

MINIMALLY SUPERVISED LOW-FREQUENCY METHODS FOR
MONITORING MAJOR ENERGY CONSUMERS IN BUILDINGS

by

Saman Mostafavi

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Electrical Engineering

Charlotte

2019

Approved by:

Dr. Robert W. Cox

Dr. Yogendra P. Kakad

Dr. Thomas P. Weldon

Dr. Erik Saule

ABSTRACT

SAMAN MOSTAFAVI. Minimally Supervised Low-Frequency Methods for Monitoring Major Energy Consumers in Buildings. (Under the direction of DR. ROBERT W. COX)

Buildings consume approximately 40% of global primary energy. Almost a fifth of this consumption arises from faulty or poorly operated systems. Monitoring the state of such systems can provide beneficial diagnostic insight in favor of more energy-efficient buildings. Non intrusive load monitoring (NILM) is the process of breaking down the aggregate power consumption of a building into its individual constituents. NILM can be performed using smart meter data to provide more granular information on building appliances. It is crucial that the algorithms involved in this task be unsupervised since scaling such a process to millions of houses with human involvement is near impossible. It is also vital that the algorithms are compatible with the current, i.e., smart meters, infrastructure if they are to be implemented at a large scale and market-friendly.

This thesis proposes a novel robust NILM algorithm capable of disaggregating the major loads for a portfolio of commercial and residential buildings from minutely recorded smart meter data. The key contribution of this thesis can be summarized in highlighting the major characteristics of power consumption in commercial and residential buildings, inspiring certain relaxing assumptions on smaller loads in favor of accurate state tracking for bigger ones. These assumptions address the seasonal and operational load variability in buildings and provide a framework for inference methods capable of producing good approximations for a computationally intractable problem.

Finally, an empirical evaluation of the proposed NILM against several other studies in the field is performed. To this end, a unique dataset with labeled appliance-level data is created using data collected from several commercial buildings. The evaluation

is presented for this dataset along with several other publicly available residential datasets. The results indicate an improvement in several evaluation metrics defined in this thesis.

ACKNOWLEDGEMENTS

As I close this chapter in my life, it is impossible not to feel sentimental about the journey I have gone through in the past five years. I count myself very lucky to have had the chance to be supervised by Dr. Robert Cox. I will forever be in his debt for the genius insights and the general directions he offered. He gave me a chance to move forward, and I remain hopeful that one day, I can provide a similar opportunity for somebody else.

I would like to thank my dissertation committee members Professor Yogendra Kakad, Dr. Thomas Weldon, and Dr. Erik Saule. I was lucky enough to be mentored by Professor Kakad in my masters. Since day one, his office door was always open to me, and he provided a certain level of wisdom transcending the realms of academia. I would also like to thank my colleague and friend, Dr. Benjamin Futrell, for keeping me company all these years, being a good listener, and offering the kind of advice that would jump-start a dead project when things were starting to look a bit stagnant.

This dissertation would not have been possible without the generous help and support offered by the Department of Electrical and Computer Engineering, the Graduate School, and the Center for Sustainably Integrated Buildings and Sites (SIBS) at UNC Charlotte. Most of my research and publications were supported by a grant from the National Science Foundation. I would also like to thank Wells Fargo, Bank of America, Ingersoll Rand, and Atrium Health, who are the members of SIBS.

Last but not least. I want to thank my dear family and friends here in Charlotte, North Carolina. My aunts, uncles, and many cousins, you all provided unconditional love and support to get me over the line. I can not even begin to say how grateful I am to have my uncle, Dr. Taghi Mostafavi, as a mentor in my life. Sina, you will always be a pain in the neck. Mom, Dad, I did it for you!

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xiii
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: BACKGROUND AND LITERATURE REVIEW	19
2.1. Building Energy Monitoring	19
2.2. Intrusive Monitoring	21
2.3. Non Intrusive Monitoring	23
2.3.1. High Frequency Methods	26
2.3.2. NILM with smart meter data	27
2.3.3. Public Datasets	28
2.4. Summary	32
CHAPTER 3: MOTIVATION	33
3.1. Setting the Standards for NILM	35
3.2. Hidden Markov Models	36
3.2.1. Factorial hidden Markov Model	39
3.2.2. Challenges of NILM with hidden Markov models	41
3.2.3. On the issue of supervision	45
3.3. Summary	47

CHAPTER 4: A ROBUST NILM ALGORITHM	48
4.1. Background filtering	49
4.1.1. Model based change of mean detection	50
4.1.2. Baseline removal	54
4.2. Edge analysis and load modeling	56
4.3. State inference and power estimation	60
4.3.1. Combination of Gaussian	64
4.4. Summary	68
CHAPTER 5: PERFORMANCE ANALYSIS ON SEVERAL PUBLIC AND PRIVATE DATASETS	69
5.1. Evaluation Metrics	69
5.2. Performance of NILM on Residential Datasets	72
5.2.1. Disaggregating REDD	72
5.2.2. Disaggregating AMPds	74
5.3. Disaggregating Commercial Buildings	76
5.3.1. Data Acquisition	79
5.3.2. Experimental results on the bank datasets	82
5.3.3. An example use case for HVAC monitoring using NILM	88
5.4. Summary	90
CHAPTER 6: SUMMARY, LIMITATIONS AND FUTURE DIRECTIONS	91
6.1. Summary	91
6.2. Limitations	93

	viii
6.3. Future directions	94
REFERENCES	96

LIST OF TABLES

TABLE 3.1: Time complexity for a single observation sequence of length T for M hidden Markov models, each consisting of K states.	42
TABLE 5.1: Performance comparison between state of the art algorithms and the proposed NILM, performed on REDD dataset.	73
TABLE 5.2: Performance comparison between SparseNILM and the proposed NILM, performed on AMDPs dataset.	76
TABLE 5.3: Performance comparison between SparseNILM and the proposed NILM, performed on mid-size commercial building dataset.	87

LIST OF FIGURES

FIGURE 1.1: Breakdown of estimated U.S. energy consumption in 2018.	2
FIGURE 1.2: Global primary energy consumption.	3
FIGURE 1.3: GDP per capita.	4
FIGURE 1.4: Effect of global carbon emissions on global average temperature.	4
FIGURE 1.5: An example of the aggregate electricity signal explained by it's individual constituents [1].	5
FIGURE 1.6: Expansion of energy use for space cooling, almost entirely shaped by electricity.	6
FIGURE 1.7: Expected growth of baseline scenario for Residential AC cooling by country.	7
FIGURE 1.8: Share of space cooling in peak load and electricity demand by country.	8
FIGURE 1.9: Example of net load on the Duke Energy Progress network.	8
FIGURE 1.10: Common types of air-conditioning system.	10
FIGURE 1.11: AC unit with a standard vapour compression refrigeration cycle.	11
FIGURE 1.12: Example fault -detection process for a heat pump in a retail bank in Charlotte, North Carolina. Left shows a lack of correlation between outdoor air temperature and HVAC energy for HVAC unit with faulty compressor.	12
FIGURE 1.13: Example fault-detection process for a heat pump in a retail bank in Charlotte, North Carolina. Left shows a lack of correlation between outdoor air temperature and HVAC energy for HVAC unit with faulty compressor.	13
FIGURE 1.14: Typical arrangement for a utility-installed smart meter along with the respective constraints for different frequencies of metering [2].	13

FIGURE 1.15: A comparison between the major individual constituents for a daily aggregate power consumption of (a) residential building and (b) commercial building.	15
FIGURE 2.1: Percentage of penetration for smart appliances in U.S. residential households.	22
FIGURE 2.2: The fundamental signal flow path in a standalone NILM.	24
FIGURE 2.3: Typical daily load profiles for various circuits in the retail environment. (a) Air conditioner, (c) exterior lighting, and (e) interior lighting and Typical transient profiles for various circuits in the retail environment (b) Air conditioner, (d) exterior lighting, and (f) interior lighting.	26
FIGURE 2.4: Daily example of aggregate power and its individual constituents for one house in REDD.	29
FIGURE 2.5: Daily example of aggregate power and its individual constituents for AMPds.	31
FIGURE 2.6: Daily example of aggregate power and its individual constituents for Pecan Street data.	32
FIGURE 3.1: Graphical representation of a hidden Markov model.	37
FIGURE 3.2: Graphical representation of a factorial hidden Markov model.	41
FIGURE 3.3: An example of TV Regularization with (a) $\lambda = 1000$ (over smoothing) and (b) $\lambda = 0.1$ (under smoothing).	43
FIGURE 3.4: A statistical demonstration of non-stationary baseline.	45
FIGURE 3.5: An example of aggregate power consumption, along with individual load constituents, for a typical weekday in a commercial building.	47
FIGURE 4.1: The diagram of the proposed algorithm.	49
FIGURE 4.2: An example of the Exterior light uncertainty along with the aggregate power in a mid-size commercial building.	51

FIGURE 4.3: An example of the output of the change of mean detection algorithm, highlighted by the grey area.	54
FIGURE 4.4: An example of (a) baseline estimation along with the (b) detrended signal.	56
FIGURE 4.5: An example of (a) Choosing the optimum number of clusters based on BIC and (b) the corresponding CKmeans clustering results.	59
FIGURE 4.6: Empirical ON Distributions of Loads in an example building.	60
FIGURE 4.7: Training a naïve gaussian Bayes classifier.	64
FIGURE 4.8: An example of (a) initial clusters (b) pruned sum of clusters (red are filtered).	66
FIGURE 4.9: Graphical representation of Gaussian mixtures with known parameters.	67
FIGURE 5.1: Confusion matrix.	71
FIGURE 5.2: Estimation accuracy results for 2 different test days compared to their respective ground truth.	74
FIGURE 5.3: Major monthly consumers in a retail bank in South Carolina. Note that HVAC electricity consumption is low in the winter because the facility relies on gas heat.	77
FIGURE 5.4: An example of a daily waveform in a commercial building.	79
FIGURE 5.5: NILM data acquisition.	81
FIGURE 5.6: Basic hardware arrangement in the custom data-acquisition system.	82
FIGURE 5.7: Pairwise distributions of three RTU units in a mid-size bank building during the first week of July.	83
FIGURE 5.8: Comparing the performance of SparseNILM (Top) vs proposed NILM (Bottom) on one month of data.	85
FIGURE 5.9: Sub-hourly load data vs. outside air temperature (one plot for each hour). Red dots are estimates, and black dots are ground truth.	89

