57.7	41.1	56.0
$\boldsymbol{D_H}$	1111.0	1097.4	1107.9
$\boldsymbol{D_I}$	271.0	255.1	268.3
$\boldsymbol{D_J}$	1283.8	1230.3	1234.2
$\boldsymbol{D_K}$	807.4	790.2	807.2
$\boldsymbol{D_L}$	394.4	409.5	410.1

Thereafter, appliances modeling and estimation of mutual interactions were performed on a two-day length sequence of measurements. Table 3.6 shows the obtained estimations of appliances two-way interactions. The unknown appliance in Table 3.6 was represents a consumption transition of an unlabeled appliance in the REDD.

Table 3.6: Estimated interactions between appliances.

Appliance D_A	Appliance D_B	Interaction AB (W)
Outlet 1	Outlet 2	-27.7
Outlet 2	Unknown (≈40 W)	-12.5
Refrigerator	Microwave	-31.5

The adaptive estimations approach proposed in this Chapter is applicable to cases of four or less simultaneously ON home appliances. Thus, number of concurrently ON home appliances was investigated during the load disaggregation process and found as shown in Table 3.7.

Table 3.7: Time occupancy for number of concurrently ON appliances.

Number of concurrently ON appliances	Occupied time (hours/day)	Percentage of time occupancy of a day
Base load only	14.45	60.22%
One appliance	5.81	24.21%
Two appliances	2.69	11.22%
Three appliances	0.81	3.37%
Four appliances	0.21	0.87%
Five appliances or more	0.03	0.11%
Total	**24 hours**	**100%**

The proposed non-intrusive load disaggregation approach then was applied on one-day sequences of the total aggregate smart meter measurements. The proposed approach was applied three times, each time by applying a different scenario as follows

- The first scenario: using the modeled appliances *without* information on appliances two-way interactions, the standard FHMM was applied to disaggregate individual home loads consumptions.
- The second scenario: using the modeled appliances *with* information on appliances two-way interactions, a FHMM that embeds information on appliances mutual interactions was applied to disaggregate individual home loads consumptions.
- The third scenario: using the modeled appliances *with* information on appliances two-way interactions, a FHMM that embeds information on appliances mutual interactions was applied to disaggregate individual home loads consumptions. In addition, the adaptive estimations approach was applied at every sample of the aggregate signal to update appliances main power effects and two-way interactions (if possible) as explained earlier in this Chapter.

Table 3.8 shows a comparison between the obtained results when approaches are applied on house 2 from the REDD using the three above presented scenarios. Table 3.8 shows results in terms of disaggregation accuracy using the metric of correct assigned power compared to the actual consumed power by an appliance as explained earlier in subsection 2.2.4. Figure 3.8 shows an example of the washer dryer that illustrate the estimated versus the actual power consumptions of the appliances.

Table 3.8: Disaggregation accuracy for each device.

Appliance	FHMM (without interactions)	FHMM (with interactions)	FHMM (with interactions and adaptive estimations)
Microwave	54.26%	57.29%	58.61%
Outlet 2	78.66%	81.40%	85.83%
Refrigerator	64.11%	64.57%	66.41%
Outlet 1	68.66%	68.80%	65.89%
Washer dryer	77.26%	77.26%	87.37%
Lights	51.62%	51.66%	52.43%
Stove	81.56%	81.56%	80.66%

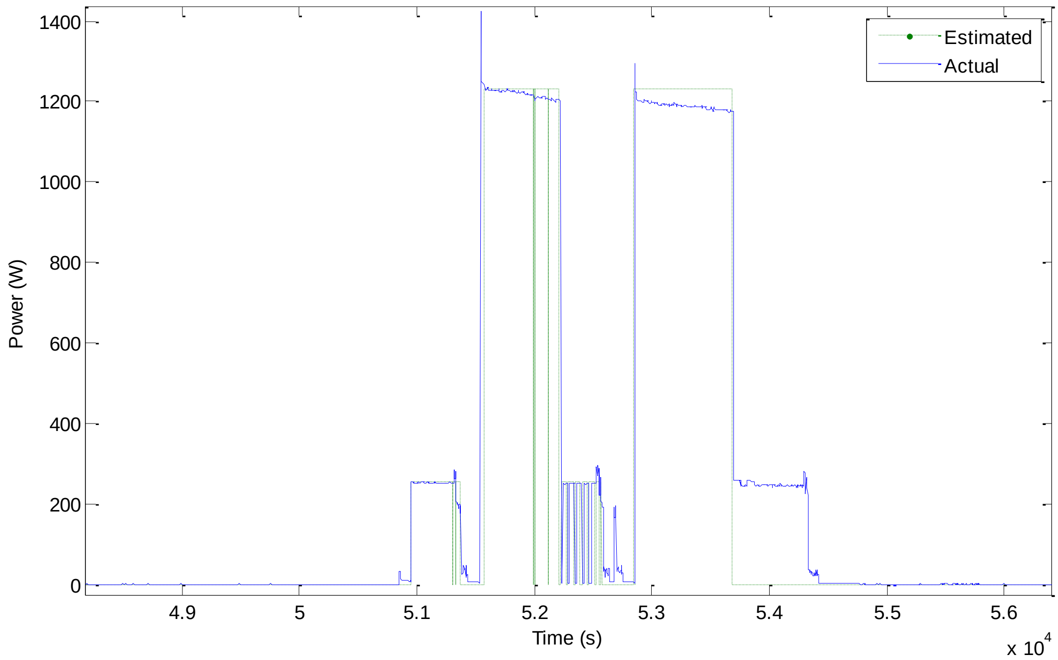

Fig. 3.7: Estimated versus actual power consumption of the washer dryer.

3.8.2 Interpretation and Discussion of Load Disaggregation Results

Including information of appliances two-way interactions (when possible) by embedding to the FHMM in the proposed approach improved the detection of appliances and load disaggregation accuracies. The rationale in this setting is to consider the behavior of an appliance when there are other appliances operating in the background. The proposed FHMM with interactions showed an improvement in the disaggregation accuracy whenever a mutual interaction between two devices was observed and estimated. In addition, the adaptive estimations used to update appliances models and possible interaction improved the load disaggregation accuracy for almost all devices except for the stove and the outlet 1.

In the learning and modeling phase, it was possible to estimate a limited number of two-way interactions between existing appliances. This reveals the fact that to estimate an interaction between two appliances, it is essential to be able to observe each appliance operating alone in different time instants, beside observing when both are operating simultaneously. These conditions, as explained earlier in this Chapter, limited the cases of possible estimations of interactions between pairs of appliances. Therefore, it was possible to estimate three cases of two-way interactions between home appliances as shown in Table 3.6. The transition of the unknown appliance in Table 3.6 was detected but there was no close match with any of the known appliances in the household. Hence, it was given a generic unknown appliance name.

The load disaggregation results conclude that embedding estimated interactions characteristics have improved the disaggregation accuracy for a set of appliances such as the microwave, refrigerator, outlet 1 and outlet 2. No significant improvement was achieved for the rest of other devices since no interaction information were extracted (to be utilized in the FHMM representation) from the measurements sequence for these appliances. Conversely, it was found that either interactions information was embedded to FHMM or not, appliances with unknown two-way interactions features were not significantly affected in terms of the disaggregation accuracy. Overall, embedding information on appliances interactions helped in improving the load disaggregation accuracies from two aspects

- Improving detection of appliances, as information of appliances interactions were utilized to investigated most probable appliances switching by considering interactions with other ON appliances. That is, when a transition in the aggregate signal is detected, not only the primary power consumptions should be searched for matching, but also how the switching appliance may interact with already ON appliances in the same household.
- Improving the estimation of power consumptions by appliances, as information of appliances interactions provide how their power consumption is affected when there are ON appliances in the background. Thus, this change of power consumption should be reflected in the process of power estimation of individual home appliances.

The adaptive estimations of main power factors and two-way interactions of operating devices, where applicable, has improved the power estimation accuracy for almost all present appliances except for the stove and outlet 1. Table 3.7 depicts that considering the adaptive updating process in cases of four or less concurrently ON appliances is reasonable. As shown in Table 3.7, observing five appliances or more to operate at the same time was found to be rare in this case study on a typical household (which is house 2 from the REDD public data set).

The adaptive estimations techniques were applied during the disaggregation process which provided a method to follow up and mimic the changes of appliances power consumption (main power effects) and possible inter-appliances two-way interactions.

This adaptive updating approach re-estimated corresponding appliances models (power consumptions profiles) by utilizing available information from the aggregate total signal during the load disaggregation process. From Table 3.8, it was found that updating the primary power consumption (main factor effect), even when no interaction information is available, may provide a good enhancement in the load disaggregation accuracy as in the case of the washer dryer. The lights in the selected house were modeled as a lumped model but they were actually several distributed light bulbs as can be seen from the individual consumptions of the REDD public data set. This resulted in a poor modeling of lights which explains the low disaggregation accuracy results obtained even in the considered standard configuration of the FHMM.

The two cases of the stove and outlet 1 showed lowered disaggregation accuracy that is not better than those obtained in cases of no adaptive estimations methods were applied. The lowered disaggregation accuracy happened since there were still errors in appliances detections or power consumptions models, which have likely accumulated and contributed to the obtained disaggregation results. In addition,

- From the individual actual consumption data of outlet 1, it was observed that it sometimes consumed different amounts of power which indicates the possibility to have more than one appliance (or FSMs) connected to the outlet 1. Moreover, there has been a low power consumption level (about 15 W) for some periods, which is undetectable by our approach (due to being a too small transition) but could affect the adaptive estimation process. The adaptive estimations techniques are affected by small transitions in the total signal because the total consumptions measurements always play a role in the adaptive estimations process as clarified in section 3.7.
- The stove was observed in operations only in day 2 and day 7 of the disaggregation period. The drop in the disaggregation accuracy for the stove is because there have

been smooth gradual transitions (instead of clear or sharp transition) that happened in day 2 of the disaggregation period. These smooth gradual transitions are not detectable by the proposed modeling approach as they basically depend on significant ON/OFF switching events.

For the purpose of comparison of approaches, the following results were obtained when applying FHMMs on house 2 of the REDD and using the same power estimation accuracy metric

- Kolter and Johnson [Kolter and Johnson, 2011] obtained energy disaggregation accuracies in the range of 50.8% to 59.6% when applying a FHMM on house 2 from the REDD.
- Johnson and Willsky [Johnson and Willsky, 2013] obtained energy disaggregation accuracies in the range of 50.8% to 84.8% when applying a FHMM on house 2 from the REDD.

These results show good performance of the proposed methods in Chapter 3, when compared to methods in literature that were applied on house 2 from the REDD and adopted the same accuracy metric (accuracy of power estimations).

3.8.3 Unseen Appliances and Computational Complexity

There are two appliances in house 2 from the REDD, the dishwasher and the disposal, not appearing in Table 3.8 showing the load disaggregation results. From the individual power consumptions data of the REDD, it was observed that the dishwasher operated only once during the entire two-week period at time when data was collected. This single appearance could be sufficient to build a model for the appliance but the appliance itself was not switched ON again during the rest of the two-week period. Hence, it was not detected by our proposed disaggregation algorithm. On the other hand, the disposal consumed varying power amounts during the learning phase. Therefore, appliance with varying power was not detectable by the proposed load disaggregation approach.

For further assessments of the proposed approach, the computational complexity of the proposed adaptive estimations approach was investigated. The approaches were executed on one-day sequences of samples using Matlab 2013a installed on a PC with Intel Core i5 3230M dual CPU 2.60 GHz and 4.00 GB of RAM. The average simulation and execution time increased from 13.1 s (for FHMM without adaptive estimations) to 15.4 s (for FHMM with adaptive estimations). With increased capabilities of processors and computers in recent years, this increase in complexity is acceptable for the improved disaggregation accuracies gained from the adaptive approach. The executional time for the algorithm serves as an indication in the practical deployment of NILM where the availability of disaggregation outcomes to household occupants in *real time* are important to lead to potential impacts in energy saving.

3.9 Remarks on the Proposed Approaches in this Chapter

The proposed approaches utilized information on two-way interactions between home appliances aiming to enhance the overall load disaggregation accuracy. Some remarks were noticed in the proposed approaches, which can be summarized as follows

1. To be able to extract information on a two-way interaction between two appliances, it is necessary to observe the total aggregate measurements in cases when both are operating (ON) together and when each of them is the only operating (ON) appliances. These conditions may limit the number of extractable information of appliances interactions. Longer sequences of measurements (e.g. extra few days of aggregate data) could be used in the learning phase (if available) in order to capture as much information on appliances interactions as possible.

2. The proposed approach makes use of only the main power effects and two-way interaction of home appliances. Three-way and higher order interactions were neglected in accordance with the principle of sparsity. However, in some factorial design experiments, three-order or higher interactions could show some significance. In cases three-order interactions are significant, they could contribute to the overall enhancement of the disaggregation accuracy. For examples on using three-way interactions in various domains in science and engineering, refer to [Montgomery, 2001; Montgomery and Runger, 2011]. However, incorporating three-way interactions between appliances would incur added computational complexity in both learning and testing phases.
3. The adaptive estimations approach was applied on cases when there are simultaneously four or less ON appliances in the household. Cases of having five or more simultaneously ON appliances was found unlikely to happen. However, number of simultaneously ON appliances depends mainly of the number of appliances, number of occupants and their usage behavior. Thus, applying an adaptive estimation method for cases when there are five or more ON appliances could be significant for some households. However, expanding the adaptive estimations approach would incur extra computational complexity.
4. The adaptive estimations approach updates appliances models that are detected by the load disaggregation approach. This means that if some appliances were incorrectly detected, it would lead to further inappropriate updating to these appliances models (i.e. their main power effects and two-way interactions).

3.10 Summary

This chapter presented the first proposed approach that aims to enhance the overall load disaggregation accuracy by embedding information of appliances mutual interactions.

In addition, the proposed approach was extended by adaptive estimations techniques that update the models of home devices and their interactions (when possible). The proposed approaches were tested on a real house from an available public data set. Results showed enhanced performance of the load disaggregation methods when applying the proposed approaches, compared to the reference FHMM techniques for modeling and estimation of individual appliances power consumptions.

Chapter 4

Disaggregating Continuously Varying Loads and a Framework for the Hybrid Continuous/Discrete States FHMM

This chapter presents the second proposed approach that aims to enhance the overall performance of non-intrusive load disaggregation. Modeling and disaggregating continuously varying home loads is a principal challenge in the field of load disaggregation. Continuously varying home loads are those consuming power at infinite possible number of states such as light dimmer, electronic devices, power tools, etc. In this Chapter, a method is proposed to model continuously varying home loads using a quantized continuous-state hidden Markov model (CS-HMM) that estimates the transition matrix in a way to mitigate two possible extreme cases of transitions. Finally, the proposed CS-HMM was consolidated with the factorial hidden Markov model (FHMM) to produce a hybrid continuous/discrete state FHMM that capable of modeling and disaggregating various types of home loads.

4.1 The Continuous-State Hidden Markov Model

The continuous-state hidden Markov model (CS-HMM), or sometimes called the infinite HMM, is that with the belief of having a continuous space of hidden states (i.e. possible infinite number of hidden states) which influence the sequence of observable emissions or measurements [Beal *et al.*, 2002]. It is notable that increasing complexity of the model or unavailability of the transitions likelihoods of hidden states in the CS-HMM are essential challenges that arise when applying a CS-HMM on real-world problems. To articulate the basic structure of the CS-HMM, Table 4.1 shows a comparison between the used notation for both the discrete-state HMM (DS-HMM) and the continuous-state HMM (CS-HMM).

Table 4.1: Notation for the CS-HMM and the DS-HMM.

Component	DS-HMM	CS-HMM
State notation	i, j	z_{t-1}, z_t
Observation	x_t	x_t
Transition probability	$a_{i,j}$	$P(z_t \mid z_{t-1}, f_{Z,Z})$
Observation probability	$b_j(x_t)$	$P(x_t \mid z_t, f_X)$
Initial state probability	π_i	$P(z_0)$

where $f_{Z,Z}$ is a joint continuous function that represents likelihoods of transitions in the continuous space from state z_{t-1} at time $t-1$ to state z_t at time t. When the continuous transitions probability likelihood ($f_{Z,Z}$) is unavailable in prior, some approximating methodologies, such as nonparametric statistics or quantizing the continuous states range, could be applied. The function f_X in Table 4.1 defines the probability of having an observation x_t over possible hidden state z and $P(z_0)$ defines the likelihood of a state z at time $t = 0$.