LIST OF ABBREVIATIONS

AMPds Almanac of Minutely Power Data Set

CT Current Transducer

FHMM Factorial Hidden Markov Model

HMM Hidden Markov Model

NILM NON Intrusive Load Monitoring

REDD Reference Energy Disaggregation Data set

RTU Rooftop Unit

VRF Variable Refrigerant Flow

CHAPTER 1: INTRODUCTION

Figure 1.1 shows the typical view of the energy system. Energy stored for millennia in carbon bonds is released to provide thermal energy or mechanical work, driving transportation, industry, and buildings. It has been posited, however, that this technocratic view limits society's ability to control consumption. This alternative view suggests that the energy system exists because of the human need for the services derived from energy, including warmth, comfort, entertainment, mobility, hygiene, and safety¹. Thus, the system should be viewed from the perspective of the services derived on the right side of Figure 1.1. One result of this technocratic view is that significant low-cost energy reductions in the residential and commercial sectors are not achieved. It is estimated that about 20% of consumption in the US commercial and industrial sector, or 8% of all U.S. energy use and emissions, could be avoided with efficiency improvements to these building ([3, 4, 5]. Further, this estimate is derived from changes that can be achieved with little or even negative cost, making savings here particularly attractive [3]. One potential driver for achieving such reductions is to provide detailed, appliance-level usage information to consumers. Obtaining such information in an effective manner can be challenging and costly. One potentially achievable idea that has failed to find widespread adoption is the idea of energy disaggregation, which refers to a set of statistical approaches for extracting end-use and/or "appliance level" data from an aggregate, or whole-building, energy signal [2]. The potential similarity of this approach to that of speech recognition is clear as the laws of physics cause clear patterns in electricity demand that "communicate" a load's current state or activity. Unlike speech recognition, however, the electricity

¹<http://www.ieadsm.org/wp/files/Rotmann-BEHAVE-2016.pdf>

signal is not intended to be a communications medium â it just happens to be one. When using a voice-activated personal assistant, for instance, common sense leads one to minimize all background noise when providing the system with an input. When disaggregating consumptive patterns from an electricity signal, however, one must simply accept the realities of overlapping patterns with distinctly different sizes. This thesis provides an approach that has been demonstrated to overcome these issues in a number of cases, providing appliance-level information for the largest consumers in several buildings.

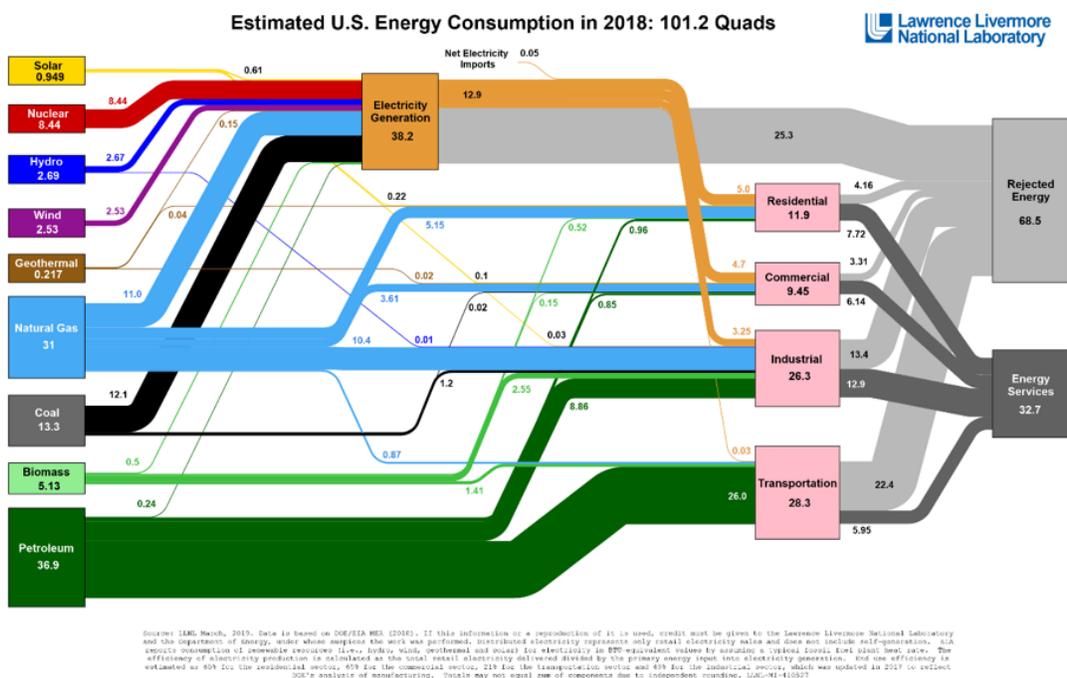


Figure 1.1: Breakdown of estimated U.S. energy consumption in 2018.

By unlocking the energy stored in fossil fuels, mankind unleashed human advancement at a rate unimaginable prior to the industrial revolution. Figure 1.2 shows the nearly exponential growth in primary energy consumption that has occurred since 1800. Figure 1.3 shows the growth in GDP per capita over the same period. The correlation is clear and not unexpected since mankind required energy to provide mobility and generate electricity. These innovations were essential drivers for economic growth and the human ability to harness creativity into invention. However, bene-

Since the growth has been, there are some unintended consequences. The reactions that unleash energy from fossil fuels also produce byproducts that contribute to the greenhouse effect. Figure 1.4 shows how carbon emission rates caused by the burning of fossil fuels have grown since 1900. Over that same time period, global temperatures have risen in nearly the same manner. The impact of this change is significant. The world's oceans have absorbed 93% of the excess heat created since the mid-20th century and about 25% of the carbon dioxide. The higher acidity and surface temperatures are changing overall climate patterns and have caused 7 to 8 inches of sea level rise since 1900. The potential impacts are significant. With continued growth in emissions at historic rates, annual losses in some economic sectors are projected to reach hundreds of billions of dollars by the end of the century—more than the current gross domestic product (GDP) of many U.S. states.

Addressing the megatrends noted above requires a myriad of solutions, including

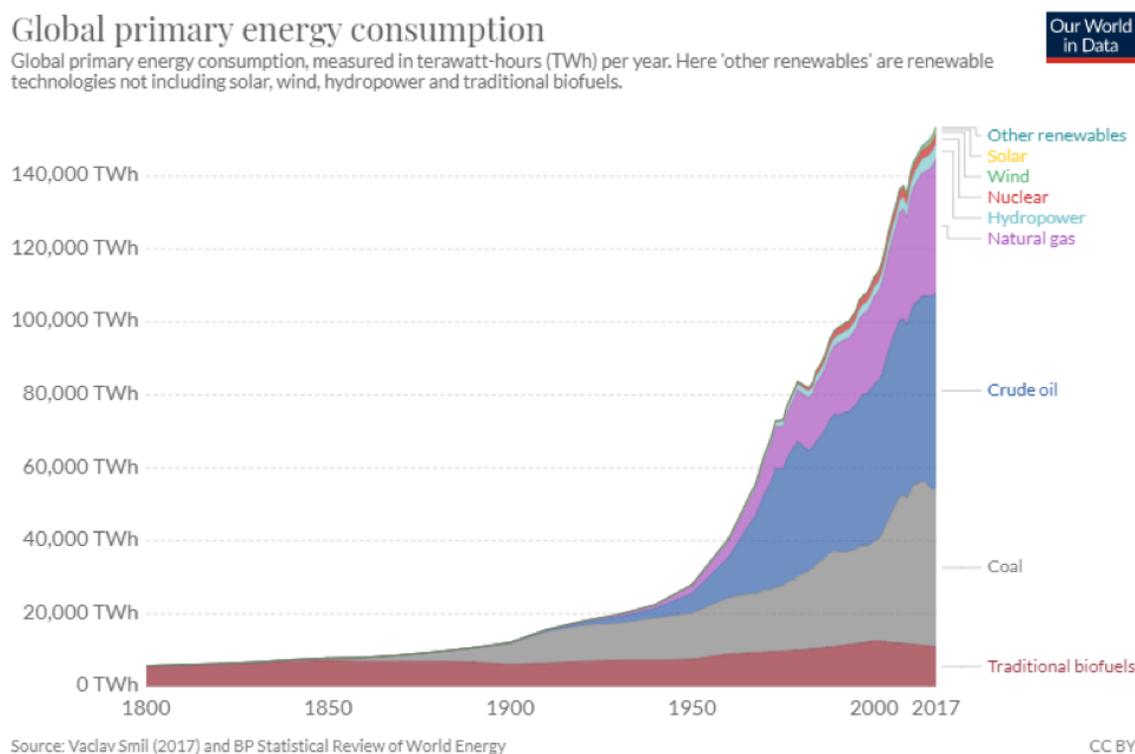


Figure 1.2: Global primary energy consumption.

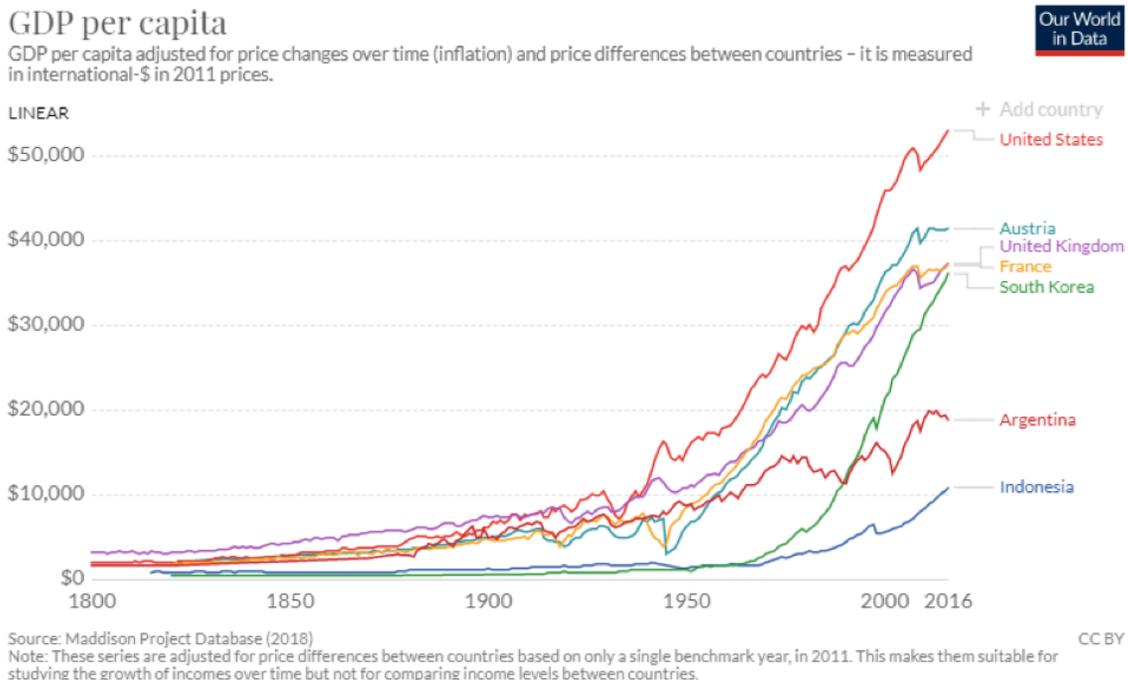


Figure 1.3: GDP per capita.

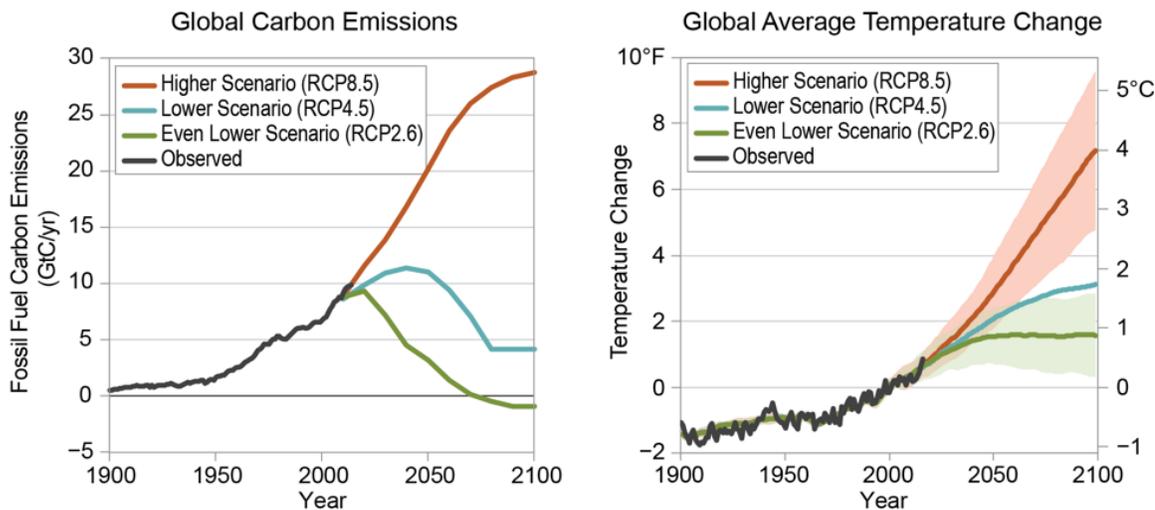


Figure 1.4: Effect of global carbon emissions on global average temperature.

technologies, policies, and behaviors. This thesis focuses on one potential solution that can help to reduce the electricity needed to provide human services. The specific innovation focuses on an ability to disaggregate appliance-level energy consumption data from an aggregate electricity signal. Figure 1.5 illustrates the basic idea. As loads operate, they impact the electricity flowing into a building. For instance, heating

water on a stove occurs by heating a resistive element. To do so safely, the power flowing into a stove burner is typically pulse-width modulated. A refrigerator, on the other hand, cycles periodically as a result of thermostatic control occurring inside the storage compartment. These patterns are clearly observed in Figure 1.5, and thus there is a desire to teach a computer to detect them automatically.

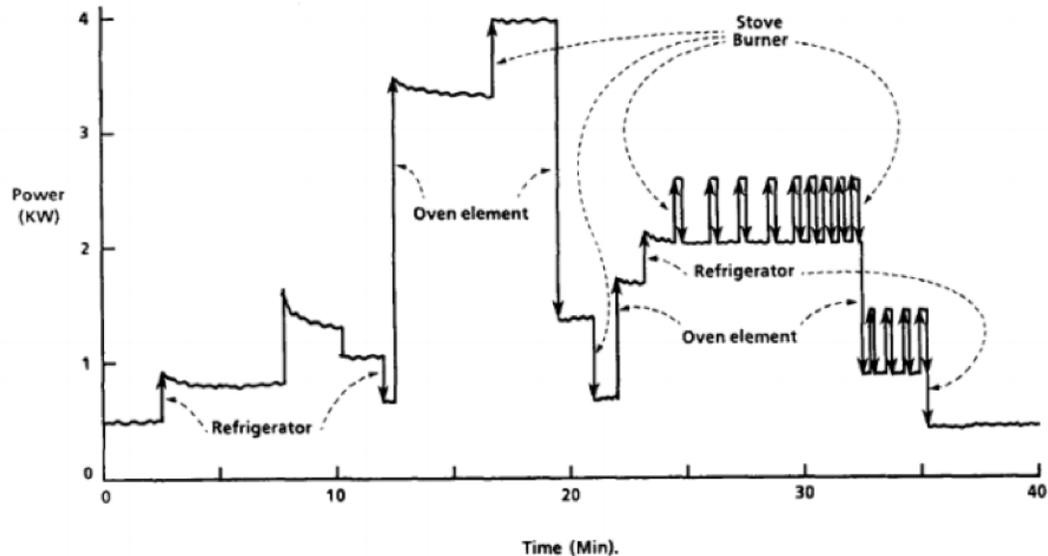


Figure 1.5: An example of the aggregate electricity signal explained by its individual constituents [1].

Armel et al. [2] provide a clear discussion on the potential benefits of disaggregation. One particularly poignant example is that of monitoring the energy required to heat and cool buildings. Figure 1.6 shows the growth in global energy consumption for space cooling². Final energy use for space cooling in residential and commercial buildings⁵ worldwide more than tripled between 1990 and 2016 to 2020 terawatt hours (TWh). The electricity used for space cooling required around 400 million tonnes of oil equivalent of primary energy in 2016 $\hat{=}$ or 3% of world total primary energy use $\hat{=}$ taking account of the large amounts of energy lost in transforming primary energy sources into electricity. This is equivalent to all the energy used for international

²https://webstore.iea.org/download/direct/1036?fileName=The_Future_of_Cooling.pdf

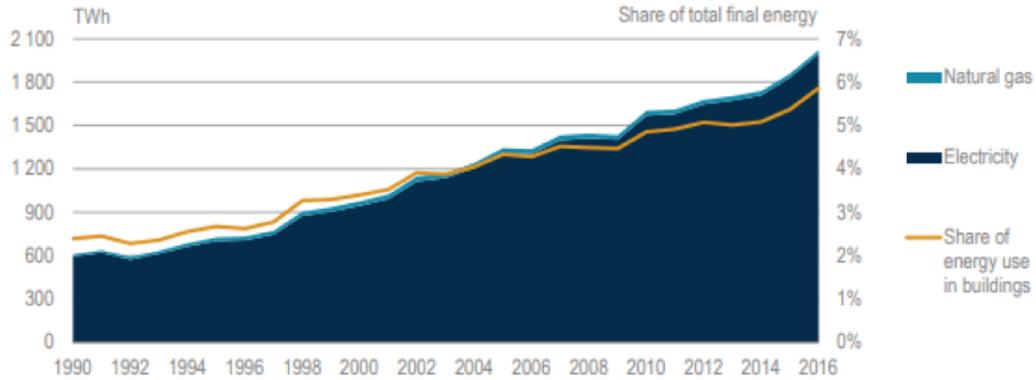


Figure 1.6: Expansion of energy use for space cooling, almost entirely shaped by electricity.

aviation and shipping worldwide.

Active air conditioning, as opposed to building designs that keep indoor temperatures down, is a relatively recent phenomenon. Although mechanical techniques for cooling indoor air were developed as early as the 19th century and the first modern electrical air-conditioning unit was invented at the beginning of the 20th century, the widespread use of ACs only started to take off, initially in the United States, in the 1950s with improvements in the performance of commercial devices, lower prices, and growing prosperity. The IEA recently labeled the growth in global demand for space cooling as one of the most overlooked energy issues of our time. If left unchecked, energy demand from air conditioners will more than triple by 2050, equal to China's electricity demand today. One of the major drivers is that of the 2.8 billion people living in the hottest parts of the world, only 8% currently possess ACs, compared to 90% ownership in the United States and Japan. Figure 1.7 shows the expected growth to 2050 in the residential sector.

Space cooling is not just a major driver of consumption, but also of demand, which is the rate of consumption. Figure 1.8 shows that air conditioning resulted in nearly 1/3 of peak load in the United States. Because air-conditioners drive peak loads, they present significant challenges in an electricity sector that relies heavily on distributed

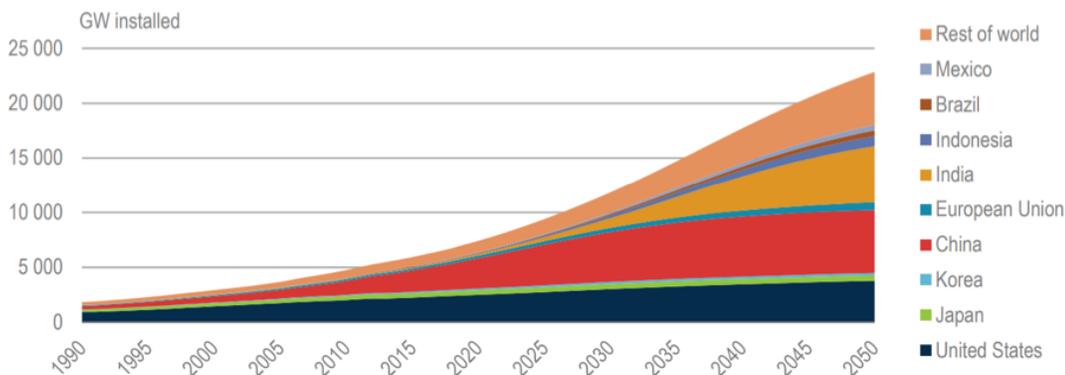


Figure 1.7: Expected growth of baseline scenario for Residential AC cooling by country.

renewable sources. Solar PV, for instance, is an excellent source of energy when it is hot and air-conditioning demand is high. During the winter, however, when space-cooling needs are lower, solar PV generation can be excessive and problematic. Figure 1.8, for instance, shows net load on the Duke Energy Progress network on January 24, 2018. As a result of the solar energy provided to the network during the day, the total load in the afternoon was significantly lower than the network was designed to provide. The area shown in yellow is load that is effectively removed from centralized generators as a result of distributed solar. An increase in solar generation, which is projected, would cause the net load to dip even lower, and force utilities to pay others to take electricity from base assets or invest in significant amounts of energy storage. Effective management of electricity consumption from space cooling is thus essential in a changing electricity sector.

Figure 1.10 shows the most common types of air-conditioning systems found in the world today. The majority of units found in homes and most small-to-medium-sized offices are the central ducted split system and the packaged rooftop unit (RTU). The former is commonplace in most single-family US homes, while the latter is common in commercial offices, restaurants, and other facilities. These units are based on vapor-compression refrigeration technology, which exploits a basic law of physics: when a liquid converts to a gas (in a process called phase conversion), it absorbs heat; and

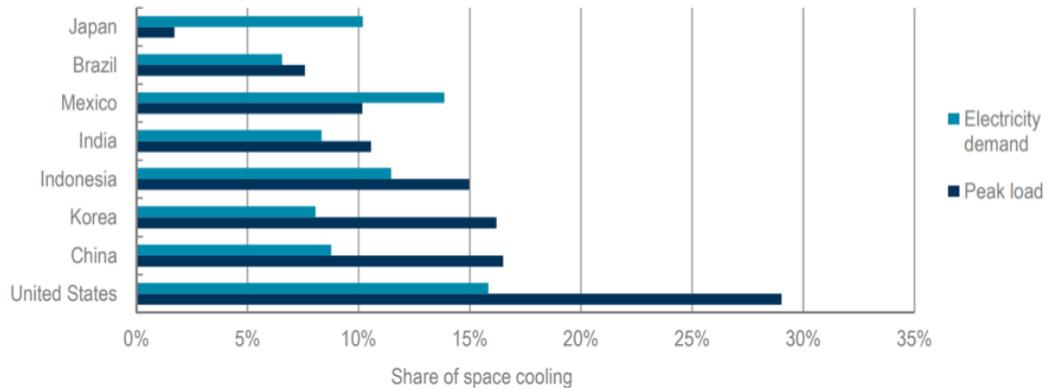


Figure 1.8: Share of space cooling in peak load and electricity demand by country.