A typical CS-HMM is then often characterized by the joint density functions of both observations and states sequences which can be given as

$$P(X_T, Z_T) = P(z_0) \prod_{t=1}^{T} P(x_t \mid z_t) P(z_t \mid z_{t-1}) \tag{4.1}$$

where T represents the length of the modeled sequences of observations (measurements sequence, X_T) and the hidden states sequence (Z_T). The prediction of a possible observation x_t at time t can be obtained by marginalization of equation (4.1) as follows [Ainsleigh, 2001]

$$P(x_t) = \int P(x_t, z_{0:t})\, dz_{0:t} \tag{4.2}$$

where $z_{0:t}$ represents all possible sequences of hidden states from z_0 to z_t. The estimation of the maximum *a posteriori* (MAP) of hidden state sequence $\check{Z}_T$, given a sequence of observed measurements is another interesting goal in the CS-HMM, which is expressed as follows

$$\check{Z}_T = \underset{Z_T}{\arg max}\, P(Z_T|X_T) \tag{4.3}$$

Equation (4.3) can be solved efficiently by the Viterbi algorithm starting with the initialization of the forward optimal path with $\emptyset(z_0) = P(z_0)$, and the forward recursion cab be defined for $t = 1, 2, \dots T$ as follows [Ainsleigh, 2001]

$$\emptyset(z_t) = P(x_t|z_t) \underset{Z_{t-1}}{max}\, \{P(z_t|z_{t-1})\emptyset(z_{t-1}) \tag{4.4}$$

4.2 The Quantized Continuous-State Hidden Markov Model

Quantization of a continuous space or signal in its broad concept have been applied in various fields, see e.g. [Widrow *et al.*, 1996; Lloyd, 1982; Lookabaugh and Gray, 1989]. The discretization of a continuous space often produces a finite number of levels, that range from a minimum point L_{min} to maximum point L_{max}, with each quantization level has a width Δ that can be given by

$$\Delta = \frac{L_{max} - L_{min}}{N} \tag{4.5}$$

where N is the total number of levels. To distinguish a quantized space from spaces that are finite by nature, quantized levels are denoted here as either k or l, with respective $k \rightarrow l$ transition probability of $a_{k,l} = P(s_t = l|s_{t-1} = k)$, where the state model switches from state k at time $t-1$ to state l at time t. If the center point of a level k, denoted as c_k, is used to estimate or substitute the actual value of the signal at any time t, denoted as x_t, then there is a potential relative estimation error of

$$Error = \frac{|c_k - x_t|}{x_t} \tag{4.6}$$

Since this approach approximates the CS-HMM to a version of a DS-HMM, the estimation of the transition matrix and prior probabilities of the quantized CS-HMM are still required to model real-world applications. Two common classical methods to estimate prior state probabilities and the transition matrix are as follows

1. Using uniform distributions over possible transitions and state priors, in cases no information is available [Parson, 2014]. In this method, for each element in the $N \times N$ transition matrix, the assigned probabilities are $a_{k,l} = 1/N$. This is also applicable to the initial or prior states probabilities π_k. The essential shortcoming of this method is that some probabilities could be overestimated, and some could be underestimated. Hence, errors in the inference process or the prediction process using this HMM configuration are likely to occur. In general, using a uniform distribution over hidden states transitions degrades the importance of modeling the transitions between states, which is an essence when modeling a case application using HMMs.
2. Using the empirical method that estimate transitions probabilities according to the time occupied in these states [Kolter and Jaakkola, 2010]. This method estimates transitions probabilities based on counting the number of detected occurrences for

each transition. The basic shortcoming of this method is that while some transitions may have been never visited during the learning phase, other transitions may dominate, especially if the sampling time is too short (i.e. too shorter than reasonable time to expect a transition). If a specific transition has never been visited in the learning phase (i.e. $a_{k,l} = 0$), then such a transition will never be detected in the testing phase even if it actually has occurred in the testing or the validation phase. Consequently, a HMM leant by this method is prone to errors in the testing phase since some states transitions are unlikely to be detected by the Viterbi algorithm as the algorithm will always assign zero likelihood for these transitions. On contrary, dominating transitions may get extra detections that they actually occur, which also likely to lead toward errors during the testing and validation phase.

4.3 A Quantized CS-HMM to Model Continuously Varying Home Loads

A quantized CS-HMM is proposed to model and disaggregate home loads with varying power consumptions. A method to estimate the transition matrix for the proposed CS-HMM is proposed aiming to mitigate the effect of the two extreme cases that commonly appear in the learning phase of standard HMMs

- Non-visited transitions in which a transition (e.g. $k \rightarrow l$) is never observed during the learning phase. In such cases, the empirical time occupancy method concludes that the corresponding transition probability $a_{k,l} = 0$ due to zero occurrences of the transition during the learning phase.
- Too frequent transitions which result in high transition probabilities (e.g. $a_{k,l} \approx 1.0$) that dominate other transitions probabilities in the same row of the transition matrix.

To model a varying home load using a quantized CS-HMM, it is reasonable to firstly quantize the range in which the varying load is fluctuating in power consumptions into N equally-spaced levels, where each level center is denoted as c_k. The power consumption at a specific consumption level can be assumed to follow a normal distribution pattern as presented and discussed later in section 5.1. Hence, the observed measurements of power consumption of an appliance (p_m) operating at some quantized level k can be modeled using the following normal distribution pattern

$$p_m \sim \mathcal{N}(c_k, \sigma_k) \tag{4.7}$$

where $\mathcal{N}(c_k, \sigma_k)$ represents a normal distribution fitting of the consumed power by the appliance p_m with mean value of c_k and a standard deviation of σ_k.

To estimate the transition matrix A, the following framework in the learning phase are applied with consideration to the above-mentioned concerns

1. Initialize both the initial prior probabilities (π_i) and elements of the transition matrix (A) with uniformly distributed probabilities as presented in section 2.3. To some extent, this step supports the possibility to detect whatever transitions appear during the testing phase, even if they were never visited during the learning phase. Overall, using an initialization with a uniform distribution mitigates the extreme cases of never-visited transitions during the learning phase of the proposed CS-HMM.

2. Thereafter, it is proposed to modify the method of time occupancy to estimate transitions probabilities in a way that mitigates the cases of domination of one transition over others in the same row of the transition matrix. The modified method aims to smoothen the assigned probabilities by sharing a portion of the transition probability to the direct neighboring states.

To illustrate the effect of domination of a specific transition probability, the case of two adjacent levels k and $k+1$ from the quantized CS-HMM is considered as shown in Figure 4.1. A transition is assumed to be done by the varying load when it varies its power consumption from x_t at time t to x_{t+1} at time $t+1$, and there is a distance of δ between x_{t+1} and the upper edge of the level k as shown in Figure 4.1.

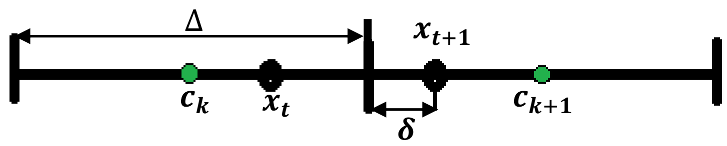

Fig. 4.1: Adjacent levels in the quantized CS-HMM.

To show how a transition probability domination may affect the inference process, it is assumed that $a_{k,k} \gg a_{k,k+1}$ which may lead the model be unable to detect the transition from level k to level $k+1$. Figure 4.2 shows an example with a reasonable value of $\Delta = 10\ W$ and the maximum potential chance for state switching when $a_{k,k+1} = 1 - a_{k,k}$ (i.e. other transitions in the same row of the transition matrix are neglected).

The likelihood to detect the new state at $k+1$ is given by (which also appears as area 2 in Figure 4.2)

$$P(s_t = k+1 | s_{t-1} = k) = a_{k,k+1}\, \mathcal{N}(\frac{\Delta}{2} - \delta, \mu = c_{k+1}, \sigma = \sigma_{k+1}) \qquad (4.8)$$

Likewise, the likelihood to detect the same state at k is given by (which also appears as area 1 in Figure 4.2)

$$P(s_t = k | s_{t-1} = k) = a_{k,k}\, \mathcal{N}(\frac{\Delta}{2} + \delta, \mu = c_k, \sigma = \sigma_k) \qquad (4.9)$$

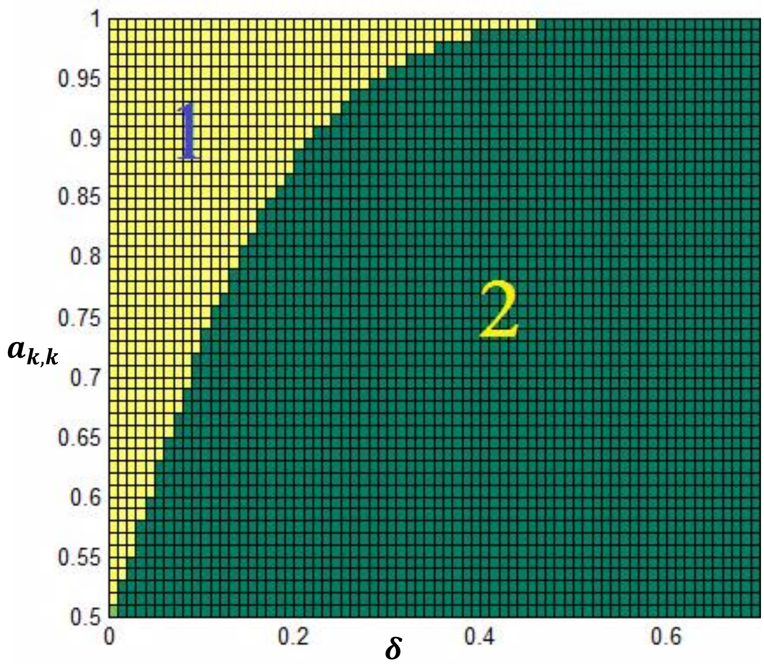

Fig. 4.2: Likelihoods of transiting from level state k to k, or to $k + 1$.

From Figure 4.2, the following could be concluded:

- It becomes more unlikely to detect a transition in cases of higher values of $a_{k,k}$ for farther deviations of an observed x_t away from the edge of the neighboring level. This observation concludes that the more a transition $a_{k,k}$ is dominant (bigger than others in same row of the transition matrix), the more unlikely to detect the correct transition by the model.
- In the extreme case where $a_{k,k} = 1$, the model would be stuck in this case and may never be able to detect future states transitions. Likewise, the more a transition $a_{k,k}$ is close to 1, the more likely a model transition is not detected.

From the above, a method is needed to smoothen such possible domination of a transition probability. However, such method should keep transitions probabilities indicative of the actual cases of transitions within the CS-HMM. In general, a state transition may dominate other transitions (in the same row of the model transition matrix) in the following cases:

- The transitions itself is too frequent that its corresponding transition probability (e.g. $a_{k,l}$) is noticeably much larger than other transitions probabilities in the same row of the transition matrix. Even in these cases, there is still a need to prevent extreme transitions from affecting the detection of other transitions, as illustrated earlier.
- The effect of a too compact sampling time that is much shorter than the average time occupied by the model at a specific state. This is likely to increase the self-state transitions probabilities (i.e. the diagonal items $a_{k,k}$ for $k = 1, 2, \ldots N$, of the transition matrix).

To perform a reasonable smoothing technique, a triangular probability redistribution is proposed with consideration of sharing a transition probability with direct neighboring states, whenever a transition $k \to l$ is detected as illustrated in Figure 4.3.

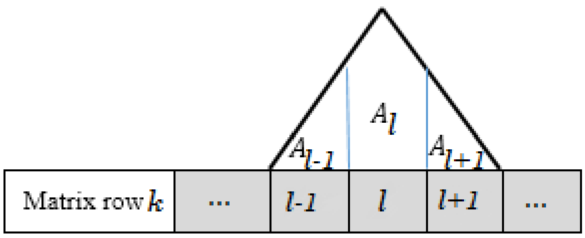

Fig. 4.3: Estimating a specific transition $k \to l$.

Figure 4.3 shows an example transition $k \to l$, where area of the triangular probability redistribution at a given iteration in the learning phase is given by

$$A_{l-1} + A_l + A_{l+1} = \frac{1}{n_k} \tag{4.10}$$

where the areas A_{l-1}, A_l and A_{l+1} are the additional probabilities to be added to transitions $a_{k,l-1}, a_{k,l}$ and a_{l+1} respectively, n_k is the total number of times when the model operated at state k. The essence of equation (4.10) is to redistribute the time occupancy probability (which is escalated each iteration by a value of $\frac{1}{n_k}$) by sharing a portion with direct neighboring states and same time keeping a substantial value for the estimate of $a_{k,l}$ to maintain its indication of the ground truth probability of the corresponding transition. The time occupancy method, on contrary, assigns the whole probability to the corresponding transition based on time of occurrences (e.g. the case of equation (4.10) becomes $A_l = \frac{1}{n_k}$ for each iteration in the learning phase).

Thereafter, the new estimates of the transitions probabilities (denoted as $\dot{a}_{k,l-1}, \dot{a}_{k,l}$ and $\dot{a}_{l+1}$) are obtained by adding the respective probabilities portions to the current corresponding values as follows

$$\dot{a}_{k,l-1} = a_{k,l-1} + A_{l-1} \tag{4.11}$$

$$\dot{a}_{k,l} = a_{k,l} + A_l \tag{4.12}$$

$$\dot{a}_{k,l+1} = a_{k,l+1} + A_{l+1} \tag{4.13}$$

Finally, it is crucial to normalize the affected matrix row (k) to maintain transitions probabilities that sum to unity. Hence, for $l = 1, 2, \dots N$, the following is applied

$$\tilde{a}_{k,l} = \frac{\dot{a}_{k,l}}{\sum_{l=1}^{N} \dot{a}_{k,l}} \tag{4.14}$$

Therefore, it is concluded that

$$\sum_{l=1}^{N} \tilde{a}_{k,l} = 1 \tag{4.15}$$

The above approach, presented by equation (4.10) through equation (4.14), is then iteratively applied during the learning phase so as to estimate the transition matrix of the quantized CS-HMM.

To compare the performance of the proposed methods above to some reference methods, three possible quantized CS-HMMs are presented

- ***Model A*** *is a quantized CS-HMM that uses the Viterbi algorithm (at maximum likelihood) for power estimation and uses the time occupancy method to learn the transitions matrix.* This model is a standard quantized CS-HMM that can be used as a reference model for the purpose of benchmarking and comparisons with the proposed models that incorporate modified methods to estimate the transition matrix.
- ***Model B*** *is a quantized CS-HMM that uses the Viterbi algorithm (at maximum likelihood) for power estimation and uses the proposed modified method to learn the transitions matrix.* This model is proposed by incorporating the proposed modified method to estimate the transition matrix of the quantized CS-HMM. The interest here is to test the performance of this model compared to the performance of the standard ***Model A*** as described above. In brief, the model B adds the layer of the modified method learn the transition matrix to the standard model A.
- ***Model C*** *is a quantized CS-HMM that uses the Viterbi algorithm (collective mean) for power estimation and uses the proposed modified method in this Chapter to learn the transitions matrix*. This model is proposed by incorporating the proposed modified method to estimate the transition matrix of the quantized CS-HMM. The interest here is to test the performance of this model compared to the performance of the previous ***Model B*** by applying the collective mean method to estimate the power consumption of the modeled continuously varying home load. In brief, the model C is similar to model C except in the method that is used for estimations (which the collective mean in model C).

4.4 Estimation of Power Consumption of Continuously Varying Loads Using Quantized CS-HMM

In general, Viterbi algorithm is used to infer the most probable hidden sequence of states, given a sequence of measurements or observations. However, in the domain of non-intrusive load disaggregation, there is an additional vital role for the Viterbi algorithm which is to estimate the power consumption of the modeled loads at every time instant (i.e. in parallel to each item of the observation sequence). Estimation of power consumption of a varying load, that is modeled by a quantized CS-HMM, by means of the Viterbi algorithm can be carried out using two methods

1. Using the maximum likelihood from the Viterbi outcome to assign the corresponding level center as an estimate of the consumed power at that time instant. That is, whenever a state or a quantized level k is inferred at time t for the appliance m, then the estimated consumed power can be assigned as $\hat{p}_m(t) = c_k$.
2. Using the collective mean (expectation) from the outcomes of the Viterbi algorithm. That is, at a time instant t, the Viterbi outcomes indicate the likelihoods of hidden states as a vector V, which can be written as

$$V = [v_1 = p(s_t = 1), v_2 = p(s_t = 2), \dots, v_N = p(s_t = N)] \quad (4.16)$$

The collective mean is the convex combination that uses the normalized version of V and levels centers to estimate the power consumed by a specific appliance m at time t. That is,

$$\hat{p}_m(t) = \sum_{k=1}^{N} c_k \frac{v_k}{\sum_{n=1}^{N} v_n} \quad (4.17)$$

where the term $\frac{v_k}{\sum_{n=1}^{N} v_n}$ is essential to normalize the values in the Viterbi vector V. Overall, the estimated power consumption $\hat{p}_m(t)$ can be regarded as the expectation of a probability function whose items are the quantized level centers and corresponding probabilities of the normalized values from the outcome of the Viterbi algorithm at some time t.

4.5 A Proposed Framework for Learning in the Hybrid CS/DS-FHMM Applied to Load Disaggregation

The hybrid CS/DS-FHMM combines both the CS-HMM and the DS-HMM into one FHMM, where some of hidden states chains model discrete loads (i.e. ON/OFF appliances and FSMs) and some other hidden states chains model continuously varying loads. Dynamic Bayesian networks, which generalizes the structure of HMMs, provide modeling and inference of cases of hybrid or mixed discrete-continuous systems [Murphy, 2002]. One of the main contributions presented here is the framework of learning and estimation applied by the proposed hybrid FHMM to NILM rather than its hybrid or mixture structure. The proposed learning and estimation framework applied by the hybrid FHMM is tailored to be applied on the problem of load disaggregation. The observed aggregate measurements represent the total sum of individual loads consumptions apart from their type. In the learning phase, the transition matrix of the CS-HMM component of the hybrid FHMM are learned. Besides, it is also necessary to model the profiles of power consumption of the existing appliances. The proposed learning framework can be summarized as follows

- For discrete loads such as ON/OFF devices and finite state machines (FSMs), the appliance is modeled by its characterizing transitions from the total aggregate signal (i.e. switching ON/OFF) same as applied in Chapter 3. These ON/OFF (i.e.

rising/falling) transitions from the aggregate measurements are essential features to model and detect the corresponding appliance.

- For the continuously varying loads, the model is learnt and the transition matrix of the CS-HMM is estimated using portions of the aggregate signal where the continuously varying load is the *only* ON device. This condition of being the *only* ON devices is necessary to avoid the interference or fluctuations of power consumptions from other discrete loads (e.g. ON/OFF loads and FSMs).

4.6 Estimation of Appliances Power Consumptions Non-Intrusively Using the Hybrid CS/DS-FHMM

The principal objective of an approach for load disaggregation is to estimate the power consumption of individual household appliances merely from the total aggregate measurements signal. The hybrid CS/DS-FHMM can be used to model a mixture of discrete and continuously varying loads as explained earlier. The estimation process depends on the types of the detected loads that operate at a specific time.

For ON/OFF loads and FSMs, instead of using the average consumptions, a first order hold (FOH) approach to estimate the power consumption of each load over the period of operation was used. The average of consumptions of appliances was not used because fluctuations in measurements while a discrete appliance is ON are possibly affected by the variations due to the operation of the varying loads, which exist together beside the discrete loads. Hence, using averaged models for appliances from measurements samples may not be accurate in representing the behavior of the discrete loads only. The zero-order hold (ZOH) approach was not used because real appliances consumptions profiles rarely show straight horizontal consumption patterns over the operation period (i.e. ZOH results in zero-order models, which are horizontal lines). Many appliances, when they are switched ON, showed an overshoot starting power consumptions that is usually above their settled power consumption level. In addition, an appliance power consumption may increase or decrease during their operations based on the type of the appliance and its operation cycle as explained earlier in section 3.4. To illustrate this method, an example is presented for an appliance m that was detected at time t_x with an ON transition of value $p_m(t_x)$, and with an OFF transition at time t_y of value $p_m(t_y)$, then an estimate $\hat{p}_m$ at time t may be considered as follows

$$\hat{p}_m(t) = \frac{p_m(t_y) - p_m(t_x)}{t_y - t_x}(t - t_x) + p_m(t_x), \qquad for\ \ t_x \leq t \leq t_y \tag{4.18}$$

On the other hand, the estimation process for the continuously varying load depends on whether there are other appliances operating in the ON state simultaneously with the varying load. If the varying load is the only ON device, the Viterbi algorithm can then be applied to the modeled quantized CS-HMM as presented in section 4.4. In cases when there are other operating (ON) discrete loads at same time instant, it is proposed to predict the operation of the varying load using a method analogous to that in equation (4.2) for the discrete domain with replacement of the integral to a summation operator. This prediction method is proposed because it becomes too challenging to distinguish between normal signal fluctuations/noise due to discrete loads and fluctuations due to the operation of the varying load.

Generally, there are two basic remarks about the proposed hybrid CS/DS-FHMM which are summarized as follows

- To model a continuously varying home load from the total aggregate signal, observed portions of measurements or sequences where it is the *only* ON device are required. Such observations are not required for discrete loads modeled by FHMMs as they can be detected and modeled from significant transitions in the total aggregate measurements.

- The method of the learning and inference in the CS-HMM can be applied to one continuous varying home load only. If there are two or more varying loads, they can be modeled as one lumped varying load. However, several varying loads lumped together may produce irregular consumption patterns that it becomes more challenging to be learnt or estimated by the CS-HMM.

4.7 Case Study Application on Varying Home Loads and Applying the Hybrid CS/DS-FHMM on Load Disaggregation

The quantized CS-HMM approach proposed in this Chapter to model and disaggregate continuously varying home loads was applied on synthetic data and on real home loads data from the REDD public data set. Comparison of performance of different possible estimations methods is presented and results are discussed in detail. The proposed framework of the hybrid CS/DS-FHMM was tested by applying to modeling and disaggregating of various types of home loads merely using their total aggregate measurements. A 9- sample median filter on both training and testing phases was applied to remove noisy signals and outliers. The used accuracy metric to assess models' performance is the accuracy of estimating an appliance power consumption compared to the actual power consumption as presented earlier in subsection 2.2.4

4.7.1 Applying CS-HMM on a Single Continuously Varying Home Load

The proposed approaches for both learning and power estimation of continuously varying home loads were tested on synthetic data and real home loads. Firstly, a set of synthetic data was generated to simulate a light dimmer that represents a typical varying home load with power consumptions ranging from $0\ W$ to $150\ W$. The transitions were generated to switch to both direct neighboring and far states. The model was learnt from a portion of the generated data and then validated on the remaining portion.

Table 4.2 compares the performance, three possible quantized CS-HMMs (as described earlier) were tested

- ***Model A*** *is a quantized CS-HMM that uses the Viterbi algorithm (at maximum likelihood) for power estimation and uses the time occupancy method to learn the transitions matrix.*
- ***Model B*** *is a quantized CS-HMM that uses the Viterbi algorithm (at maximum likelihood) for power estimation and uses the proposed modified method to learn the transitions matrix.*
- ***Model C*** *is a quantized CS-HMM that uses the Viterbi algorithm (collective mean) for power estimation and uses the proposed modified method in this Chapter to learn the transitions matrix.*

Table 4.2: Comparison of models' performance applied on a generated data set.

Number of quantization levels	**Model A**	**Model B**	**Model C**
5 levels	95.8%	95.9%	96.7%
10 levels	97.6%	97.7%	98.6%

To apply the CS-HMM on real home loads from the REDD public data set, the power consumptions of the electronics loads were observed and found to be considerable as a continuously varying load. Table 4.3 shows the performance results in terms of accuracy of power estimation obtained when applying different models (as described earlier) on real loads from the REDD.

Table 4.3: Comparison of models' performance applied on real loads.

Load	Model A	Model B	Model C
Load A	95.4%	95.4%	96.2%
Load B	72.4%	72.8%	73.1%

In Table 4.3, load A represents an electronics load from house 3 and load B represents an electronics load from house 5 from the REDD public data set.

4.7.2 Applying the Hybrid CS/DS-FHMM on Various Types of Home Loads

The proposed framework for the hybrid CS/DS-FHMM for both learning and power estimation of various types of home loads was tested real home loads. The proposed framework for the hybrid CS/DS-FHMM was applied on real data from house 3 from the REDD public data set. Measurements of four appliances in house 3 were used which include the electronics load as a varying load. In addition, the refrigerator, the washer-dryer and the microwave were used to represent discrete loads (i.e. FSMs).

The proposed framework of the hybrid CS/DS-FHMM learnt appliances models from a portion of the total lumped signal (aggregate consumptions). Thereafter, they were tested on subsequent sequences. To illustrate the different methods that were used to estimate the transition matrix of the quantized CS-HMM of the varying loads, a case is presented where the size of the transition matrix is 5×5.

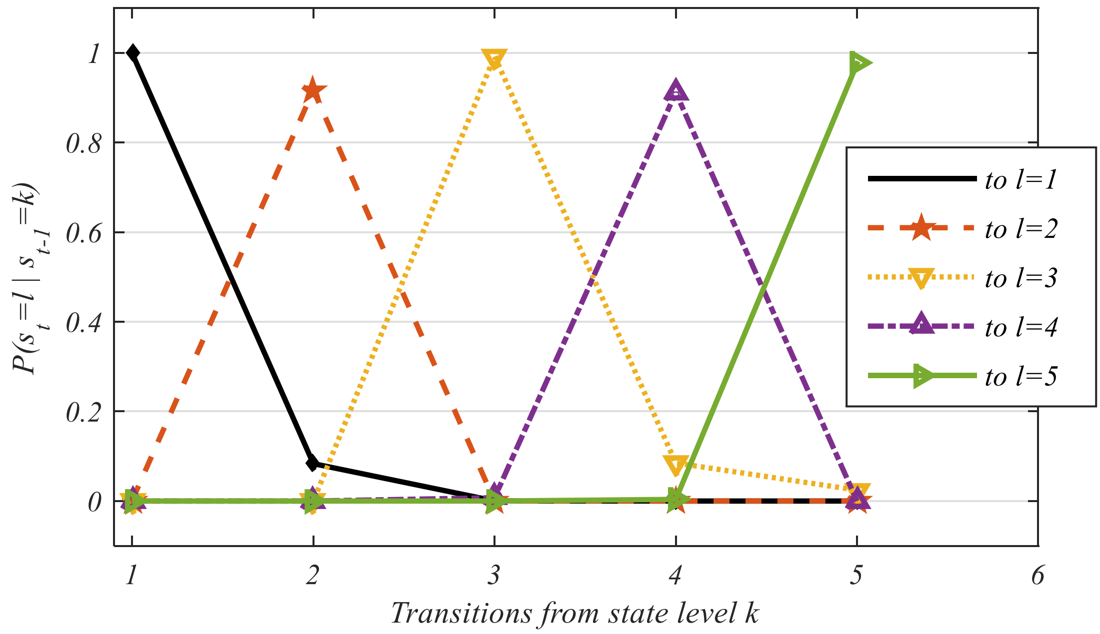

Fig. 4.4: Transition matrix using empirical time occupancy method.

Figure 4.4 shows a lines graph of the transition matrix representing $P(s_t = l|s_{t-1} = k)$ estimated using the empirical time occupancy method. Applied on the same measurements sequence as of Figure 4.4, Figure 4.5 shows a lines graph of the transition matrix representing $P(s_t = l|s_{t-1} = k)$ estimated using the modified method to estimate the transition matrix of the CS-HMM.

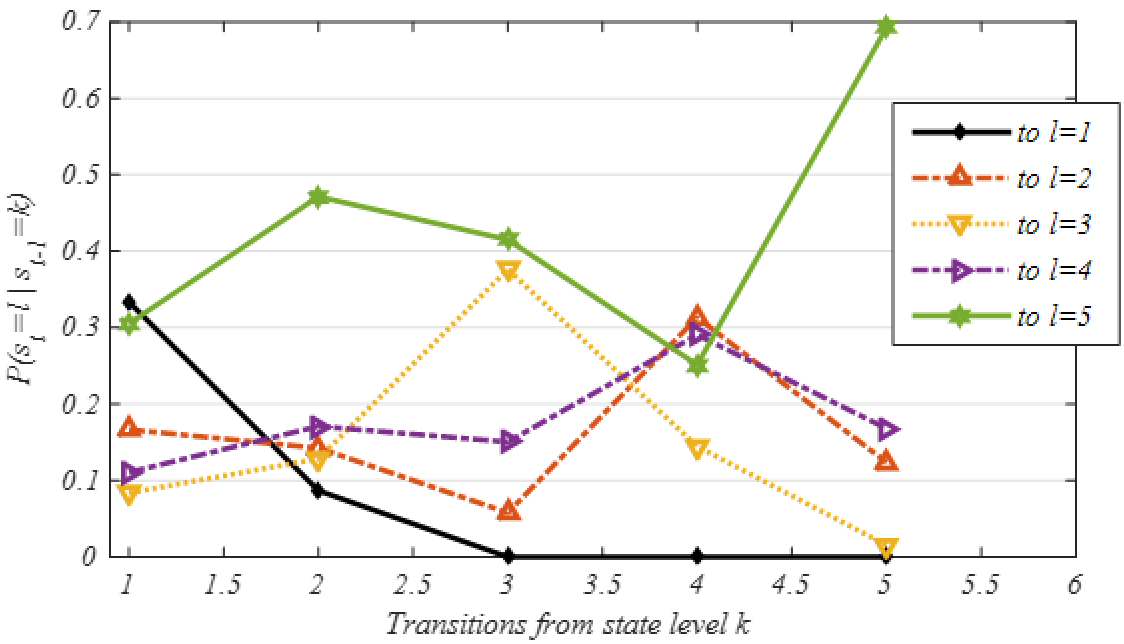

Fig. 4.5: Transition matrix using the modified method.

The proposed hybrid CS/DS-FHMM approach was then tested by disaggregating the mixture of appliances merely from the total aggregate signal. The power consumption of the continuously varying load (electronics) was obtained as shown in Table 4.4 which compares the performance of different possible estimation methods.

Table 4.4: Disaggregation accuracies of the varying load using different methods.

Method of estimating the transition matrix	**Viterbi algorithm (maximum likelihood)**	**Viterbi algorithm (collective mean)**
Time occupancy	78.86%	78.87%
Modified method	79.81%	79.55%

The discrete loads (i.e. FSMs) were disaggregated using the method explained earlier in section 4.6 and the results obtained were as shown in Table 4.5.

Table 4.5: Disaggregation accuracy of discrete loads.

Refrigerator	**Washer dryer**	**Microwave**
67.3%	57.9%	52.3%

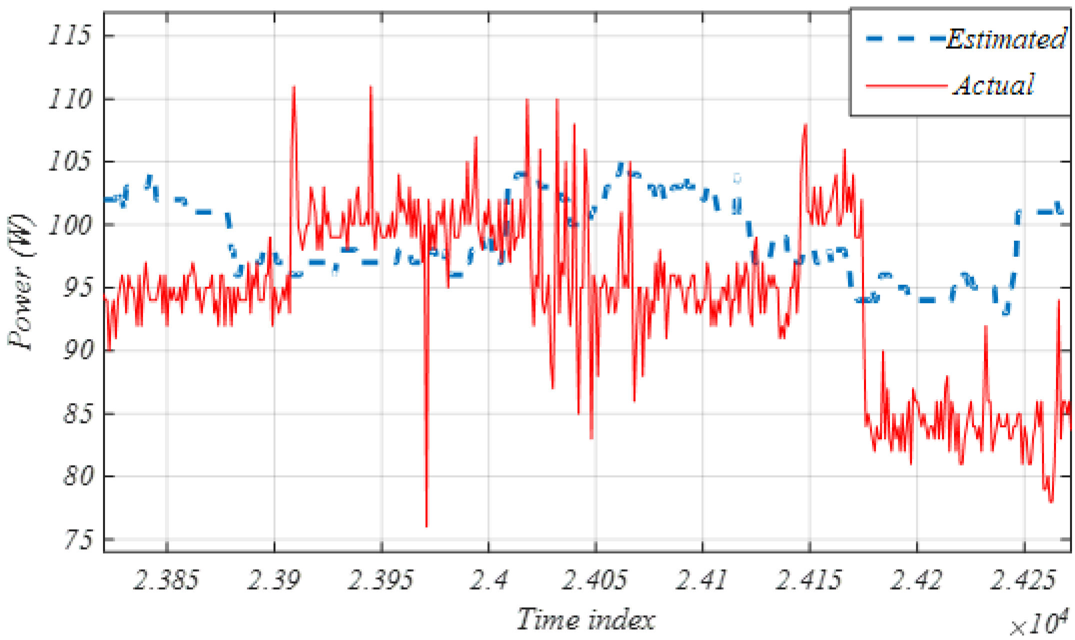

Fig. 4.6: Estimated versus actual power consumption of the varying load.