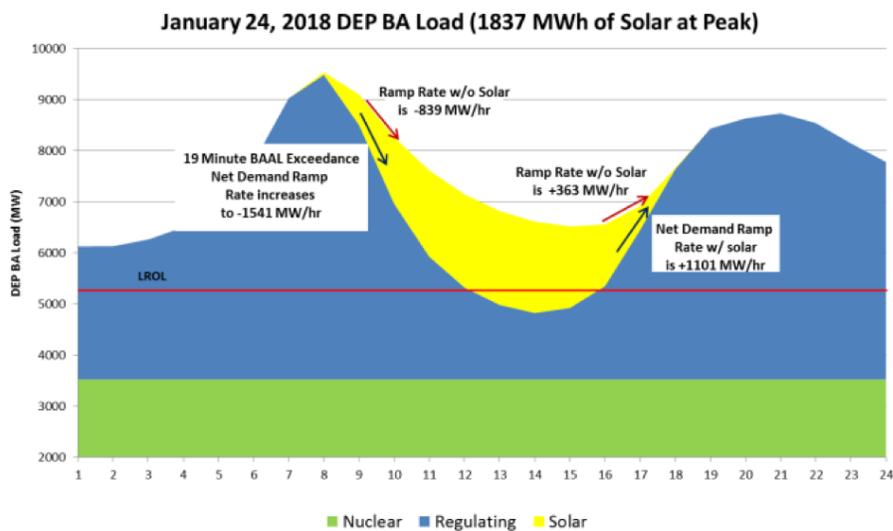


Figure 1.9: Example of net load on the Duke Energy Progress network.

when it condenses again (to a liquid), it releases heat. ACs exploit this feature of phase conversion by forcing either natural or special chemical compounds, known as refrigerants, to evaporate and condense repeatedly in a closed loop of coils. Figure ?? shows the basic process and equipment. A fan moves warm interior air over the cold, low-pressure evaporator coils. The refrigerant inside the coils absorbs heat as it changes from a liquid to a gaseous state, and thus cools the air. To keep cooling, the AC converts the refrigerant gas back to a liquid again. To do that, a compressor puts the gas under high pressure â a process that releases heat. The heat created by compressing the gas is then evacuated to the outdoors with the help of a second set

of coils called condenser coils, and a second fan. This cools down the gas in the coils and turns it back into a liquid, and the process starts all over again in a constant cycle: liquid refrigerant, phase conversion to a gas and heat absorption, compression, and phase transition back to a liquid again.

Energy disaggregation can provide significant benefits in detecting defective performance in vapor-compression systems. Such air-conditioners typically operate thermostatically to remove heat from occupied spaces. Figure 1.12 represents a scenario that can frequently occur. Both graphs show the hourly energy consumption by a single air-conditioning unit operating in a small bank branch in Charlotte, North Carolina, as a function of average hourly temperature. The graph on the right shows a behavior similar to what one might expect; in general, consumption rises as outdoor temperature rises. The graph on the left, however, shows minimal correlation with outside air temperature. The data shown on the left were recorded prior to the detection of a faulty compressor and thermostat; the data on the right were recorded after these issues were fixed. Insights such as those provided by Figure 1.12 are easily obtained when appliance-level data can be monitored. Figure 1.12 also highlights another potential issue. Without such data, faults in many systems can go undetected for months or even years. In the absence of data, the primary sensor for faults in air conditioners are occupant complaints. If occupants are comfortable â or at least not complaining about being uncomfortable â a unit can potentially operate in a significantly inefficient manner, such as the one shown in Figure 1.12.

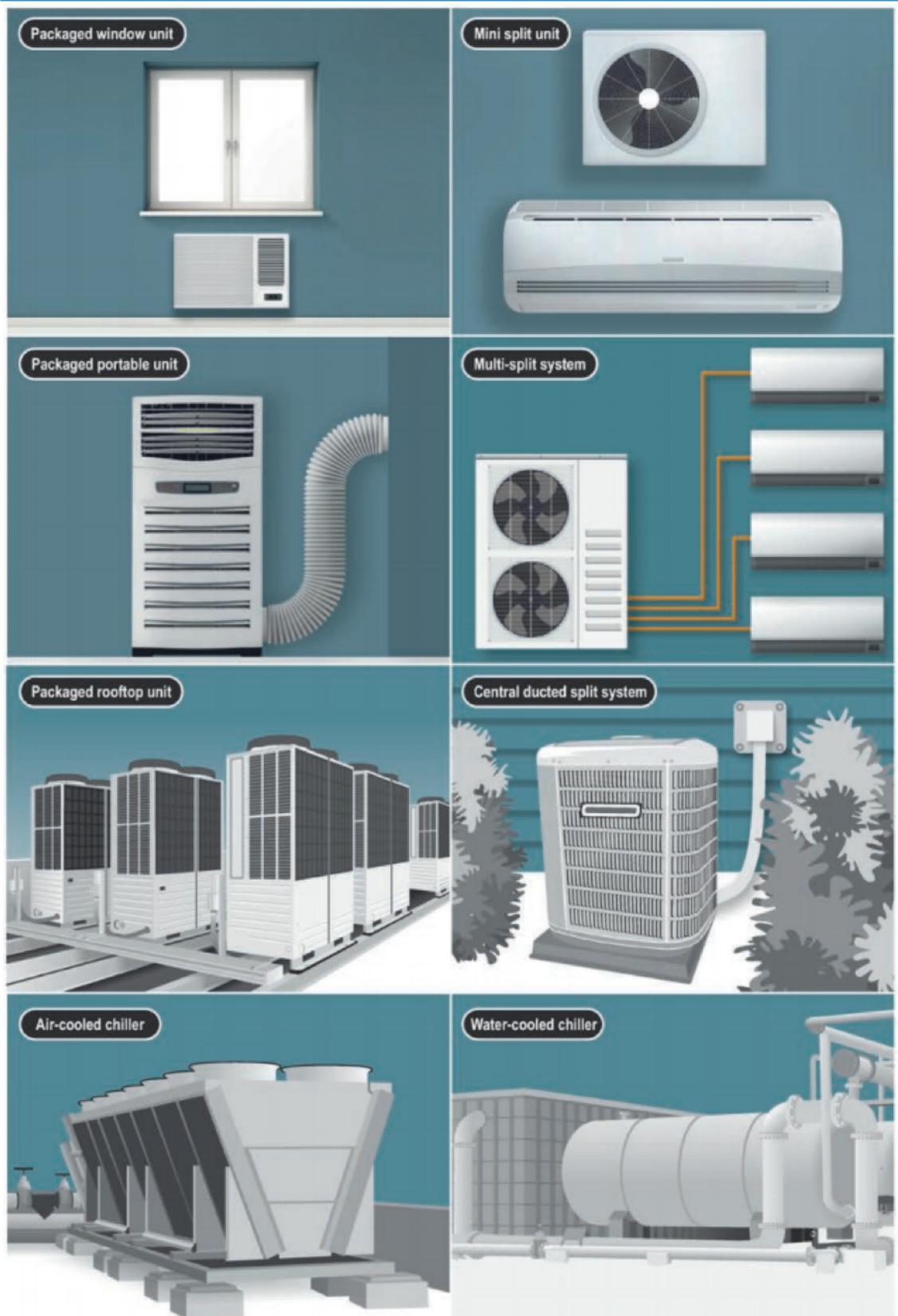


Figure 1.10: Common types of air-conditioning system.

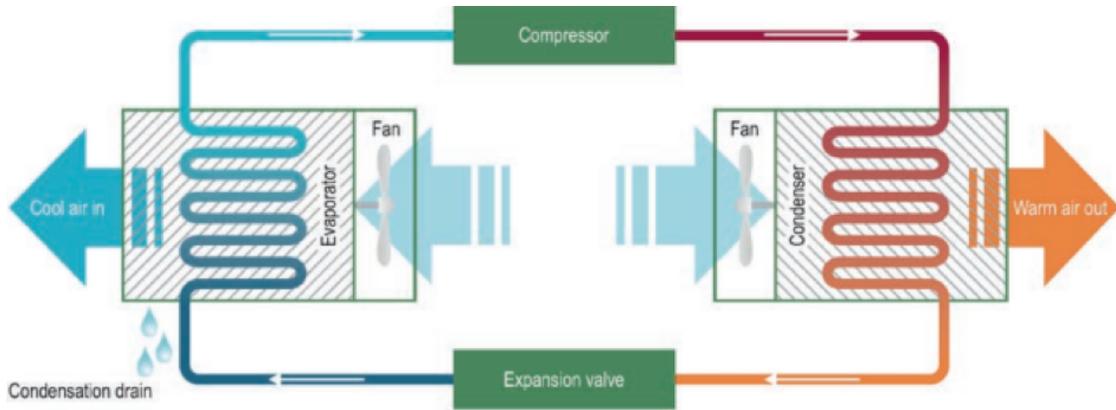


Figure 1.11: AC unit with a standard vapour compression refrigeration cycle.

The powerful insights that can be obtained using appliance-level data was highlighted previously. Other researchers have noted other opportunities [6]. One challenge is the ability to gather such information. Figure 1.13 compares cost, installation effort, and adoption level of various technologies capable of providing appliance-specific information. The first row focuses on hardware solutions, which include appliances capable of providing consumption data automatically, and add-on measurement devices. Note that both such solutions have clear drawbacks, including cost and slow adoption rates. Software-based solutions focus on algorithmic approaches to disaggregating loads from a whole-building energy signal such as the one shown in Figure 1.5. The first of these are consumer-purchased devices and the latter are smart meters provided by utilities. As of 2017, nearly 79 million such meters had been deployed by US utilities [7].

Smart meters provide an interesting opportunity to perform appliance-level disaggregation. The block diagram on the left side of Figure 1.14 shows the typical arrangement for a utility-installed smart meter. These devices have a modem that communicates over a utility-owned wide-area network (WAN), typically at 15-minute or hourly intervals. These devices can also communicate over a home-area network (HAN) to send information to homeowners. Previous research has shown [2] that most large loads, including air conditioners can be detected using minutely data.

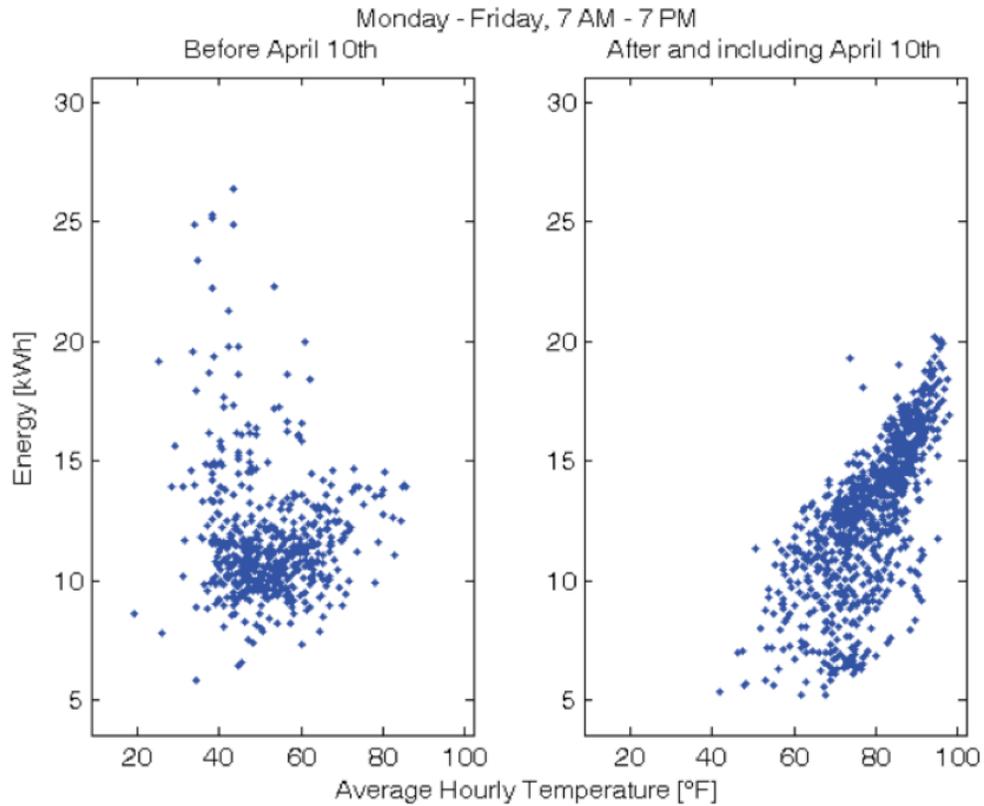


Figure 1.12: Example fault -detection process for a heat pump in a retail bank in Charlotte, North Carolina. Left shows a lack of correlation between outdoor air temperature and HVAC energy for HVAC unit with faulty compressor.

The image on the right in Figure 1.14 indicates that most smart meters – which are reaching massive deployment levels – can provide the minutely data needed for disaggregation if firmware upgrades are provided.

If utility smart meters can provide the data needed for appliance-level disaggregation and such meters are deployed at scale, a natural question arises as to why disaggregation is not common. The answer lies in the fact that disaggregation is a computationally challenging problem and the relative value is not immediately clear at the level of any one individual homeowner.

	Sensing Technology	Cost to Consumer	Installation Effort	Adoption
Hardware Disaggregation	Plug Level Hardware Monitors (e.g., Kill-A-Watt, EnergyHub)	\$30-\$50/plug; \$300-600/home	Most plugs – Med 240V plugs - Hard	Low; in existence for past 7-8 years
	Smart Appliances	\$100+ additional compared to non-Smart appliances	Easy	10-15 years after introduction for mass adoption
Software Disaggregation	House Level Current Sensor (example - TED, Blueline, Egauge etc.)	\$200+/house	Very Hard	Low (high cost + high effort)
	Smart Meter	None	None	Very High & fast (installed by utilities)

Figure 1.13: Example fault-detection process for a heat pump in a retail bank in Charlotte, North Carolina. Left shows a lack of correlation between outdoor air temperature and HVAC energy for HVAC unit with faulty compressor.

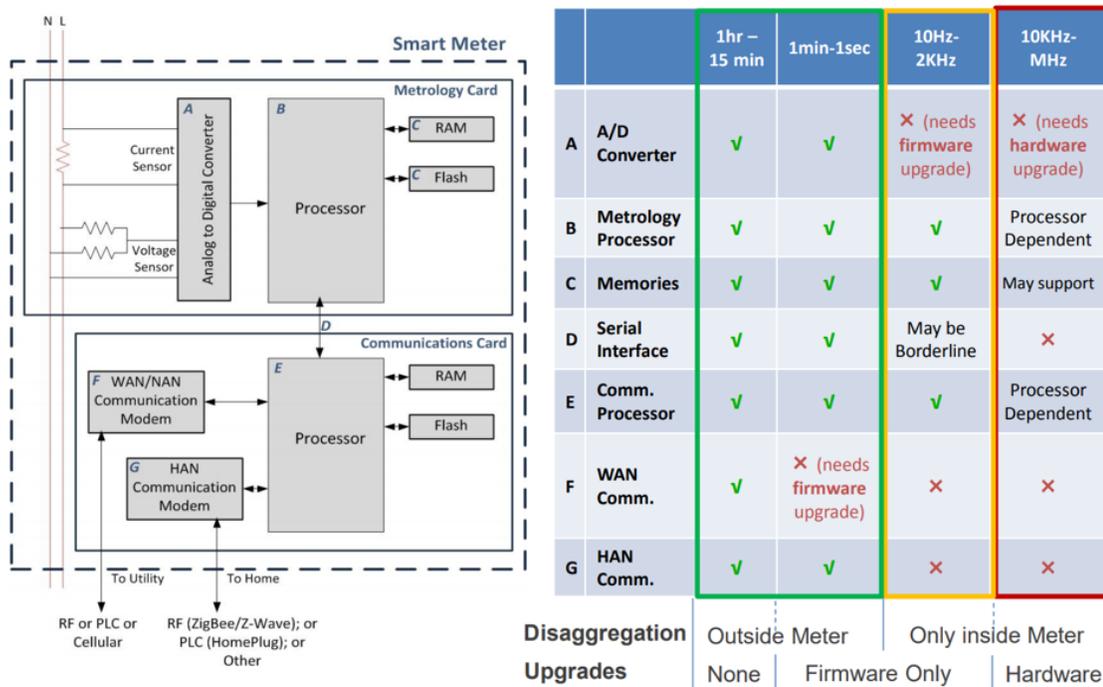


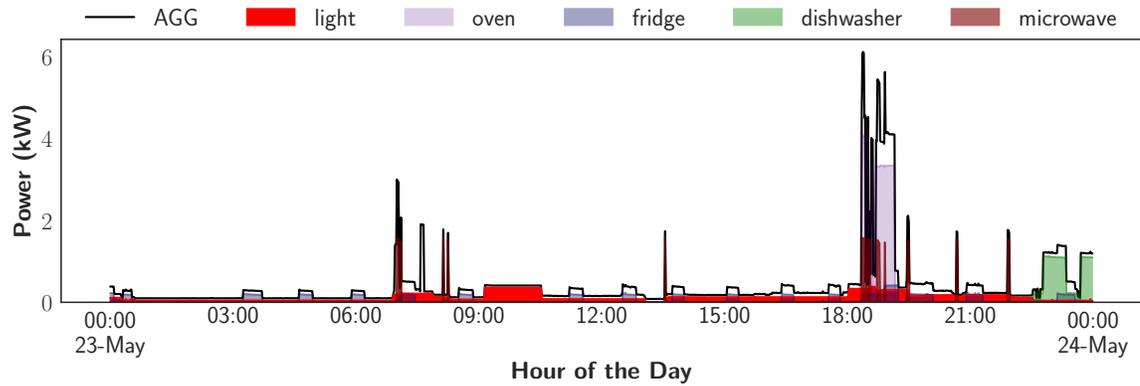
Figure 1.14: Typical arrangement for a utility-installed smart meter along with the respective constraints for different frequencies of metering [2].

This thesis is concerned with the research problem of inferring individual appliance consumption of a building for a given observation of the power at the meter. This

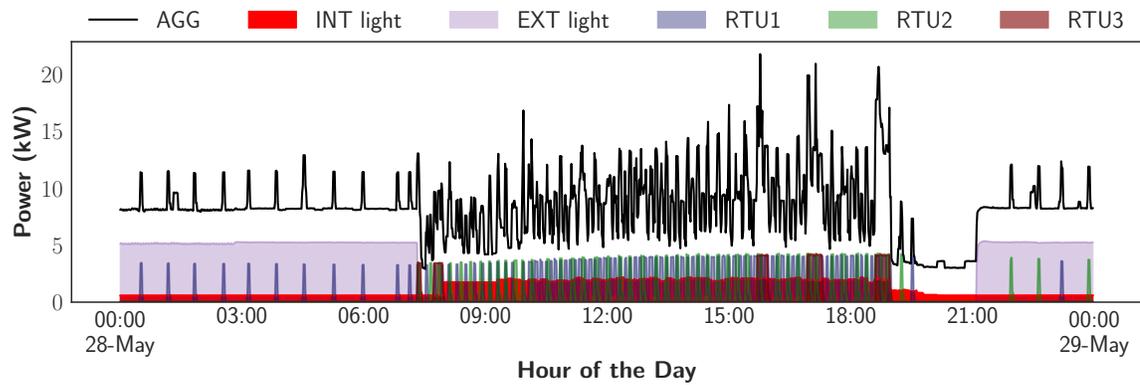
practice is called energy disaggregation and it was first proposed more than 30 years ago by Hart [1] in his seminal work. Since then, the primary focus of disaggregation has been the residential sector. Initially most efforts involved very limited single house scenarios at university research level. More recently, the creation of public datasets collected from several residential buildings [8, 9, 10] have provided a more unified test bed for comparing different approaches but these efforts remain limited to the residential sector.

Researchers have long acknowledged the importance of having a common dataset to define characteristics of appliances along with their operating conditions properly. The common dataset should ideally be drawn from a broader geographic portfolio for further robustness. In addition, it should account for different sizes of buildings, starting from small residential to larger commercial buildings. To this date, however, there has been very limited research done on commercial buildings, and there remain large untapped saving potentials in the commercial sector [2]. Figure 1.15 illustrates the contrasting patterns of one residential house in the REDD[8] dataset compared to that of a medium-size commercial building. In these figures, there are five major notable differences: 1. The demand is far greater in the commercial building in comparison to that of the residential apartment. 2. The day time and night time consumption patterns are very distinguishable in a commercial building in a way that is not so evident in the residential apartment. 3. There is much more concurrent activity in the commercial building, which makes disaggregation a very harder problem compared to that of the residential dataset 4. There is a varying baseline during occupied hours in the commercial building that has to be carefully considered as it can potentially greatly affect the result of disaggregation. We discuss this in more detail in the upcoming chapters. 5. HVAC and exterior light constitutes a major portion of the power consumption in the commercial building. All of these further stresses the need for tailoring an algorithm that is capable of handling the unique challenges

of disaggregation in the commercial sector but such a goal can not be achieved without first collecting a comprehensive dataset along with corresponding appliance level metadata as ground truth for validation purposes.



(a)



(b)

Figure 1.15: A comparison between the major individual constituents for a daily aggregate power consumption of (a) residential building and (b) commercial building.