Figure 4.6 and Figure 4.7 show examples for illustration of the estimated versus actual power consumptions of the varying load (electronics) and the refrigerator, respectively.

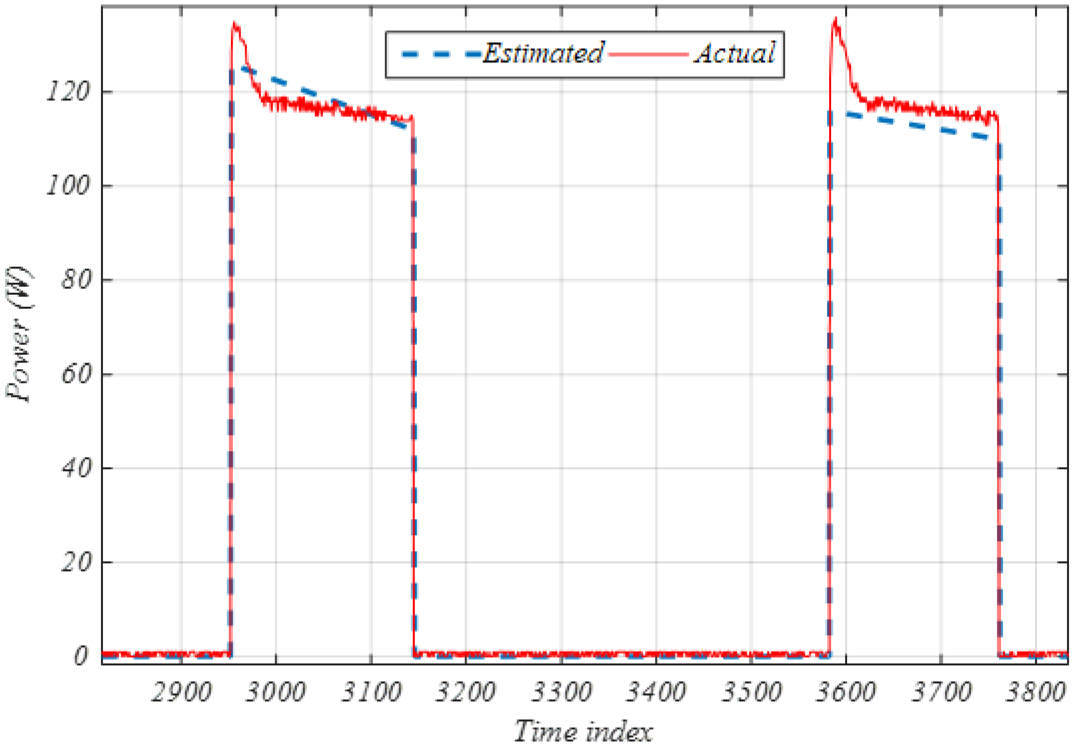

Fig. 4.7: Estimated versus actual power consumption of the refrigerator.

4.7.3 Interpretations of Varying Loads Disaggregation Results

The proposed quantized CS-HMM with the modified method to estimate the transition matrix showed better performance compared to the method using the common empirical time occupancy estimation. The framework of the hybrid CS/DS-FHMM demonstrated its capability to model and estimate the power consumptions of a continuously varying load and other discrete loads simultaneously from the aggregate total measurements.

Increasing the number of quantization levels in the CS-HMM, though it may increase the computational complexity, enhanced the obtained accuracies of estimating the power consumption of the varying load as shown in Table 4.2. As an illustration of computational complexity, increasing the number of quantization levels to 10, in Table 4.2, requires an additional 3.1% increase in computational time compared to using 5 levels for the same training and testing data set. In Tables 4.3 and 4.4, it is shown that the modified transition matrix outperforms the empirical method based on time occupancy in terms of accuracy of power estimation. The modified method to estimate the transition matrix has mitigated two extreme possible cases

- The non-visited transitions in the learning phase.
- The too-frequent or dominating transitions cases.

These two extreme cases often reduce the performance of HMM that is learnt by the empirical time occupancy method.

The collective mean method used in the power estimation process provided better performance than using the corresponding level center where maximum likelihood occurs. This is likely to happen since using levels centers will always produce a stepping up/down signal. Meanwhile, the collective mean could estimate the actual signal without any restrictions in the assigned estimates. The accuracies obtained as shown in Table 4.2 and Table 4.3 are for only a single varying load that was used in both the learning and testing phases, which is not the case in the hybrid CS/DS-FHMM in the presence of discrete loads (i.e. FSMs).

To estimate the transition matrix of the quantized CS-HMM during the learning phase, it was essential to observe sequence of the total aggregate signal where varying load is the only ON device. Figure 4.4 shows the existence of both extreme cases of transitions: non-visited transitions during the learning phase when $P(s_t = l|s_{t-1} = k) \approx 0$ (e.g. $a_{1,2} \approx 0$ and $a_{4,5} \approx 0$), and too frequent or dominant transitions when $P(s_t = k|s_{t-1} = k) \gg P(s_t = l|s_{t-1} = k), for\ l \neq k$ (e.g. $a_{1,1} \approx 1$ and $a_{3,3} \approx 1$). These cases were generally mitigated by the modified transition matrix method as shown in Figure 4.5 where the extreme cases were smoothened if compared to the corresponding transitions probabilities seen in Figure 4.4.

Dealing with varying home loads in both the training and testing phases differs from the cases of dealing with only discrete loads because the discrete loads usually show characterizing features (i.e. significant transitions) in the aggregate signal. On contrary, the power consumptions of the varying home loads often change smoothly without noticeable edges. Hence, they should be modeled separately by a CS-HMM in order to fit their nature of varying power consumptions. The first order hold (FOH) is an estimation method that estimate an appliance power consumption using the ON and OFF transitions only. It is not recommended to estimate discrete appliances consumptions profiles using averaged consumptions from the data stream since there is a possibility that the varying load is affecting/operating in those data sequences.

Some discrete appliances such as the microwave showed a low disaggregation accuracy. In general, the accuracy of disaggregation of the discrete loads were found close to those achieved using similar methods that apply standard FHMMs to model and disaggregate the discrete appliances. For example, Johnson and Willsky [Johnson and Willsky, 2013] achieved a load disaggregation accuracy of 33.3% when applying the standard FHMM on house 3 from REDD public data set. Besides, Kolter and Johnson [Kolter and Johnson, 2011] achieved load disaggregation accuracies in the range from 33.3% to 59.6% when applying the reference FHMM on house 3 from the REDD public data set.

Overall, the proposed framework of the hybrid CS/DS-FHMM showed good performance in modeling and disaggregating various types of varying and discrete loads merely from the total aggregate measurements signal.

4.8 Remarks on the Proposed Approaches in this Chapter

The proposed approaches aimed to model continuously varying home loads using a quantized CS-HMM. In addition, a framework for learning and disaggregation phases was proposed for the general hybrid CS/DS-FHMM that models various types of loads. Some remarks were noticed in the proposed approach, which can be summarized as follows

1. Quantizing the continuous space of the hidden state of a varying home load compromises for the accuracy of appliance modeling, as it substitutes a continuous state space by a finite number of discrete states. Nonetheless, quantization of continuous spaces mitigates the potential computational complexity that arises when dealing with continuous or infinite domains.
2. In the learning phase of the hybrid CS/DS-FHMM, it was necessary to observe measurements sequences when the varying load is the only ON appliance so as to model its power consumption and to estimate the transition matrix of the CS-HMM component in the CS/DS-FHMM.
3. The proposed approach could model and disaggregate one varying load in the household. In cases where there are two or more varying loads in a household, they would be lumped together and assumed to be a one varying load by the proposed approach. However, the more there are varying loads, the more unlikely that their lumped model to have a consistent pattern of power consumption (repeatable shape/style of the power consumptions by the lumped varying loads).
4. In the disaggregation phase of the hybrid CS/DS-FHMM, it was proposed to estimate power consumptions of discrete loads (e.g. FSMs) using a first order hold (FOH) technique which assumes a linear modeling of power consumption by discrete-states home appliances.

4.9 Summary

This Chapter presented the theory and case study applications of the second proposed approach that aims to tackle the problem of modeling and disaggregation of continuously varying home loads. A quantized CS-HMM was proposed with a method to estimate the transition matrix that mitigates two possible extreme cases of transitions. The proposed quantized CS-HMM was consolidated with the standard FHMM to model and disaggregate the power consumptions of appliances only from their total aggregate measurements. The proposed models and approaches were tested on synthetic data and on real home appliances from the REDD public data set.

Chapter 5

Disaggregating Overlapping Home Appliances Using a Clusters Splitting Approach

This chapter presents the third proposed approach that aims to enhance the overall performance of non-intrusive load disaggregation approaches based on clustering methods. An approach for splitting of overlapping appliances clusters is proposed based on a test of clusters cohesion. Testing of clusters cohesion is proposed by checking the degree of normality fitting of clusters using three normality tests.

Clustering methods were applied to the field of NILM, where the goal is to group home appliances into distinct clusters based on extracted features. Each resulting cluster from a clustering approach is supposed to represent one appliance. Goncalves *et al.* [Goncalves *et al.*, 2011] applied an *unsupervised* blind source separation technique to obtain appliance-level consumptions from the aggregate data. They utilized both genetic K-means and agglomerative clustering with features like $\Delta P - \Delta Q$ to cluster appliances. [Lin *et al.*, 2011] applied a fuzzy K-means clustering and optimization algorithms to identify the energizing and de-energizing statuses of each appliance. Load energizing and de-energizing transient features were extracted, and the fuzzy classifier was used to perform load identification based on these features. Shao *et al.* [Shao *et al.*, 2012] proposed a motif mining method to perform load disaggregation in an *unsupervised* fashion. Their method was mainly applicable to appliances with distinctive repeatable events. Kamoto *et al*. [Kamoto *et al*., 2017] presented a new approach based on competitive agglomeration which incorporates the good qualities of both hierarchical and partitional clustering aiming to carry out energy disaggregation to discover appliances without prior information about the number of appliances. In general, previous research investigated the usage of clustering in the domain of NILM without further investigations on the resulting clusters to test if they originated from two distinct appliances, in which a cluster splitting could be reasonable to disaggregate the overlapping appliances clusters.

Overlapping between obtained clusters of home appliances is a common challenge in the domain of NILM where two or more appliances consume close or similar amounts of power during their operation. Figure 5.1 shows some overlapping cases (particularly in low power consuming appliances) in the first NILM work by Hart [Hart, 1992].

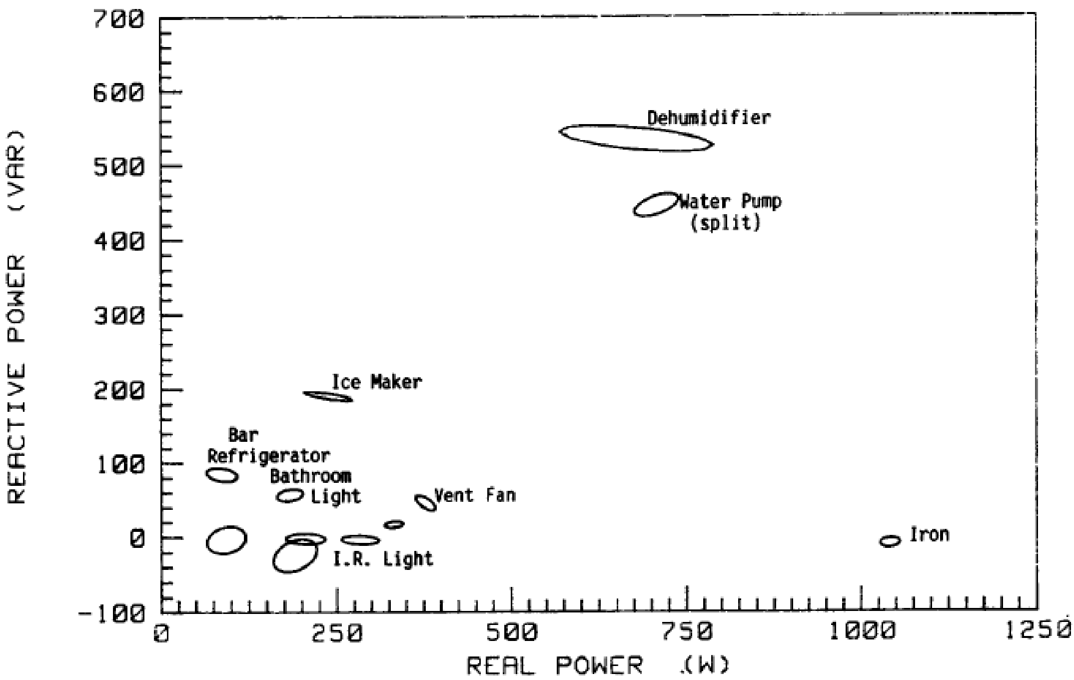

Fig. 5.1: Overlapping between appliances clusters [Hart, 1992].

Extracting additional distinct features of appliances may help in distinguishing between appliances of overlapping power consumptions. However, it may be another challenge to extract additional features for appliances especially if power measurements at low frequency rate are the only available information.

In this Chapter, a proposed approach to split overlapping clusters of home appliances is presented. The proposed approach is based on checking of clusters' cohesion to conclude if they were originating from two different sources (i.e. by two appliances); thereby to decide a cluster splitting. Testing of clusters cohesion was performed using common normality tests carried out against two confidence levels. Once a cluster splitting is decided, the splitting can be done using common clustering methods such as expectation maximization (EM).

5.1 Normality Assumption of Appliances Power Consumptions

The assumption that the pattern of power consumptions of home appliances can fit to a normal distribution was discussed and/or adopted in previous studies such as [Kim *et al.*, 2010; Kolter and Jaakkola, 2012; Parson *et al.*, 2014; Xie *et al.*, 2017]. Based on analysis of histograms of appliances consumptions, Kim *et al.* [Kim *et al.*, 2010] concluded that the normality assumption is valid for most of home appliances except for a TV and an office laptop (which fall under the category of electronics). Bases on data analysis together with normality tests like Kolmogorov–Smirnov (K–S test), Xie *et al.* [Xie *et al.*, 2017] found that it is reasonable to adopt the assumption of normality fitting in applications in modern grids (e.g. probabilistic load forecasting), especially when there are no enough resources to build comprehensive underlying models. Besides, applying the normality assumption (taken as undisputed) in previous research (e.g. as in [Kolter and Jaakkola, 2012; Parson et al., 2014]) yielded in models and/or approaches that achieved good performances in the domain of NILM. With the normality assumption of appliances power consumptions, finite state machines (FSMs) can also be modeled using a Gaussian Mixture Model (GMM) [Benaglia *et al.*, 2010]. Nonetheless, the power consumption of a household appliance could be complicated in its pattern and some appliances may show poor normality fitting. Thus, the following considerations were taken into account to maintain good normality fitting of home appliances for the purpose of load disaggregation

1. Only appliances (including FSMs) with stable power consumptions were considered in this approach in order to comply with the normality assumption. Continuously varying loads (e.g. light dimmers and electronics) were neglected as they may show varying or inconsistent power consumptions during their operation.
2. Some appliances show higher/lower power consumption levels within a short time of their switching to ON/OFF as presented in section 3.4. For these cases, such periods of rising or falling in power consumptions were neglected and only periods of stable power consumptions were considered. The significant rising/falling in power consumption level of an appliance may distort its normality fitting, thus they were ignored in this study.
3. Measurements of a household total aggregate consumption usually include noisy signals. A 9-sample median filter was applied to remove noisy signals and outliers which may alter the normality of the power consumed by home appliances.