It is in this context that we define the following research objective: To develop a novel algorithm capable of disaggregating the major loads in a portfolio of commercial buildings, namely HVAC and exterior lighting, from minutely recorded smart meter data. The goal of this algorithm would be to provide feedback, i.e., a breakdown of power consumption that can be used for diagnostics purposes and/or informed decision making. We aim for a minimum level of supervision in the algorithm, limited

to few hyperparameter tuning and load labeling.

The contributions of this thesis are summarized as:

- **A detailed review of NILM:** Providing an in-depth look at the state of the art NILM algorithms, specifically those using hidden Markov models. We point out the shortcomings of each algorithm and potential ways to improve the performance, especially in the areas of minimal supervision and robustness.
- **Data acquisition for commercial buildings:** Creating a dataset from several commercial buildings through monitoring 80 plus individual channels in each building. The dataset contains several months of data in each building, and to the best of our knowledge, it is one of a kind in NILM research.
- **A novel robust NILM:** We have created a novel NILM algorithm, capable of performing disaggregation on multiple residential and commercial buildings. The proposed algorithm can handle a varying baseline as well as being robust to un-observed loads. Our algorithm does not need metadata for training and is also computationally efficient and therefore suited for edge computing.
- **Detailed comparison against state of the art algorithms:** We show an improvement in the result of disaggregation against the state of the art NILM research. We empirically demonstrate the superiority of our approach by running the algorithm on public datasets, namely AMPds[9] and REDD[8], achieving better precision and recall rates.
- **Disaggregating major loads in commercial buildings:** Finally; we accurately disaggregate the HVAC and exterior light in a couple of commercial buildings for several months.

In the remaining chapters of this thesis, we address the proposed work as follows:

Chapter 2 presents a historical overview of the different types of monitoring in buildings. We begin by discussing the consumption patterns in buildings, both in the residential and in the commercial sector and the merits of tracking in general. We then discuss different methods to carry this task. Among these, we emphasize non-intrusive low-frequency methods, which is the scope of the problem defined in this thesis. We give a brief description of the available public datasets and also discuss the challenges of NILM and why it is still a relevant problem.

Chapter 3 outlines the limitations of different HMM based approaches for NILM in detail. We discuss inference for NILM as an NP complete problem and talk about the approximation methods used for inference in HMMs. We also discuss the problem of supervision, addressing the need to move away from training NILM using metadata. Most important of all, we discuss how the current methodologies are incapable of producing accurate results when challenged by an unobserved variation in the electricity signal.

In Chapter 4, we propose a novel algorithm capable of doing *robust NILM*, by which we refer to a methodology that is capable of monitoring the power demand of major consumers of energy for a large portfolio of buildings. To this end, we break down the proposed approach into three major steps that include background filtering, edge analysis and load modeling, and state inference and power estimation.

Chapter 5 is an in-depth look at the results of deploying the proposed algorithms into the problem of energy disaggregation. To this end, a comparison is made for the performance of the proposed NILM against that of the state of the art methodologies discussed in Chapter 2 and 3. We first define the grounds of comparison in detail. Using the defined metrics, a performance comparison of different algorithms on the REDD [8] and AMPds[9] datasets is presented. Finally, and most notably, we make a performance comparison between the proposed NILM and SparseNILM [11] for data collected from a couple of mid-size Commercial buildings. A thorough description of

data accusation and the general characteristics of the dataset is provided alongside these results.

Finally, Chapter 6 gives a summary of the findings presented in this dissertation. We summarize the research findings and elaborate on the limitations of the current work. In the concluding remarks of the thesis, we discuss the potential ways to extend this research to be able to make NILM a practical solution for the market.

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

This chapter presents a historical overview of the different types of monitoring approaches in buildings. We begin by discussing the consumption patterns in buildings, both in the residential and in the commercial sector, and argue the merits of tracking certain appliances. We then discuss different methods to carry this task. Among these, we emphasize nonintrusive low-frequency methods that closely relate to the goals set in this thesis. Finally, we give a brief description of multiple available public datasets and their individual load characteristics.

2.1 Building Energy Monitoring

According to the latest reports by the U.S. Energy Information Administration[7], residential and commercial buildings account for approximately 40% (or about 40 quadrillion British thermal units) of the primary energy consumed in the United States each year.¹ Data suggests that faulty or poorly operating systems are responsible for approximately 20% of this consumption [6]. It has long been suggested that intelligent data collection methods can detect such inefficiencies and thus drive savings [1, 2].

Specifically, in the commercial sector, which is the main focus of this thesis, several companies have recently attempted to capitalize on this potential, and tools for large and medium-sized commercial buildings are growing in acceptance [12]. This new development arises from the fact that larger buildings typically feature hundreds or thousands of small components that work in concert to provide thermal and visual comfort. A centralized automation system manages these many subsystems and ac-

¹Energy consumption by the commercial sector also includes energy consumption for street and other outdoor lighting, and for water and sewage treatment. However, these energy uses are relatively small contributors to the commercial sector's total energy consumption.[7]

quires a vast trove of information in the process. The use case is not as clear in small commercial buildings; however, because of their relative simplicity. For instance, most small offices and retail locations only require a handful of air-conditioning units, and only the most advanced among them include lighting control systems. The aggregate impact of these small and simple facilities, however, is enormous. Buildings occupying less than 50,000ft² comprise 94% of the commercial sector and consume about 47% of the energy used by buildings overall [13]. Given the comparatively small footprint and a vast number of such buildings, most property-management firms have relatively few maintenance technicians forced to operate in response to occupant complaints. As a result, maintenance issues are rampant. One major study focused on 4,168 operational commercial air-conditioners found that about 44% had improper airflow, and some 72% had improper refrigerant charge [14]. Consensus suggests that as many as 70% of all small commercial air conditioners have some form of fault [15]. These findings highlight the challenging reality: the opportunity to achieve savings at low cost is real, but doing so requires countless small actions in numerous buildings by a limited number of trained professionals.

Networked data tools represent the most elegant technical solution to the problem noted above, providing the opportunity to monitor many locations from one central command center. The problem, of course, is that decision-making requires detailed information about individual loads and their performance as a function of time, weather, and occupancy. The most common commercial solutions usually take one of two forms. The first of these is benchmarking tools that compare the annual, monthly, or daily consumption between different buildings [16]. The latter utilizes more granular information provided by some combination of sub-meters and unit control signals (i.e., for lighting and space-conditioning systems). The latter can provide robust and useful analytics, but the cost can be extremely hard to justify, given the relatively low probability of finding a major problem in any one facility. Therefore,

a low-cost alternative solution to gather granular information on buildings is highly valuable to achieve the goals mentioned above.

Different methods of intelligent data collection or monitoring as we would refer to it in this thesis can be divided into two main subcategories:

- **Intrusive Monitoring:** referring to those monitoring methods that require a point of measurement from each appliance inside a building whose tracking is of interest.
- **Non Intrusive Monitoring:** referring to monitoring methods inferring the status of several appliances in a building from a single (or at most a few) point of measurement.

2.2 Intrusive Monitoring

The main advantage of using intrusive methods is their high accuracy and limited need for post-processing. However, this comes at a high cost of installment and maintenance. There are more than a few ways to do intrusive monitoring, and recently there has been an interest shown in intrusive methods with a reduction in the cost of sensing technologies and the emergence of IoT. Historically speaking, intrusive methods suggested the installation of an electrical meter for each appliance in the building. This can be either a clamp meter, i.e., a current transducer (CT), calculating the current by measuring the electromagnetic field around the wire or a plugin meter that you plug into a power socket. The plugin meter has a delicate installation procedure because it requires a restart of the appliances to locate their presence by the meter. This is not always practical, especially in commercial buildings and for more sensitive loads. Additionally, both of these methods are highly costly in installation and maintenance, as they often require the presence of electrical technicians.

An alternative to direct electrical metering is smart appliances, which are expanding their presence in the households by the day. The smart appliance is capable of

reporting its energy use and status if connected to a WiFi network. Figure 2.1 depicts the current numbers and future estimates of smart homes in the United State [17]. As shown by this chart, the smart appliances are nowhere near the number required to reach the goals of monitoring, even in a very tech-friendly country such as the United States. This fact only leaves the door open to retrofitting old appliances in favor of making them *smart*. However, retrofitting can often be even more costly and impractical than direct electrical monitoring.

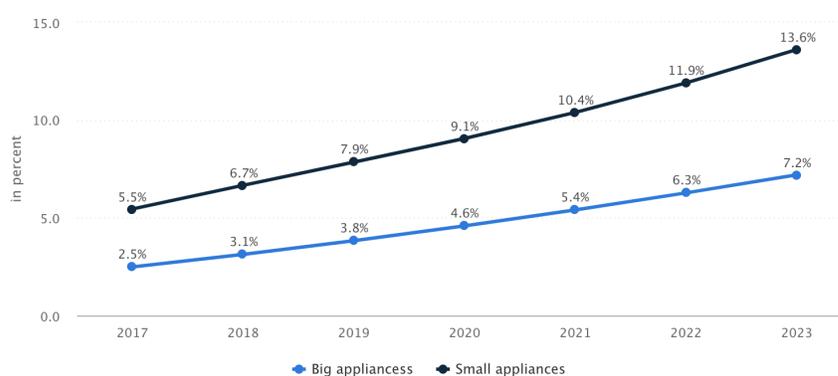


Figure 2.1: Percentage of penetration for smart appliances in U.S. residential households.

It is worth noting that in the commercial sectors, and specifically in more prominent buildings, state of the art Building Management Systems (BMS) have provided more granular information on buildings in the past decade or so. This provides very useful side information that can be leveraged in monitoring scenarios, but there remain many problems that need to be properly addressed before making use of such data. These issues include data quality, sensor tagging, sensor calibration, and maintenance, etc. Furthermore, the industry is yet to adopt a unified code for the process, and there are ongoing efforts to standardize BMS data [18, 19] Last, but not least, even if the majority of households were to be equipped with such a technology, there are still numerous security and privacy issues that have to be addressed. This fact alone can create a whole new domain of problems for energy monitoring.

Over the past thirty years, there have been a few other suggestions for intrusive monitoring mostly conducted in limited scope at the university research level. McWilliam [20] suggests a tagging process with RFID signals using the electrical cable for communication. Some have suggested using other sensor measurements such as temperature, acoustics, lights density, etc., [21, 22]. All of these methods would likely involve some form of retrofitting when it comes to broad-scale implementation.

2.3 Non Intrusive Monitoring

History of Non Intrusive Appliance Load Monitoring (NIALM)² goes back to more than 30 years ago. Initially proposed by Hart [1], it only concerned about the problem of tracking appliance consumption inside a residential building. Hart did not pursue the problem much further as he shifted his focus in research, but much of the research in this field has been built on the framework that he proposed in his seminal paper. This involved an edge clustering algorithm as the basis of building a Finite State Machines (FSM) to model the loads. A Viterbi algorithm [23] is then used to infer the state of each load from a single power meter measurements. In the years that followed, NILM received limited attention. More recently, a roll-out of smart meters, along with the creation of public datasets, advances in publicly available software, and the dawn of the machine learning era, has generated much buzz around this problem. According to a recent study [24], the number of publications alone in this field has grown exponentially over the past 15 years.