As majority of common smart meters mainly provide measurements of real power consumptions of residential households, an appliance power consumption (p_m) can be modeled by a one-dimensional normal distribution as follows

$$p_m \sim \mathcal{N}(\mu_m, \sigma_m) \quad (5.1)$$

where μ_m is the mean value of the normal distribution and indicates the average power consumed by the appliance m, and σ_m is the standard deviation of the power consumption which indicates fluctuations in power consumption of an appliance. Kim *et al.* [Kim *et al.*, 2010] found that the power consumption of an appliance often fluctuates within $\pm 20\%$ from its average value.

5.2 Checking Clusters Cohesion to Decide Splitting

The normality fitting as expressed earlier can be used to describe the behavior or consumptions pattern of an appliance. The clusters cohesion test aims to investigate if the cluster could be originating from two different sources (i.e. two overlapping appliances) by checking the degree of normality distortion in the cluster. That is, if the cluster of an appliance has originated purely by one appliance, it should show a good normality fitting (i.e. good level of cohesion between items in the cluster). On contrary, if two appliances with overlapping power consumptions have merged in one cluster, the cluster may show poor level of cohesion due to altered normality by the merge of items (i.e. measurements) from two appliances into one cluster.

To encompass different aspects of normality fitting when carrying out cohesion tests on clusters, it is recommended to apply more than one normality test like the following

- Kolmogrov-Smirnov (KS) test [Massey, 1951; Marsaglia *et al.*, 2003].
- Anderson-Darling (AD) test [Anderson and Darling, 1954].
- Jarque-Bera (JB) test [Jarque and Bera, 1987].

These tests are used to encompass possible variabilities within the tested samples. For example, while KS test checks normality by testing the distance between the cumulative distribution function (CDF) of the reference distribution and the empirical distribution of the sample, the AD test gives more weight to tails of the distribution. Besides, JB test detects normality by checking the skewness and kurtosis of the tested sample [Razali and Wah, 2011].

Though the mechanisms of normality tests may differ, a statistical hypothesis is usually performed to check for normality as following

- *Null hypothesis, $H0$: the tested sample fits a normal distribution.*
- *Alternate hypothesis, $H1$: the tested sample does not fit a normal distribution.*

The outcome of the hypothesis testing can be given in several ways, to decide the conclusion of the test, such as the test statistic, confidence intervals and the *P-value* [Montgomery and Runger, 2011]. In general, there should be some threshold confidence level $(1 - \alpha)$ used for the purpose of obtaining a test conclusion. Note that α is also called the probability of committing type I error which indicates the likelihood of rejection of $H0$, while $H0$ is correct. Common used values for α are in the range from 0.001 to 0.10. For more details on confidence levels and hypothesis testing, refer e.g. to [Montgomery and Runger, 2011].

Using the *P-value* gives a good indication of how likely the tested sample follows a normal distribution. Theoretically, a *P-value* can fall in the following range

$$0.00 \leq P - value \leq 1.00 \tag{5.2}$$

where larger *P-values* mean better normality fitting of the tested sample. The ideal case of $P - value = 1.0$ represents the perfect ideal normal distribution. Hence, the *P-value* may be used as an indicator of the degree of a cluster cohesion, where a cluster cohesion is understood as the degree that its items show a good normality fitting. Therefore, it can be written that

$$Cohesion \approx P - value \tag{5.3}$$

The conclusion of the hypothesis test is obtained as follows

$$Test\ conclusion = \begin{cases} Fail\ to\ reject\ H0, & P-value > \alpha \\ Reject\ H0, & P-value \leq \alpha \end{cases} \tag{5.4}$$

In order to perform a cohesion check based on normality tests, it is proposed to run the three normality tests on each cluster, then averaging the outcome *P-values*, denoted later as $\overline{PV}$, so as to encompass major variability aspects of the tested cluster. Thereafter, it is suggested to compare the resulting $\overline{PV}$ against two confidence levels (denoted as α_1 and α_2) to decide clusters splitting as follows

$$Splitting\ decision = \begin{cases} Split, & \overline{PV} \leq \alpha_1 \\ Test\ sub-clusters, & \alpha_1 < \overline{PV} \leq \alpha_2 \\ Do\ not\ split, & \alpha_2 < \overline{PV} \end{cases} \tag{5.5}$$

Equation (5.5) consider the convention that $\alpha_1 < \alpha_2$ to decide a cluster splitting based on the cohesion of the cluster which is represented by its degree of normality fitting. The three cases in equation (5.5) can be articulated in the following

- Cases when $\overline{PV} \leq \alpha_1$, which lead to decide a cluster splitting because of too poor normality fitting in the tested cluster. The distorted normality within a cluster indicates a poor cohesion degree, which could be caused by two different sources (i.e. home appliances) that originated one merged cluster. Hence, a decision to split the cluster is reasonable in this case.
- Cases when $\alpha_2 < \overline{PV}$, which lead to decide no splitting on the tested cluster. In these cases, the tested clusters show a good normality fitting which indicates a good degree of cohesion among cluster elements. Thus, it is reasonable not to decide a splitting on the cluster as it is likely has originated from one source (i.e. an appliance).
- Cases when $\alpha_1 < \overline{PV} \leq \alpha_2$, further normality tests on the possible sub-clusters are suggested to be performed before a decision is made. The motivation of testing the

sub-clusters is to check if better normality fitting in the sub-clusters is achievable more than in the original cluster. Testing both sub-clusters is proposed to be done in three subsequent steps as follows

1. Perform the cluster splitting as depicted in the section 5.3.
2. *Reconstruct* the sub-clusters to substitute the sharp cut of the splitting.
3. Apply cohesion test on both sub-clusters as performed earlier.

It is crucial to note that splitting in step 1 above is not a final decision till results from step 3 lead toward a splitting decision. The splitting of a cluster will often make a sharp cut in the assumed normal distributions as shown in Figure 5.2. Therefore, it is recommended to run normality tests on *reconstructed* small clusters as mentioned in step 2 above. A *reconstructed* cluster consists of the split cluster merged to a portion placed in the cut region that reflects the other side portion of the cluster as illustrated in Figure 5.2.

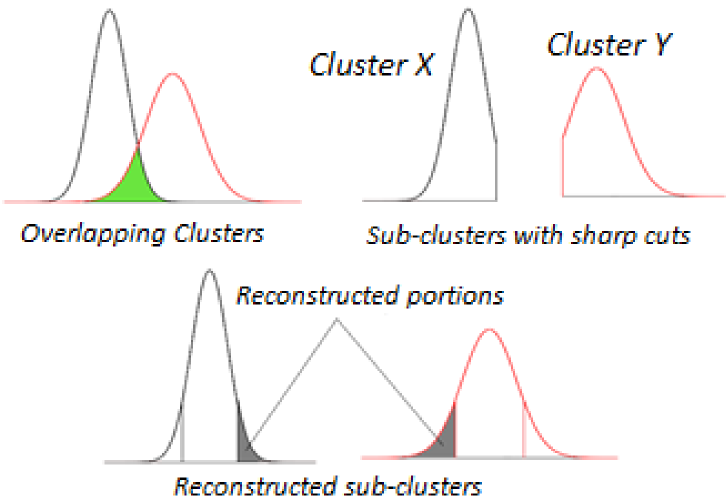

Fig. 5.2: Sharp cuts in the sub-clusters and reconstructed sub-clusters.

For illustration on how to obtain the reconstructed portions, we consider the overlapping cluster X and cluster Y as shown in Figure 5.2. The reconstructed portion of cluster X is defined as the items group x which is a subset of the cluster X where each item x_i satisfies the following constraint

$$\min(X) \le x_i \le \mu_X - |\max(X) - \mu_X| \tag{5.6}$$

where $\min(X)$, μ_X and $\max(X)$ are the minimum item value in the cluster X, the mean of the cluster X and the maximum item value in the cluster X after the temporary splitting, respectively. Similarly, the reconstructed portion of the cluster Y is defined as the items group y which is a subset of the cluster Y where each item y_i satisfies the following constraint

$$\mu_Y + |\mu_Y - \min(Y)| \le y_i \le \max(Y) \tag{5.7}$$

where $\min(Y)$, μ_Y and $\max(Y)$ are the minimum item value in the cluster Y after the temporary splitting, the mean of the cluster Y and the maximum item value in cluster Y, respectively. These portions of the clusters are then copied in the *inverse* order and appended to the other hand of the corresponding cluster so as to obtain the *reconstructed* version of the clusters.

The rationale behind obtaining these *reconstructed* small clusters is due to the fact that the sharp cut in one of sides of a cluster is more likely to lead to poor normality fitting within that cluster. The poor normality fitting is not due to how data is distributed among the cluster, but it is mainly due to the sharp splitting. Hence, it is proposed to perform normality tests on a *reconstructed* version of these sub-clusters. The reconstructed portion of these sub-clusters is taken typically the same as the corresponding present portion on the other side of the same sub-cluster. Consequently, normality tests results could be more reasonable in indication the degree of clusters cohesion; thereby guide toward better decision about the splitting.

Thereafter in step 3, the $\overline{PV}$'s of both *reconstructed* sub-clusters should be obtained separately to compare if any of these $\overline{PV}$'s is larger than the $\overline{PV}$ of the original cluster before the temporary splitting. In cases that any $\overline{PV}$ of the *reconstructed* sub-clusters is larger than that of the original clusters before the temporary splitting, it means that the sub-cluster shows a better normality fitting than the original cluster. Hence, each *reconstructed* sub-cluster could be originating from a different source and a splitting decision should be taken. In cases the contrary result appears, no splitting decision should be taken since the original cluster before the temporary splitting shows a better degree of cohesion among its items.

5.3 Executing Clusters Splitting

Once a decision is taken to split a cluster into two smaller clusters, it is proposed to perform some inner clustering method like expectation maximization (EM) which will assign the original cluster elements into the two small clusters. It is proposed to initialize the EM with two centroids C_1 and C_2 to be 2σ distance from the mean value of the original cluster. That is,

$$C_1 = \mu + 2\sigma \tag{5.8}$$

$$C_2 = \mu - 2\sigma \tag{5.9}$$

where μ and σ are the mean and the standard deviation of the original cluster before splitting, respectively. To initialize each small cluster with some elements, it is reasonable to assign elements according to which minimum distance from the centroids. Finally, an EM clustering can be applied to obtain the final version of the two split clusters (the sub-clusters that are supposed to represent two distinct home appliances).

The correct assignment of elements to the resulting small clusters is good measures to assess the accuracy of the splitting methodology. Since the assignment task is like a binary decision as we need only to assign an element to one of the two split clusters, it is reasonable to use the *Precision*, *Recall* and *F-measure* to assess the accuracy of elements assignment after splitting as explained earlier in subsection 2.2.4.

5.4 Case Study Applications on Overlapping Clusters

The approach proposed in this Chapter to disaggregate overlapping cases of home appliances was applied on overlapping cases of six real houses data from the REDD public data set. The cases of overlapping appliances consumptions were investigated from the individual consumption (sub-metered at appliance-level) data from the REDD data set. Thereafter, the cohesion of appliances clusters (a merged cluster from two overlapping appliances) was checked and clusters splitting was carried out as explained earlier in this Chapter. Same as applied to other proposed approaches, a 9-sample median filter to was applied on the data to filter out noisy signals and outliers.

5.4.1 Investigating Cases of Overlapping Appliances Clusters

In this work, it was focused only on cases where there is a possibility of overlapping between appliances clusters. This usually occurs when two appliances or more have close power consumptions or overlapping power fluctuations at some states of their operation. From appliance-level data, possible cases of overlapping between appliances were searched in each house in the REDD separately. Table 5.1 through Table 5.6 show the detected possible cases of overlapping between clusters (which represent appliances) in house 1 through house 6 from the REDD public data set, respectively.

Table 5.1: Overlapping clusters in house 1 from the REDD.

Cluster number	Appliance type	Power consumption $\sim \boldsymbol{N(\mu, \sigma)}$ (Watts)	Overlaps with
Cluster 1	Oven	$N(1657.3, 11.6)$	Cluster 2
Cluster 2	Oven	$N(1687.3, 20.4)$	Cluster 1
Cluster 3	Microwave	$N(1506.9, 43.6)$	Cluster 4, cluster 5
Cluster 4	Bathroom gfi	$N(1597.4, 12.1)$	Cluster 3
Cluster 5	Kitchen outlet	$N(1529.9, 7.7)$	Cluster 3

Table 5.2: Overlapping clusters in house 2 from the REDD.

Cluster number	Appliance type	Power consumption $\sim \boldsymbol{N(\mu, \sigma)}$ (Watts)	Overlaps with
Cluster 1	Lighting	$N(155.7, 7.5)$	Cluster 2
Cluster 2	Refrigerator	$N(163.0, 6.7)$	Cluster 1

Table 5.3: Overlapping clusters in house 3 from the REDD.

Cluster number	Appliance type	Power consumption $\sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$ (Watts)	Overlaps with
Cluster 1	Electronics	$N(155.7, 7.5)$	Cluster 2, cluster 6
Cluster 2	Refrigerator	$N(98.8, 9.5)$	Cluster 1, cluster 6
Cluster 3	Lighting	$N(181.7, 18.3)$	Cluster 4, cluster 5
Cluster 4	Lighting	$N(189.9, 5.9)$	Cluster 3, cluster 5
Cluster 5	Lighting	$N(196.9, 11.8)$	Cluster 3, cluster 4
Cluster 6	Kitchen outlet	$N(127, 0.7)$	Cluster 1, cluster 2

Table 5.4: Overlapping clusters in house 4 from the REDD.

Cluster number	Appliance type	Power consumption $\sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$ (Watts)	Overlaps with
Cluster 1	Kitchen outlet	$N(137.1, 6.7)$	Cluster 2
Cluster 2	Lighting	$N(101.5, 15.5)$	Cluster 1, cluster 3
Cluster 3	Lighting	$N(93.9, 4.9)$	Cluster 2

Table 5.5: Overlapping clusters in house 5 from the REDD.

Cluster number	**Appliance type**	**Power consumption $\sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$ (Watts)**	**Overlaps with**
Cluster 1	Unknown outlet	$N(169.3, 7.6)$	Cluster 2
Cluster 2	Refrigerator	$N(164.2, 6.2)$	Cluster 1
Cluster 3	Lighting	$N(576.7, 17.1)$	Cluster 4
Cluster 4	Furnace	$N(553.4, 12.5)$	Cluster 3
Cluster 5	Sub-panel	$N(226.7, 4.2)$	Cluster 6
Cluster 6	Sub-panel	$N(216.8, 6.0)$	Cluster 5
Cluster 7	Electric heat	$N(808.7, 19.5)$	Cluster 8
Cluster 8	Electric heat	$N(799.7, 19.0)$	Cluster 7
Cluster 9	Lighting	$N(112.8, 2.7)$	Cluster 10
Cluster 10	Lighting	$N(116.8, 1.3)$	Cluster 9

Table 5.6 Overlapping clusters in house 6 from the REDD.

Cluster number	Appliance type	Power consumption $\sim \boldsymbol{N(\mu,\sigma)}$ (Watts)	Overlaps with
Cluster 1	Unknown outlet	$N(96.6, 13.0)$	Cluster 2
Cluster 2	Lighting	$N(121.3, 8.8)$	Cluster 1
Cluster 3	Electric heat	$N(463.3, 18.4)$	Cluster 4
Cluster 4	Air conditioner	$N(416.5, 6.8)$	Cluster 3
Cluster 5	Bathroom gfi	$N(945.9, 3.4)$	Cluster 6, cluster 7
Cluster 6	Air conditioner	$N(985.2, 31.0)$	Cluster 5, cluster 7
Cluster 7	Air conditioner	$N(973.5, 31.2)$	Cluster 5, cluster 6

Each cluster in the above tables (Table 5.1 through Table 5.6) represents a single stable state in an appliance. The appliance itself may have more than one state of operation, e.g. a finite state machine (FSM), but only states where overlapping between appliances is possible were considered. In addition, each cluster comes from a specific appliance even if they have the same appliance type. For example, cluster 1 and cluster 2 in house 1 come from two independent ovens and so on. Some clusters are defined as outlets or unknown outlets where there is no more information from the REDD about the exact appliance plugged to those outlets. The real power consumption by each appliance was modeled using a normal distribution from the sub-metered data from the REDD.