Figure 2.2 presents a block diagram of a modern NILM system (This example shows a three-phase installation as might be found in a small commercial building). Voltage and current transducers are installed at the utility service entrance, where a preprocessor algorithm computes useful data streams such as real or reactive power. Subsequent steps focus on event detection, classification, and diagnostics [25].

²In this thesis, the terms NILM and NIALM are used interchangeably. Energy disaggregation is another commonly used phrase as a substitute for these terms.

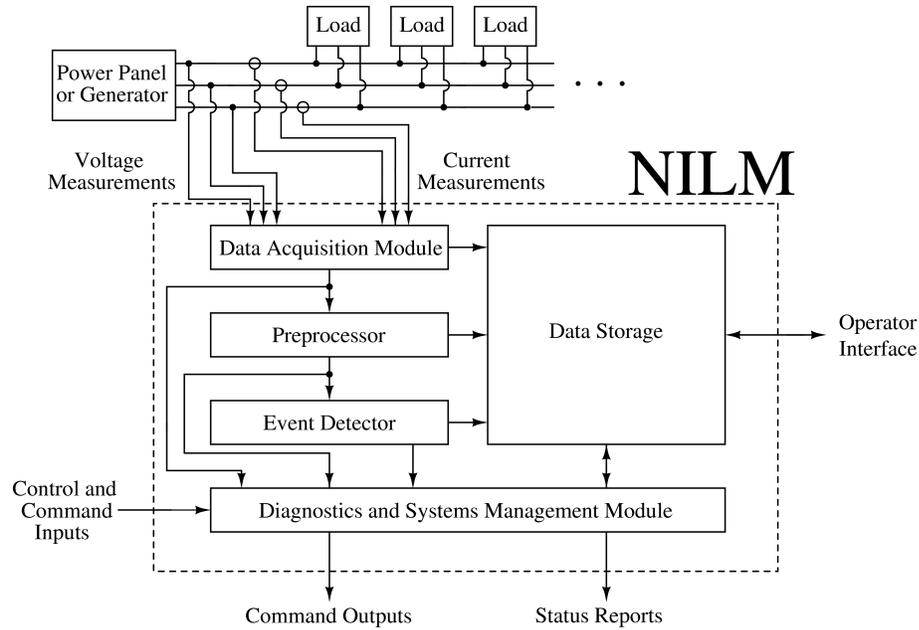


Figure 2.2: The fundamental signal flow path in a standalone NILM.

The classification of events can be done using many different characteristics. These can be put into two major subcategories, namely high frequency, and low frequency, that involves information contained over various time scales [26]. This information is best understood by considering several examples. Figure 2.3 a, c, and e show normal daily load profiles for typical circuits found in the retail environment. Figure 2.3 a, for instance, shows the load profile for a normally operating air conditioning unit; Figure 2.3 c shows the profile for an exterior lighting circuit; and Figure 2.3 e shows the profile for an interior lighting circuit. Note that the air conditioner shows intermittent cycling that decreases its frequency at night when it is cooler outside, and the interior set point is higher. Similarly, the exterior lighting circuit energizes after sundown and de-energizes after sunrise. The interior lights, on the other hand, show numerous on-off patterns during daytime hours and again for several hours after 8 PM when the janitorial contractors enter the building. These patterns, as well as the specific average power levels for each load, are well correlated with expected physical behaviors. One can collect data from various locations and understand how actual

operating conditions affect these patterns. Similar physical intuition can be applied at much shorter time scales. Figure 2.3 b, for instance, shows the typical behavior of a packaged rooftop air-conditioning unit. Note that the air handling fan first energizes and draws a relatively large in-rush current as the motor accelerates the fan blades from rest. Several minutes later, the compressor energizes, and a similar in-rush ensues. For a short time after the second in-rush transient, the steady-state power fluctuates as the refrigerant experiences transient changes in temperature and pressure. If one were to examine other air-conditioning units, similar yet different behaviors would be observed. In-rush times, for instance, vary based on the compressor type (i.e., either reciprocating or scroll) because of the physical differences between them [27], and the specific power levels vary based on the refrigerant state, outdoor weather conditions, and system size [28]. By comparison, Figure 2.3 d shows the transient performance of an exterior metal halide lamp. Note that there is an initial in-rush as the argon arc forms. Subsequently, there is a long transient as the temperature and pressure in the inner arc chamber change, and the lamp begins to reach its full light capacity. Again, behaviors for specific lamps would be different, but the general behavior is similar because of the internal physics. Finally, Figure 2.3 f shows the transient behavior of an interior lighting circuit with switched fluorescent lamps. In this case, one observes a very short inrush as the internal gases ionize, followed by an almost immediate steady behavior.

The examples shown in Figure 2.3 illustrate the various time-scale issues at play when examining load profiles. Not only are there notable and useful patterns at daily scales, but there are also useful transient patterns for the same load that happen at the sub-second, second, and minute scales. Most importantly, all of these behaviors are physically justified, and thus the general patterns can be used for training. Note that all of these patterns, however, may be different from one specific load to the next, and possibly from one start to the next because of effects such as heating and

gas pressures.

Having discussed these examples, we now review some of the research that has been done to leverage the many characteristics of power waveform for NILM.

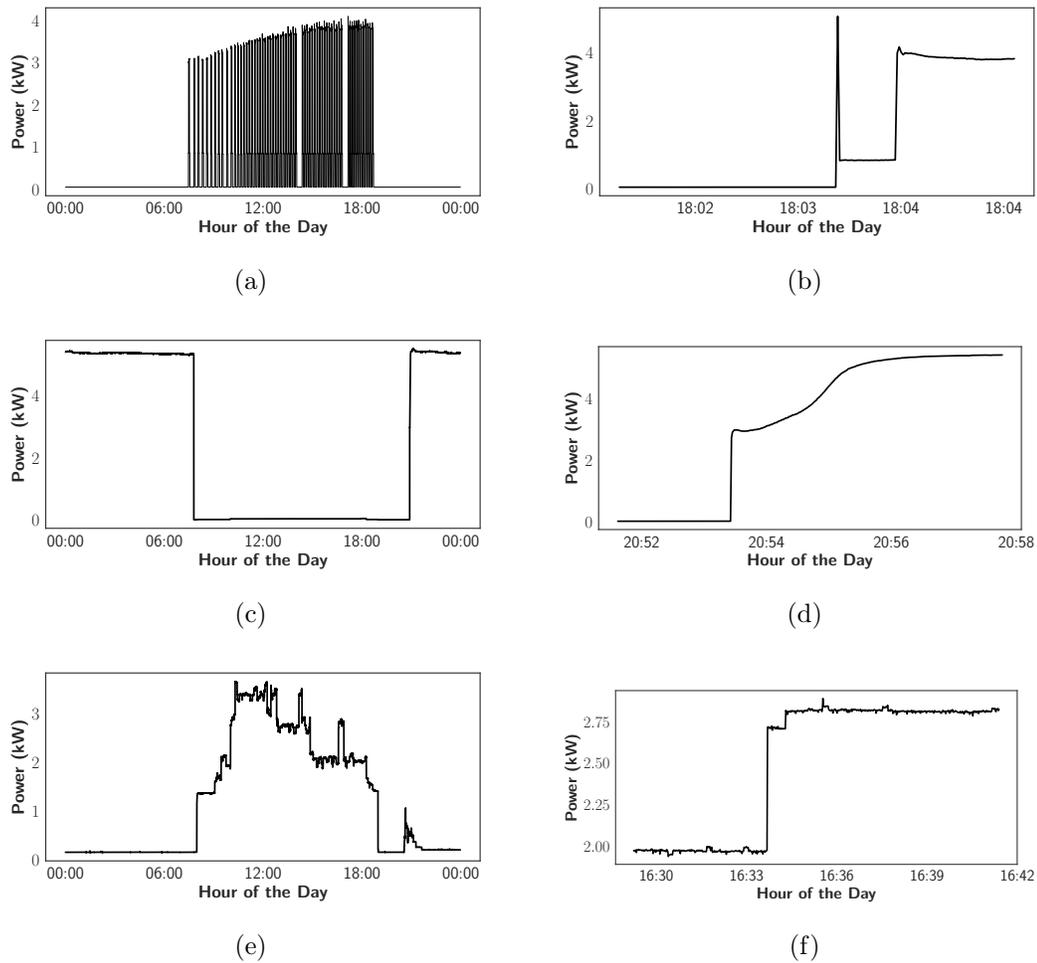


Figure 2.3: Typical daily load profiles for various circuits in the retail environment. (a) Air conditioner, (c) exterior lighting, and (e) interior lighting and Typical transient profiles for various circuits in the retail environment (b) Air conditioner, (d) exterior lighting, and (f) interior lighting.

2.3.1 High Frequency Methods

There are various useful information that can be considered in a high-frequency NILM scenario. The real and reactive power features can distinguish loads with similar steady-state power[1]. Loads that have a continuously variable power consumption

and produce higher-order harmonics can be classified by analyzing their Fourier series coefficients [29].

Transient shape information also assists in load recognition. Most loads observed in the field have repeatable transient profiles or at least sections that are repeatable [25, 30, 31]. The classifier attempts to match previously defined transient shapes known as exemplars to any transients observed in the incoming aggregate power stream. These exemplars are historically determined during a one-time training process. When events occur, the resulting transient is fit to each exemplar using different methods such as a least-squares process [30, 31], SVM classifier [32] etc. The duration of a transient can also be effective in the "turn-on" event classification [33].

Typical modern smart meters are designed to store measurement at once per minute frequency. This is far from the resolution needed to do disaggregation based on harmonics and transient features. While the importance of analyzing these harmonics in differentiating loads with similar steady-state can not be denied, it is simply impossible to leverage such information without further retrofitting of buildings. These methods are considered to be out of the scope of the goals of this research and are therefore not pursued any further in this work, and we instead shift the focus on NILM using the state of the art smart meters.

2.3.2 NILM with smart meter data

Event detection and classification, which is sometimes called "disaggregation," has long been the focus of academic literature. In this problem, one seeks to detect events in the aggregate power signal and then associate these events with particular loads. Several different techniques have been proposed over the years, all of which are predicated on the notion that the power drawn by an individual load is related to the physical task it performs [25]. Most approaches proceed in two steps, with the first focused on locating changes in the power signal. This step has been implemented using changes in steady-state power levels [1, 34]. In the second step, features extracted from

the power signal are fed to a classifier that associates them with a particular load. Artificial neural networks [35], support vector machines [36], clustering algorithms [37, 38], traditional, bayesian and additive factorial hidden Markov and semi Markov models [39, 40, 41], deep learning [42] have all been applied.