To illustrate the positive impact of using the median filter, the following example is given for an air conditioner from house 6 from the REDD. Figure 5.3 shows a portion of the raw power signal of an air conditioner. Spikes in power signal appearing in Figure

5.3 occupy one sample only. After applying a 9-sample median filter, the power signal was filtered as shown in Figure 5.4. Applying normality tests presented in section 5.2 resulted in average P-values of $\overline{PV} = 0.0005$ for the signal in Figure 5.3 (which means non-normality fitting) and a $\overline{PV} = 0.1676$ for the signal in Figure 5.4 (which means a good normality fitting).

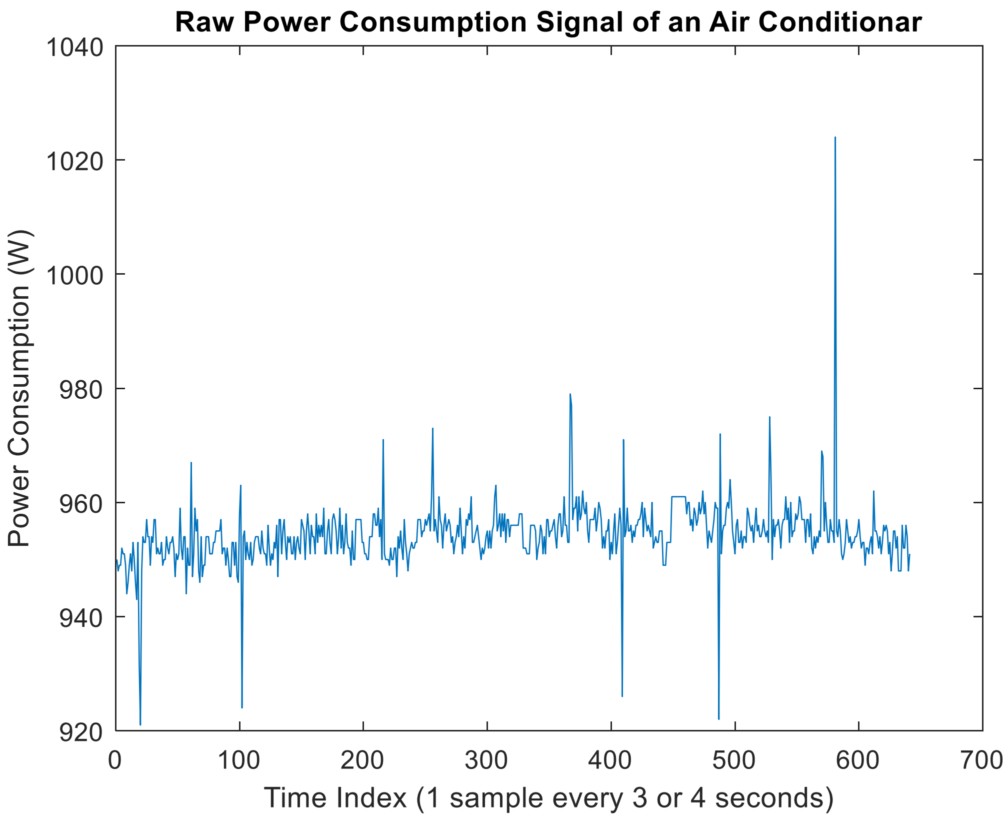

Fig. 5.3: Raw power consumption signal of an air conditioner.

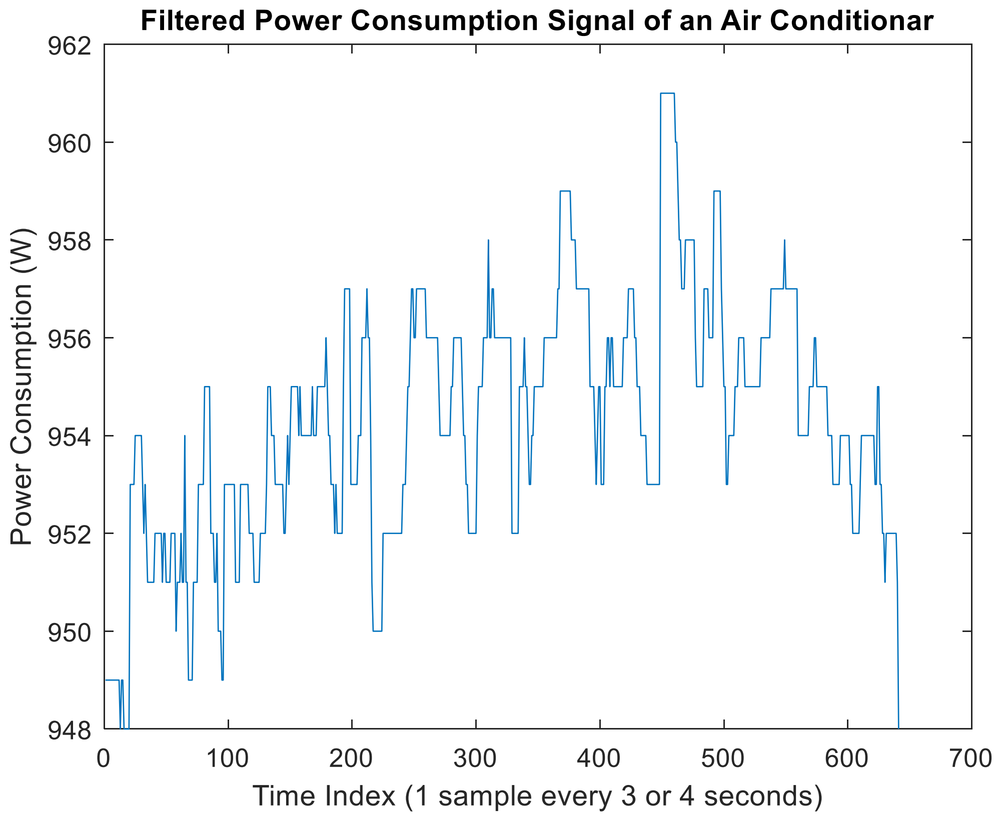

Fig. 5.4: Filtered power consumption signal of an air conditioner.

5.4.2 Splitting of Overlapping Appliances Clusters

The proposed approach in this Chapter was tested on each case of overlapping clusters. For the proposed splitting decision criteria, two confidence levels as $\alpha_1 = 0.01$ and $\alpha_2 = 0.10$ were applied to provide reasonable indications about clusters cohesion. Table 5.7 through Table 5.12 show the obtained splitting accuracies in terms of *Precision*, *Recall* and *F-measure* for all tested cases in house 1 through house 6, respectively. The *Precision*, *Recall* and *F-measure* accuracy metrics were used basically since the splitting process aims to assign items to clusters correctly. Therefore, this metric is suitable to assess the assignments accuracy as explained earlier in subsection 2.2.4.

Table 5.7: Accuracy of clusters splitting in house 1 from the REDD.

Cluster number	*Precision*	*Recall*	*F-measure*
Cluster 1	94.5%	79.1%	86.1%
Cluster 2	74.3%	92.9%	82.6%
Cluster 3 (versus cluster 4)	100%	72.2%	83.8%
Cluster 4	82.4%	100%	90.3%
Cluster 3 (versus cluster 5)	100%	51.1%	67.7%
Cluster 5	68.9%	100%	81.6%

Table 5.8: Accuracy of clusters splitting in house 2 from the REDD.

Cluster number	*Precision*	*Recall*	*F-measure*
Cluster 1	96.2%	59.8%	73.8%
Cluster 2	27.4%	86.4%	41.6%

Table 5.9: Accuracy of clusters splitting in house 3 from the REDD.

Cluster number	*Precision*	*Recall*	*F-measure*
Cluster 1 (versus cluster 2)	100%	95.7%	97.8%
Cluster 2 (versus cluster 1)	90.4%	100%	95.0%
Cluster 3 (versus cluster 4)	59.9%	66.9%	63.2%
Cluster 4 (versus cluster 3)	75.3%	69.2%	72.1%
Cluster 3 (versus cluster 5)	80.7%	66.9%	73.2%
Cluster 5 (versus cluster 3)	39.4%	57.4%	46.8%
Cluster 4 (versus cluster 5)	76.3%	29.8%	42.9%
Cluster 5 (versus cluster 4)	19.1%	64.2%	29.4%
Cluster 6 (versus cluster 1)	99.3%	100%	99.6%
Cluster 1 (versus cluster 6)	100%	96.8%	98.4%
Cluster 6 (versus cluster 2)	99.3%	100%	99.6%
Cluster 2 (versus cluster 6)	100%	96.8%	98.4%

Table 5.10: Accuracy of clusters splitting in house 4 from the REDD.

Cluster number	*Precision*	*Recall*	*F-measure*
Cluster 1	75.5%	100%	86.1%
Cluster 2 (versus cluster 1)	100%	75.9%	86.3%
Cluster 3	63.0%	99.6%	77.2%
Cluster 2 (versus cluster 3)	99.3%	45.6%	62.5%

Table 5.11: Accuracy of clusters splitting in house 5 from the REDD.

Cluster number	*Precision*	*Recall*	*F-measure*
Cluster 1	80.3%	57.5%	67.0%
Cluster 2	54.8%	78.5%	64.5%
Cluster 3	98.6%	23.3%	37.7%
Cluster 4	14.6%	97.6%	25.4%
Cluster 5	99.5%	70.9%	82.8%
Cluster 6	66.4%	99.4%	79.6%
Cluster 7	60.6%	54.7%	57.5%
Cluster 8	58.6%	64.2%	61.3%
Cluster 9	3.4%	89.4%	6.5%
Cluster 10	99.6%	49.4%	66.0%

Table 5.12: Accuracy of clusters splitting in house 6 from the REDD.

Cluster number	*Precision*	*Recall*	*F-measure*
Cluster 1	27.2%	99.9%	42.7%
Cluster 2	100%	38.4%	55.5%
Cluster 3	89.7%	95.1%	92.3%
Cluster 4	99.3%	98.5%	98.9%
Cluster 5 (versus cluster 6)	8.6%	100%	15.8%
Cluster 6 (versus cluster 5)	100%	48.9%	65.7%
Cluster 5 (versus cluster 7)	8.4%	100%	15.5%
Cluster 7 (versus cluster 5)	100%	47.8%	64.7%
Cluster 6 (versus cluster 7)	59.1%	53.3%	56.0%
Cluster 7 (versus cluster 6)	57.5%	63.1%	60.1%

It is important to note that although some clusters overlap with more than one cluster, the proposed splitting approach was applied separately on each pair of overlapping clusters. For illustration about the placement of devices clusters, an example of devices clusters in house 1 is shown in Figure 5.5 which provides indications about the degree of overlapping between clusters. Moreover, it shows whether a device cluster is tight or loose by figuring out the spread their items. For example, cluster 4 and cluster 5 are tight but cluster 3 is loose. Besides, cluster 3 and cluster 4 overlap in a quite small degree but cluster 5 totally overlaps with cluster 3.

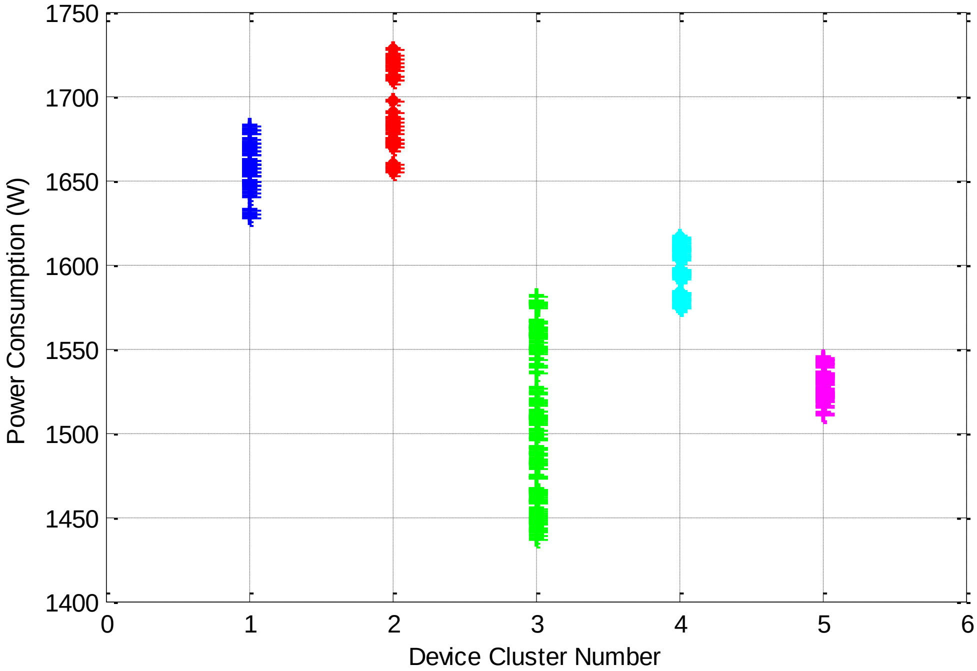

Fig. 5.5: Devices clusters in house 1 from the REDD.

As illustration about cohesion tests and $\overline{PV}$'s of the sub-clusters as explained in this Chapter, the case of cluster 3 and cluster 4 in house 1 is presented. The normality tests on the merged cluster showed a $\overline{PV}$ approximately 0.00, which concluded a splitting decision. After performing the cluster splitting, the normality fitting of the sub-cluster 3 was found to be better than the big cluster with $\overline{PV} = 0.033$. The $\overline{PV}$ of sub-cluster 4 has not changed significantly above that for the big merged cluster. Once either sub-clusters show better fitting for normality than the big merged cluster, it indicates that sub-clusters are likely be originating from two different sources as articulated earlier.

5.4.3 Interpretation and Discussion of Clusters Splitting Results

The results in the previous subsection showed that the accuracy of retrieving the split devices clusters from the merged one, in general, depends on three essential factors

- The degree of overlapping between the clusters of these devices.
- The tightness or looseness (i.e. the spread) of each overlapping cluster.
- The total number of items in each cluster.

The degree of overlapping can be concluded from the distance between the centers of two overlapping clusters. The standard deviation indicates whether a cluster is tight (items are close to each other) or loose (items spread over wide ranges).

Results shown in Figures 5.3 and 5.4 show the positive impact of applying the median filter on appliances power signal. It can be concluded that applying the median filter to remove outliers, spikes and noisy signals would help in presenting filtered power signals that show good normality fitting's. As methods proposed in Chapter 5 depends on the assumption of normality fitting of appliances consumptions, the median filter was applied to power signals of home appliances.

Two confidence levels where $\alpha_1 = 0.01$ and $\alpha_2 = 0.10$ were used to determine the splitting decisions as proposed in this Chapter by providing reasonable indications about clusters cohesion. The value of α_1 was chosen as small as 0.01 so as to decide a splitting once normality or cohesion of a cluster is deduced to be very poor. Besides, if a $\overline{PV}$ falls in the range between α_1 and α_2, the uncertainty case is dealt with as explained earlier in this Chapter. The value of α_2 should not be taken too higher than α_1 because this leads to become more prone to commit an error regarding the normality checking of the tested device cluster.

As shown in Table 5.7 for splitting accuracy results of house 1, high accuracies were obtained for device cluster 4 and lower accuracies for device cluster 3 (tested versus cluster 5). Cluster 4 showed a high accuracy as its center is quite far from cluster 3 and its items are close to each other (a tight cluster) as seen in Figure 5.5. Conversely, cluster 3 (tested versus cluster 5) showed a low accuracy mainly because cluster 5 is entirely overlapping with it as seen in Figure 5.5. This implies that several items which originally belong to cluster 3 will mostly be assigned to cluster 5. There is an interesting observation that if it happens to have a 100% in *Precision* for one cluster, the other overlapping cluster should have 100% in *Recall*. These results are logical as a cluster retrieved with 100% in *Precision* means all items assigned to it do belong to this cluster. That is, there is a zero element under the category of False Positive (*FP*) decisions. On contrary, a zero element in *FP* category for one cluster means a zero element in False Negative (*FN*) category for the other cluster which implies a 100% in *Recall*. To articulate this fact, we explain more about such cases in house 1 but the same justification applies to all other houses. In house 1, cluster 3 kept a 100% in *Precision* while clusters 4 and 5 kept a 100% in *Recall*. From Figure 5.5, it is observable that after performing the splitting, all assigned items to cluster 3 will most likely do belong to cluster 3 in realty due to its notable looseness and tightness of the other overlapping clusters. However, several items from cluster 3 could be assigned to cluster 4 or cluster 5. This fact interprets the low *Precision* for both cluster 4 and cluster 5. On the other hand, since cluster 4 and cluster 5 are notably tight, it is more likely that after splitting, all items those really belong to these clusters would be assigned to the corresponding cluster. This explains the 100% in *Recall* in the splitting accuracy results as there are no items missed from the assignment, i.e. zero items in the *FN* category. Such cases have repeatedly happened in the other houses for similar reasons.

In case when both overlapping devices clusters are tight and there is a noticeable distance between their centers, it is more likely to obtain high accuracies in retrieving both clusters. Cluster 1 and cluster 2 in house 3 are a good example showing this case. The converse example can be seen from the case of cluster 4 and cluster 5 in house 3, which results in low splitting accuracies. Some cases when two clusters centers are very close to each other, the obtained accuracies are very low as seen in the case of cluster 9 in house 5. Besides, clusters with limited number of items may show lower accuracy results as in the case of cluster 4 and cluster 9 in house 5. These clusters have limited number of items since the corresponding appliances were operating for less time durations than other appliances. Therefore, a cluster with fewer items is more likely to be less representative for the actual power consumption model of that specific device. In addition, an error in assigning a small number of items in small-sized clusters would result in noticeable reductions in the obtained accuracies.

An interesting case appears when the two overlapping clusters represent two identical appliances in the same house. From the REDD data, it was found that cluster 6 and cluster 7 in house 6 represent two identical air conditioners. However, each appliance showed a slightly different amount of fluctuation in power consumption which could be due to uneven noise effects. In such cases, it is reasonable to obtain *Precisions* and *Recalls* around 50% for both clusters. This can be understood as the splitting approach resulted in somehow two similar small clusters. Each of these small clusters consists of approximately a half of its original items and a half from the other cluster items. Therefore, the number of items under the categories of *TP*, *FP* and *FN* are approximately close to each other. This conclusion articulates the obtained accuracies around 50%.

For application of the developed clusters splitting method in a real-world scenario, it is proposed to be applied as a subsequent step to a *clustering-based* NILM approach taking into account all other considerations. In brief, a real-world scenario would go through the following steps

1. First of all, a *clustering-based* NILM approach should be applied. The results are supposed to be a number of distinct clusters that represent individual appliances.
2. The developed clusters splitting method is proposed to be applied to check the cohesion of appliances clusters that were obtained in step 1. The developed method should be applied to clusters where the normality fitting assumption is reasonable

(this means to avoid varying appliances and cases of poor/non-normality as presented previously in section 5.1).

3. In cases where a splitting of a cluster is decided, the two resulting clusters should be labeled to two different appliances (even though being close in power consumption). This step is done basically manually and may need some experience/knowledge.

From the above steps, the proposed clusters splitting methods can be applied to clusters that are labeled or matched with appliances and their individual consumption can be modeled by normal distribution (i.e. excluding continuously varying appliances and those with noticeable overshoot periods). Labeling appliances clusters with their corresponding appliances is usually done based on the used learning method as follows

- In the supervised learning methods, models/patterns of individual appliances are available and thus constructed clusters are easily labeled/matched with the corresponding appliances during the learning phase.

- In the unsupervised learning methods, models of appliances are built using some probabilistic rules. The resulting clusters/models are not labeled with the corresponding appliances. Hence, there must be a manual step to label/match clusters with respective appliances (this usually requires experience/knowledge about home appliances and their consumptions). The proposed clusters splitting method can be applied after having the clusters labeled with their respective appliances.

5.4.4 Contributions to Load Disaggregation Accuracy

To the best of the author's knowledge, there is no cohesion testing method that has been applied to the same REDD data set for the purpose of dealing with overlapping consumptions of the devices. For this reason, we were unable to compare results to benchmark reference methods. Clusters cohesion tests in the literature were applied in different fields but the nature of the problem (i.e. measurable features, their dimensions, fitting distributions, etc.) are basically dependent on the case under study.

The proposed approach, to deal with specific cases of overlapping clusters, is suggested to work together with NILM approaches as a subsequent step to tackle detected cases of overlapping devices clusters. The NILM method should first disaggregate individual devices and then the proposed approach refines the findings by tackling the issues of overlapped devices. Hence, it contributes to the improvement of the overall load disaggregation accuracy.

5.5 Remarks on the Proposed Approach in this Chapter

The proposed approach aimed to disaggregate overlapping home appliances using a method for clusters splitting. The approach applied a cohesion test that was based on three tests for normality fitting. Some remarks were noticed in the proposed approach, which can be summarized as follows

1. Continuously varying appliances and some FSMs showed poor normality fitting of their power consumptions. Therefore, the proposed approach becomes inapplicable to these appliances as the cohesion test is basically dependent on the normality assumption of appliances power consumptions.
2. The performance of items assignments after clusters splitting is affected by common issues such as the degree of overlapping between appliances clusters, whether the clusters are tight or loose and the total number of items in the appliances clusters.

5.6 Summary

This Chapter presented the third proposed approach which aims to split overlapping clusters of appliances, which result when applying a clustering method to load disaggregation. The splitting of overlapped clusters is basically taken when poor degrees of clusters cohesion are detected. Checking clusters cohesion was carried out by investigating the normality fitting of appliances clusters using three common normality tests. The proposed methods were applied on overlapping cases of appliances from six real houses from the REDD public data set.

Chapter 6

Conclusions and Potential Future Work

In this work, three enhanced approaches were proposed and tested aiming to improve the overall performance of non-intrusive load disaggregation. This chapter presents significant conclusions and suggests potential future work.

6.1 Conclusions

In this thesis, three enhanced approaches were proposed to tackle three challenging problems in the field of non-intrusive load disaggregation which aims to decompose the total aggregate measurements of a household power consumption into individual appliances power consumptions. The motivation of this work was augmented as it contributes to achieving energy saving in the residential sector. The problem of non-intrusive load disaggregation was first defined, and main challenges were highlighted. Characteristics of promising solutions were explained specially regarding the use of low frequency measurements and applying unsupervised learning approaches that do not require sub-metering of power consumptions at appliance-level. Literature review and previous works were investigated and presented in detail about used appliances features and adopted approaches. In addition, basic limitations in related studies were discussed. The foundations of the proposed approaches such as HMMs and their extensions and clustering methods were presented in brief.

The first proposed approach targeted improving the disaggregation accuracy of individual home appliances. The feature of mutual appliances interactions (two-way interactions) was estimated and embedded in a FHMM representation of the aggregate signal and individual home appliances. Three-way and higher order interactions were neglected in accordance with the principle of sparsity. Interactions between appliances are usually caused by poor power quality issues in the design of appliances or in electricity network components. In addition, an adaptive estimations approach was introduced to run during disaggregation process to update appliances models and their interactions (if possible). The adaptive estimations approach was applied on cases of simultaneously four or less ON appliances. The proposed approaches were tested on house 2 from the REDD public data set. Models were first built, and possible interactions were observed from a portion of aggregate total signal. Then, the disaggregation process was performed on the remaining portion. It was found that appliances with known two-way interactions were disaggregated with higher accuracies than those with unknown mutual interactions. Besides, the adaptive estimations of main power effects and two-way interactions contributed in enhancing the disaggregation results. The proposed approaches contributed to enhance the overall disaggregation accuracies of individual appliances between 0.81% to 10.11% above the disaggregation accuracies obtained by the standard FHMM methods.

The second enhanced approach aimed to disaggregate continuously varying home loads. A quantized CS-HMM was proposed to deal with continuously varying loads by applying a modified method to estimate the transition matrix. The proposed estimation method of the transitions matrix aims to mitigate two possible extreme cases in the learning phase which are cases of never-occurred transitions and too frequent transitions. The proposed methods were tested by applying to a set of generated data and real data from the REDD public data set. Thereafter, the CS-HMM was combined in the structure of the FHMM to produce a hybrid CS/DS-FHMM. In the hybrid CS/DS-FHMM, a framework for learning and estimation was carried out using the total aggregate signal of real data of four common home appliances. The proposed framework of the hybrid CS/DS-FHMM showed enhanced performance in terms of accuracy of estimated power consumption when compared to classical methods. It was found that increasing the number of quantization levels can improve the disaggregation accuracy, for a reasonable cost of complexity and extra execution time. The proposed methods could enhance the disaggregation accuracy for the varying load (in presence of other discrete loads) almost 1% above a reference common method.

The third enhanced approach aimed to split overlapping appliances clusters based on a method that investigates the degree of clusters cohesion. The cases where appliances power consumptions are expected to show good normality fitting were considered. The proposed approach analyzes the cohesion of appliances clusters to determine if a cluster should be split into two small clusters. The analysis of clusters cohesion was based on the collective outcome of three normality tests performed against two confidence levels. Thereafter, the splitting is carried out using an inner expectation maximization (EM) method. The proposed approach was tested overlapping appliances clusters from six real houses from the REDD public data set. The accuracy of splitting, i.e. assignment of elements to the small clusters, was found to be depending mainly on the degree of overlapping, whether these clusters are tight or loose and total number of items in each cluster. The proposed approach is suggested to be applied to the energy disaggregation problem to deal with situations where the power consumptions of individual devices may overlap. The NILM clustering-based method is suggested to disaggregate individual devices first and then the proposed approach may refine the findings by disaggregating the overlapped devices in power consumptions. Thereby, the proposed splitting approach contributes to the overall performance of the NILM approaches. With few cases of exceptions, it was found that the sub-clusters of individual appliances were retrieved with accuracies above 60% using the inner EM to assign items from the overlapped cluster to the respective sub-cluster.

6.2 Potential Future Work

For the future research work, it is suggested to extend the proposed approaches by incorporating added techniques or framework layers that could address some of the existing inadequacies.

The first proposed approach aimed to enhance the overall load disaggregation accuracy by embedding information on appliances mutual interaction. It is suggested to extend the presented techniques by including information on three-way or higher order interactions between home appliances as they may be significant in some applied case studies or real houses. This information could help in providing improved modeling and disaggregation of appliances when they operate simultaneously and in various combinations. Information on three-way or higher order interactions could be captured from the aggregate total signal by analyzing longer sequences of aggregate signal till it becomes possible that possible cases of combinations of appliances are observed in operation. In addition, it is suggested to expand the adaptive estimations approach to consider cases when there are five or more concurrently ON appliance. Higher orders of fractional factorial designs could be adopted to update appliances models when there are five or more ON appliances at the same time.

The second proposed approach aimed to model and disaggregate continuously varying home loads. It is suggested to extend the learning and estimation methods to be able to disaggregate more than one varying load (instead of lumped loads modeling). It is suggested to detect which of varying loads are actually operating by extracting and incorporating possible distinct features or performing deep investigations on their power consumptions patterns. For example, their pattern of fluctuations, nature or shape of associated noise signals, mutual interactions with other appliances, duration of use, could be some of useful features to distinguish between continuously varying loads.

The third proposed approach aimed to refine the outcomes of a clustering method applied to NILM by checking which appliances clusters should be split according to its degree of cohesiveness. It is suggested to extend the proposed cohesion test by including additional normality tests so as to investigate more aspects of normality fitting. In addition, it is suggested to develop similar methods to test adequacy of appliances clusters in cases of appliances that fit to non-normal distributions. Finally, it is suggested to re-apply the proposed approach on the resulting clusters after splitting, to test if there should be another reasonable splitting. This further cohesion testing could be helpful in cases when there are three or more overlapping clusters of appliances.

In general, future work on NILM should be directed toward large-scale applications that capable of showing enhanced applicability and scalability of solutions to the problem of non-intrusive load disaggregation. In addition, it is suggested to search for extra significant features for home appliances that may be extractable from the total aggregate signal. Besides, modern appliances that embed technologies of self-recording of their consumptions, should be used (when existing at a household) to assist in the overall framework of load disaggregation.

In some countries, gas and water smart meters began to be deployed alongside with electricity smart meters, an interesting research problem is to concurrently disaggregate all the three utilities. Similar to electricity disaggregation, a recent research work was applied to gas disaggregation [Cohn *et al.*, 2010] and another was applied to water disaggregation [Dong *et al.*, 2013]. An interesting recent study [Vitter and Webber, 2018] utilized the disaggregation of circuit-level electricity data to improve performance by a water end-use disaggregation tool. However, since some appliances consume two utilities (e.g. washing machine requires both electricity and water), information derived from one utility could also be used to infer the usage of another utility. Therefore, new techniques are required to consolidate the disaggregation of multiple utilities while allowing mutual information to be shared between each utility [Parson, 2014].

Author's Publications

Parts of this research appear in the following publications

Journal papers

[1] M. Aiad and P. H. Lee, "Unsupervised approach for load disaggregation with devices interactions", *Energy and Buildings*, vol. 116, pp. 96-103, 2016.

[2] M. Aiad and P.H. Lee, "Non-intrusive load disaggregation with adaptive estimations of devices main power effects and two-way interactions", *Energy and Buildings*, vol. 130, pp. 131-139, 2016.

[3] M. Aiad and P.H. Lee, "Energy Disaggregation of Overlapping Home Appliances Consumptions Using a Cluster Splitting Approach", *Sustainable Cities and Society*, vol. 43, pp. 487-494, 2018.

[4] M. Aiad and P.H. Lee, "Hybrid Continuous/Discrete States Factorial Hidden Markov Models with Applications to Non-Intrusive Load Disaggregation", Submitted to *Sustainable Cities and Society*, 2018.

Conference papers

[1] M. Aiad and P.H. Lee, "Non-intrusive monitoring of overlapping home appliances using smart meter measurements", *IEEE Power and Energy Conference at Illinois (PECI)*, Champaign, IL, USA, February 2017.

[2] M. Aiad and P.H. Lee, "Modeling and Power Estimation of Continuously Varying Residential Loads Using a Quantized Continuous-State Hidden Markov Model", *Asia Conference on Machine Learning and Computing (ACMLC)*, Singapore, December 2017.

References

(Alphabetical order)

[Ainsleigh, 2001] P.L. Ainsleigh, "Theory of Continuous-State Hidden Markov Models and Hidden Gauss-Marcov Models", *Naval Undersea Warfare Center Division Newport*, Rhode Island, USA, 2001.

[Akbar and Khan, 2007] M. Akbar and Z. A. Khan, "Modified Nonintrusive Appliance Load Monitoring for Nonlinear Devices", *IEEE International Multitopic Conference*, pp. 1-5, 2007.

[Albicki and Cole, 1998] A. Albicki and A. Cole, "Algorithm for Non-Intrusive Identification of Residential appliances," *Circuits and Systems, Proceedings of the IEEE International Symposium on*, vol. 3, pp. 338-341, 1998.

[Anderson and Darling, 1954] T.W. Anderson and D.A. Darling, "A test of goodness of fit", *Journal of American Statistical Association*, vol. 49, no. 268, pp. 765-769, 1954.

[Anderson *et al*., 2012] K. Anderson, A. Ocneanu, D. Benitez, D. Carlson, A. Rowe and M. Berges, "BLUED: a fully labeled public dataset for event-based nonintrusive load monitoring research", *Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, Beijing, China, August 2012.

[Baranski and Voss, 2003] M. Baranski and J. Voss, "Non-Intrusive Appliance Load Monitoring Based on an Optical Sensor", *IEEE Power Tech Conference*, Bologna, 2003.

[Baranski and Voss, 2004] M. Baranski and J. Voss, "Genetic Algorithm for Pattern Detection in NIALM Systems", *Proceedings of IEEE International Conference on Systems Man and Cybernetics*, Hague, The Netherlands, vol. 4, pp. 3462–3468, October 2004.

[Barker *et al.*, 2012] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy and J. Albrecht, "Smart*: An open data set and tools for enabling research in sustainable homes", *SustKDD,* August 2012.

[Barker, 2014] S. Barker, "Model-driven analytics of energy meter data in smart homes", PhD thesis, *University of Massachusetts Amherst*, USA, June 2014.

[Beal *et al.*, 2002] M.J. Beal, Z. Ghahramani and C.E. Rasmussen, "The infinite hidden Markov model", *Advances in neural information processing systems*, pp. 577-584, 2002.

[Benaglia *et al.*, 2010] T. Benaglia T, D. Chauveau, D.R. Hunter DR and D.S. Young, "mixtools: An R Package for Analyzing Mixture Models", *Journal of Statistical Software*, vol. 32, no. 6, pp. 1-29, 2010.

[Bhattacharyya *et al.*, 2007] S. Bhattacharyya, J.M.A. Myrzik and W.L. Kling, "Consequences of poor power quality – an overview", *Universities Power Engineering Conference UPEC*, pp. 651-656, 2007.

[Bishop, 2006] C. Bishop, "Pattern Recognition and Machine Learning", *Springer-Verlag New York, Inc*, Secaucus, NJ, USA, 2006.

[Bollen, 2011] M.H.J. Bollen, "The smart grid: adapting the power system to new challenges", *Morgan & Claypool Publishers Series: Synthesis Lectures on Power Electronics,* 2011.

[Bonfigli *et al.*, 2015] R. Bonfigli, S. Squartini, M. Fagiani and F. Piazza, "Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview", *IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC),* pp. 1175-1180, 2015.

[BP P.L.C., 2017] BP P.L.C., "BP Statistical Review of World Energy", June 2017.

[BP P.L.C., 2018] BP P.L.C., “BP Energy Outlook”, June 2018.

[Chang, 2012] H.H. Chang, “Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses”, *Energies*, vol. 5, pp. 4569-4589, 2012.

[Chou and Chang, 2013] P. Chou and R. Chang, “Unsupervised adaptive non-intrusive load Monitoring system”, *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 3180-3185, 2013.

[Cohn *et al.*, 2010] G. Cohn, S. Gupta, J. Froehlich, E. Larson and S. Patel, “GasSense: Appliance-Level, Single-Point Sensing of Gas Activity in the Home”, P. Floren, A. Krger and M. Spasojevic, editors, *Pervasive Computing*, vol. 6030 of Lecture Notes in Computer Science, Springer Berlin/ Heidelberg, pp. 265-282, 2010.

[Ding and He, 2002] C. Ding and X. He, “Cluster merging and splitting in hierarchical clustering algorithms”, *Proceedings of the IEEE International Conference on Data Mining*, pp. 139-146, 2002.

[Dong *et al.*, 2013] H. Dong, B. Wang and C.T. Lu, “Deep Sparse Coding based Recursive Disaggregation Model for Water Conservation”, *International Joint Conference on Artificial Intelligence*, Beijing, China, 2013.

[Farinaccio and Zmeureanu, 1999] L. Farinaccio and R. Zmeureanu, “Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses”, *Energy and Buildings*, vol. 30, pp. 245-259, 1999.

[Ghahramani and Jordan, 1997] Z. Ghahramani and M. Jordan, “Factorial Hidden Markov Models”, *Machine Learning*, vol. 27, pp. 245-273, 1997.

[Gil-de-Castro *et al.*, 2014a] A. Gil-de-Castro, S.K. Rönnberg and M.H.J. Bollen, "A study about harmonic interaction between devices", *IEEE 16th International Conference on Harmonics and Quality of Power (ICHQP)*, pp. 728-732, 2014.

[Gil-de-Castro *et al.*, 2014b] A. Gil-de-Castro, S.K. Rönnberg, M.H.J. Bollen and A. Moreno-Muñoz, "Study on harmonic emission of domestic equipment combined with different types of lighting", *Electrical Power and Energy Systems,* vol. 55, pp. 116-127, 2014.

[Goncalves *et al.*, 2011] H. Goncalves, A. Ocneanu, M. Berg´es and R.H. Fan, "Unsupervised Disaggregation of Appliances Using Aggregated Consumption Data", *Proceedings of KDD Workshop on Data Mining Applications for Sustainability*, San Diego, CA, USA, pp. 21–24 August 2011.

[Gupta *et al.*, 2010] S. Gupta, M. S. Reynolds, S. N. Patel, "ElectriSense: Single-Point Sensing Using EMI for Electrical Event Detection and Classification in the Home", *Conference on Ubiquitous Computing*, pp. 139-148, 2010.

[Hart, 1992] G. Hart, "Nonintrusive appliance load monitoring", *Proceedings of the IEEE*, vol. 12, no. 80, pp. 1870-1891, 1992.

[Jain *et al.*, 1999] A.K. Jain, M.N. Murty and P.J. Flynn, "Data Clustering: A Review", *ACM Computing Surveys*, vol. 31, no. 3, September 1999.

[Henao *et al.*, 2018] N. Henao, K. Agbossou, S. Kelouwani, S.S. Hosseini and M. Fournier, "Power Estimation of Multiple Two-State Loads Using a Probabilistic Non-Intrusive Approach", *Energies*, vol. 11, no. 1, 2018.

[Jarque and Bera, 1987] C.M. Jarque and A.K. Bera, "A Test for Normality of Observations and Regression Residuals", *International Statistical Review*, vol. 55, no. 2, pp. 163–172, 1987.

[Johnson and Willsky, 2013] M.J. Johnson and A.S. Willsky, “Bayesian nonparametric hidden semi-Markov models”, *Journal of Machine Learning Research*, vol. 14, pp. 673-701, 2013.

[Kamoto *et al*., 2017] K.M. Kamoto, Q. Liu and X. Liu, “Unsupervised Energy Disaggregation of Home Appliances”, *International Conference on Cloud Computing and Security*, pp. 398-409, Springer, Cham, 2017.

[Kato *et al*., 2009] T. Kato, H.S. Cho and D. Lee, “Appliance Recognition from Electric Current Signals for Information-Energy Integrated Network in Home Environments”, *Proceedings of the 7th International Conference on Smart Homes and Health Telematics*, Tours, France, vol. 5597, pp. 150–157, July 2009.

[Kim *et al*., 2010] H. Kim, M. Marwah, M. Arlitt, G. Lyon and J. Han, “Unsupervised disaggregation of low frequency power measurements”, *SIAM International Conference on Data Mining (SDM)*, pp. 747-758, 2010.

[Kolter and Jaakkola, 2010] J.Z. Kolter and T. Jaakkola, “Approximate inference in additive factorial HMMs with application to energy disaggregation”, *Journal of Machine Learning Research*, vol. 22, pp. 1472–1482, 2012.

[Kolter and Johnson, 2011] J. Kolter and M. Johnson, “REDD: A public data set for energy disaggregation research”, *SustKDD*, San Diego, CA, USA, 2011.

[Lai *et al*., 2012] Y.X. Lai, C.F. Lai, Y.M. Huang and H.C. Chao, “Multi-appliance recognition system with hybrid SVM/GMM classifier in ubiquitous smart home”, *Information Sciences*, vol. 230, pp. 39-55, 2013.

[Lam *et al*., 2007] H.Y. Lam, G.S.K. Fung, W.K. Lee, “A Novel method to construct taxonomy electrical appliances based on load signatures”, *IEEE Transactions on Consumer Electron*ics, vol. 53, pp. 653–660, 2007.

[Lee *et al.*, 2004] W.K. Lee, G.S.K Fung, H.Y. Lam, F.H.Y. Chan and M. Lucente, “Exploration on Load Signatures”, *Proceedings of International Conference on Electrical Engineering (ICEE),* Sapporo, Japan, pp. 1-5, July 2004.

[Lee *et al*., 2005] K. D. Lee, S. B. Leeb, L. K. Norford, P. R. Armstrong, J. Holloway and S. R. Shaw, “Estimation of Variable-Speed-Drive Power Consumption from Harmonic Content”, *IEEE Transactions on Energy Conversion*, vol. 20, pp. 566-574, 2005.

[Liang *et al.*, 2010] J. Liang, S.K.K. Ng, G. Kendall and J.W.M. Cheng, “Load signature study Part I: Basic concept, structure, and methodology”, *IEEE Transactions on Power Delivery*, vol. 25, pp. 551–560, 2010.

[Lin *et al.*, 2010] G.Y. Lin, S.C. Lee, Y.J. Hsu and W.R. Jih, “Applying Power Meters for Appliance Recognition on the Electric Panel”, *Proceedings of the 5th IEEE Conference on Industrial Electronics and Applications*, Melbourne, Australia, pp. 2254–2259, June 2010.

[Lin *et al.*, 2011] Y.H. Lin, M.S. Tsai and C.S. Chen, “Applications of fuzzy classification with fuzzy c-means clustering and optimization strategies for load identification in NILM systems”, *IEEE International Conference on Fuzzy Systems (FUZZ),* pp. 859-866, June 2011.

[Lloyd, 1982] S. Lloyd, “Least squares quantization in PCM”, *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129-137, 1982.

[Lookabaugh and Gray, 1989] T.D. Lookabaugh and R.M. Gray, “High-resolution quantization theory and the vector quantizer advantage”, *IEEE Transactions on Information Theory*, vol. 35, no. 5, pp. 1020-1033, 1989.

[Lu and Leen, 2007] Z. Lu and T.K. Leen, “Penalized probabilistic clustering”, *Neural Computation*, vol. 19, no. 6, pp. 1528-1567, 2007.

[Lu *et al*., 2017] T. Lu, Z. Xu and B. Huang B, “An Event-Based Nonintrusive Load Monitoring Approach: Using the Simplified Viterbi Algorithm”, *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 54-61, 2017.

[Maasoumy *et al*., 2014] M. Maasoumy, B. Sanandaji, K. Poolla, and A. S. Vincentelli, “BERDS - berkeley energy disaggregation data set”, *Proceedings of the Workshop on Big Learning at the Conference on Neural Information Processing Systems (NIPS 2013)*, 2014.

[Makonin *et al*., 2013] S. Makonin, F. Popowich, L. Bartram, B. Gill and I. V. Bajic, “AMPds: A public dataset for load disaggregation and eco-feedback research”, *Proceedings of the 2013 IEEE Electrical Power and Energy Conference (EPEC)*, 2013.

[Marsaglia *et al*., 2003] G. Marsaglia, W. Tsang and J. Wang, “Evaluating Kolmogorov's Distribution”, *Journal of Statistical Software*, vol. 8, no. 18, pp. 1-4. 2003.

[Massey, 1951] F.J. Massey FJ, “The Kolmogorov-Smirnov Test for Goodness of Fit”, *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[Montgomery and Runger, 2011] D. Montgomery and G. Runger, “Applied statistics and probability for engineers”, fifth edition, *John Wiley and Sons, Inc*, 2011.

[Montgomery, 2001] D. Montgomery, “Design and analysis of experiments”, fifth edition, *John Wiley and Sons, Inc*, 2001.

[Murphy, 2002] K.P. Murphy, “Dynamic Bayesian Networks: Representation, Inference and Learning”, Ph.D. dissertation, University of California, Berkeley, USA.

[Orbanz and Teh, 2011] P. Orbanz and Y.W. Teh, "Bayesian nonparametric models", *Encyclopedia of Machine Learning*, Springer US, pp. 81-89, 2011.

[Parson *et al.*, 2014] O. Parson, S. Ghosh, M. Weal and A. Roger, "An unsupervised training method for non-intrusive appliance load monitoring", *Artificial Intelligence*, vol. 217, pp. 1-19, 2014.

[Parson, 2014] O. Parson, "Unsupervised training methods for non-intrusive appliance load monitoring from smart meter data", PhD thesis, *University of Southampton*, UK, April 2014.

[Patel *et al.*, 2007] S. N. Patel, T. Robertson, J.A. Kientz, M.S. Reynolds and G.D. Abowd, "At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line", *Conference on Ubiquitous Computing*, pp. 271-288, 2007.

[Pavas *et al.*, 2012] A. Pavas, H. Torres-S and V. Staudt, "Statistical analysis of power quality disturbances propagation by means of the method of disturbances interaction", *Innovative Smart Grid Technologies (ISGT Europe), 2012 3rd IEEE PES International Conference and Exhibition on*, pp. 1-9, October 2012.

[Powers, 2011] D.M. Powers, "Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation", *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37-63, 2011.

[Preston *et al.*, 2010] D. Preston, C. Brodley, R. Khardon, D. Sulla-Menashe and M Friedl, "Redefining class definitions using constraint-based clustering: an application to remote sensing of the earth's surface", *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 823-832, 2010.

[Razali and Wah, 2011] N.M. Razali and Y.B. Wah, “Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests”, *Journal of statistical modeling and analytics*, vol. 2, no. 1, pp. 21-33, 2011.

[Ruzzelli *et al.*, 2010] A.G. Ruzzelli, C. Nicolas, A. Schoofs and G.M.P. O’Hare, “Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor”, *Proceedings of the 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, Boston, MA, USA, pp. 1–9, June 2010.

[Shao *et al.*, 2012] H. Shao, M. Marwah and N.A. Ramakrishnan, “Temporal Motif Mining Approach to Unsupervised Energy Disaggregation”, *Proceedings of the 1st International Workshop on Non-Intrusive Load Monitoring*, Pittsburgh, PA, USA, May 2012.

[Srinivasan *et al.*, 2006] D. Srinivasan, W. Ng and A. Liew, “Neural-network-based signature recognition for harmonic source identification”, *IEEE Transactions on Power Delivery*, vol. 21, pp. 398–405, 2006.

[Suzuki *et al.*, 2008] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura and K. Ito, “Nonintrusive Appliance Load Monitoring Based on Integer Programming”, *Proceedings of SICE Annual Conference*, Tokyo, Japan, vol. 174, pp. 2742–2747, August 2008.

[Teh *et al.*, 2005] Y.W. Teh, M.I. Jordan, M.J. Beal and D.M. Blei, “Sharing clusters among related groups: Hierarchical Dirichlet processes”, *Advances in neural information processing systems*, pp. 1385-1392, 2005.

[Teh, 2011] Y.W. Teh, “Dirichlet process”, *Encyclopedia of machine learning*, Springer US, pp. 280-287, 2011.

[Vitter and Webber, 2018] J.S. Vitter and M.E. Webber, “A non-intrusive approach for classifying residential water events using coincident electricity data”, *Environmental Modelling & Software*, vol. 100, pp. 302-313, 2018.

[Wagstaff *et al.*, 2001] K. Wagstaff, C. Cardie and S. Schroedl, “Constrained K-means clustering with background knowledge”, *Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 577-584, 2001.

[Wichakool *et al*., 2009] W. Wichakool, A.T. Avestruz, R. W. Cox and S. B. Leeb, “Modeling and Estimating Current Harmonics of Variable Electronic Loads”, *IEEE Transactions on Power Electronics*, vol. 24, pp. 2803-2811, 2009.

[Widrow *et al.*, 1996] B. Widrow, I. Kollar and M.C. Liu, “Statistical theory of quantization”, *IEEE Transactions on instrumentation and measurement*, vol. 45, no. 2, pp. 353-361, 1996.

[Xie *et al.*, 2017] J. Xie, T. Hong, T. Laing and C. Kang, “On normality assumption in residual simulation for probabilistic load forecasting”, *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1046-53, May 2017.

[Yu, 2010] S.Z. Yu, “Hidden semi-Markov models”, *Artificial intelligence*, vol. 174, no. 2, pp. 215-243, 2010.

[Zeifman and Roth, 2011] M. Zeifman and K. Roth, “Nonintrusive appliance load monitoring: review and outlook”, *IEEE Transactions on Consumer Electronics*, vol. 57, pp. 76-84, 2011.

[Zoha *et al.*, 2012] A. Zoha, A. Gluhak, M. Imran and S Rajasegara, “Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey”, *Sensors*, vol. 12, pp. 16838-16866, 2012.

[Zucchini *et al.*, 2016] W. Zucchini, I.L. MacDonald and R. Langrock, “Hidden Markov models for time series: an introduction using R”, vol. 150, *CRC press,* 2016.