In the previous chapter, we made a case for minimal supervision in NILM. Supervised algorithms, such as various supervised machine learning algorithms named above, violate this assumption. For better accuracy, these algorithms often require a significant amount of labeled metadata. They are also limited by what they learn in training examples, and their performance can be highly sensitive to unseen data [43].

It is, therefore, our intention to further exploit HMMs in this work as they have provided a backbone for unsupervised NILM methods over the years, and have also had major success when deployed in similar types of problems in other domains such as speech recognition. The fundamental theory of HMMs is explained in detail in the next chapter.

2.3.3 Public Datasets

Despite approximately 30 years of work by NILM researchers, the technique has achieved relatively limited commercial success. A big factor in this has been the lack of sufficiently large enough datasets to allow researchers to understand all of the nuanced signal-processing challenges that can arise when many loads are operated simultaneously in the same building. Significant advances have recently addressed these issues in residential settings, and several large open-source data sets have become available. These include:

- **Reference Energy Disaggregation Data Set (REDD):** REDD is collected by Kolter and Johnson [8] from 6 households in Massachusetts, U.S. The measurements are done on individual circuits in each household and also on the main meter. The mains are recorded at 15 *KHz*, and lower frequency data (3-second intervals) is available for the rest of the circuits in each house. The

list of appliances in each house is different. A few examples include lighting, refrigerator, dishwasher, furnace, washer dryer, kitchen outlets, and microwave. REDD dataset has been widely used for evaluation purposes in NILM. This includes most of the HMM approaches discussed in this thesis. Therefore, we chose to use this algorithm as a benchmark for performance analysis of our proposed work against the state of the art NILM algorithms. It should be noted that REDD is a limited dataset both from a geographical perspective and the duration of data collected.

Figure 2.4 depicts a daily example of aggregate power and individual circuits in one house of REDD.

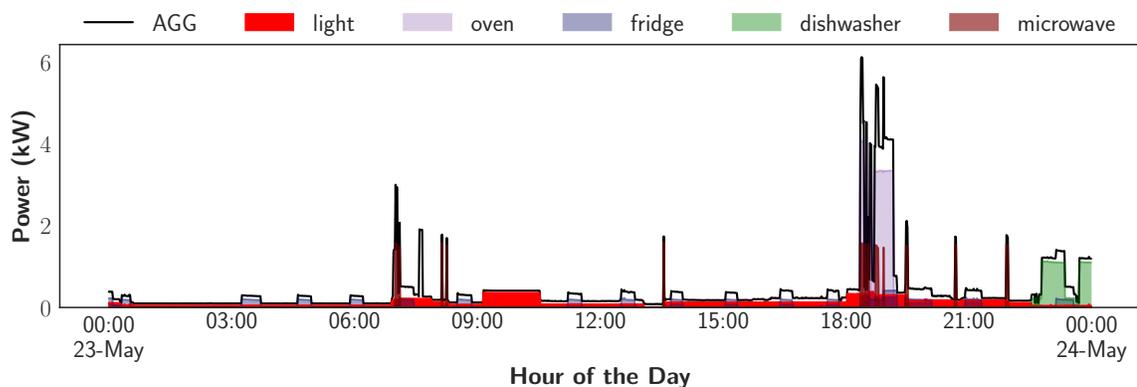


Figure 2.4: Daily example of aggregate power and its individual constituents for one house in REDD.

- **Building-Level Fully Labeled Electricity Disaggregation Data Set (BLUED):**

Anderson and his group at Carnegie Mellon University [10] have collected this dataset containing both voltage and current measurements for a house in Pittsburgh, U.S. This data is sampled at 12 kHz . The period of collection is minimal at one week but contains event information and can be potentially useful for signature extraction or creating a repository of load characteristics. We have not used this dataset in our analysis since we find it out of the scope of the aims of our research.

- **Tracebase Repository:** Is a very unique dataset from Darmstadt University in Germany[44]. It contains a growing dataset of individual appliances that are collected for characterizing appliances for identification purposes. The team claims to have collected data from more than 100 different devices from 43 different appliances types ranging from a refrigerator, computer, dishwasher all the way to play station, and Ethernet switches. It does not contain any whole-house meter information, and the houses from which the appliances are collected are unspecified. Due to these reasons, this dataset is not suitable for validation purposes of this thesis, but it provides a very good platform to create a load model repository.
- **The Almanac of Minutely Power Data Set (AMPds):** Makonin[9] and his team at Simon Fraser University have released a whole year of minute data containing the power measurements from a single house in Vancouver, Canada. They have collected data that includes measurements for 19 individual appliances, e.g., plugs lights, washer/dryer, fridge, HVAC, furnace, heat pump, oven, etc. and one whole-house meter. This dataset is unique in that it also provides gas and water measurement, something that is not present in other available datasets. Since we compare the results of our algorithm to SparseNILM [11], also proposed by the same team, we find it beneficial to use this dataset for validation purposes. However, it should be noted that we find this dataset to be very limited in that it is only collected from a single house, and even considering residential datasets, it has very little load action compared to other available datasets.

Figure 2.5 depicts a daily example of aggregate power and individual components power in AMPds.

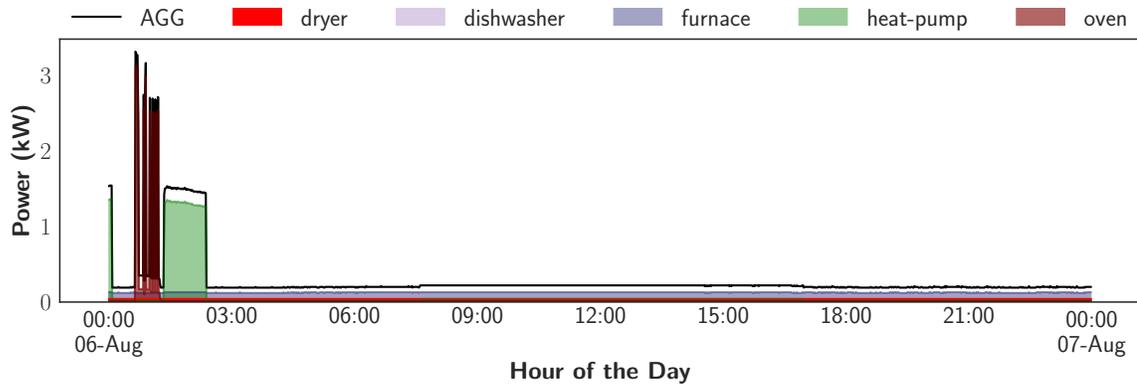


Figure 2.5: Daily example of aggregate power and its individual constituents for AMPds.

- Pecan Street project:** This project started at a limited scope back in 2013 but has expanded massively over the years. As of 2018, Pecan Street Inc. dataset [45] contains one year of 1Hz data collected through its volunteer residential research network. The dataset includes measurements from 40 homes for whole-home electricity use, solar generation, electric vehicle charging, HVAC, major appliances, and other in-home circuits. We find this dataset to be hugely beneficial for residential NILM research going forward, but since most of the available HMM approaches have validated their results on REDD, we chose not to include this dataset in the performance analysis of our work but would like to acknowledge the ongoing efforts in bench-marking this dataset [46].

Figure 2.6 depicts a daily example of aggregate power and individual components power in Pecan Street data.

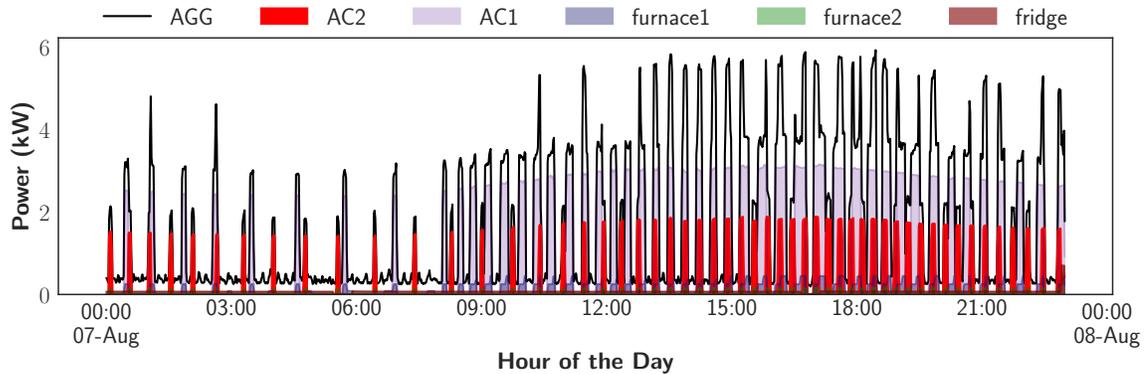


Figure 2.6: Daily example of aggregate power and its individual constituents for Pecan Street data.

The commercial sector has received far less attention from the NILM community than residential buildings, and the primary reason has been the lack of sufficient data [47]. The authors have recently sought to address this issue for small retail facilities. To do so, we have worked with some of the largest banks in the United States and have started to monitor a number of small branches. In Chapters 4 and 5, we describe critical characteristics of power usage in such environments by providing a comprehensive dataset collected from several commercial buildings. We also describe a methodology for tracking the usage of major loads.

2.4 Summary

In this chapter, we discussed the merits of energy monitoring in buildings and introduced a history of different monitoring approaches. We elaborated on the characteristics of building energy data in multi-time resolution for commercial buildings and discussed various types of information that can be leveraged in these scales. As nonintrusive load monitoring is the subject of this thesis, we provided a review of the several approaches to NILM at different levels of data granularity. Finally, we presented a list of publicly available datasets for NILM and made a case for using a couple, for evaluation purposes, that are more suited to the scope of this thesis.

CHAPTER 3: MOTIVATION

Despite approximately 30 years of work by NILM researchers, the technique has achieved relatively limited commercial success. One main reason has to do with the fact that researchers have not had access to datasets sufficiently large enough to allow them to understand all of the nuanced signal-processing challenges that can arise when many loads are operated simultaneously in the same building. In this chapter, we try to identify such challenges with a focus on commercial building power consumption data, an area in NILM that has been largely overlooked.

The research problem addressed in this thesis is related to how the aggregate power consumed in a single building and read by an individual meter can be explained by the individual constituent units inside the building. This problem was mathematically formulated by Hart [1] in 1992 as:

$$P(t) = \sum_{i=1}^n S_i(t)A_i + e(t) \quad (3.1)$$

where $P(t)$ is power observed at the main meter at time t , and $e(t)$ represents for measurement noise. If all possible n appliances are known (and this in itself is already a stretched assumption), minimizing $|e(t)|$ for n -dimensional Boolean *state* vector S_i results in a good estimate of each appliance state (ON/OFF). This can be formulated as the following combinatorial optimization problem:

$$\hat{S}_t = \arg \min_S |P(t) - \sum_{i=1}^n S_i A_i| \quad (3.2)$$

As pointed out by Hart in his original paper, this is an NP-Complete *weighted set* problem and computationally intractable, and therefore optimal solution can be

only achieved for scenarios where the aggregate is explained by a very small set of constituents.

Nevertheless, computational complexity is not the only issue of disaggregation. The nature of appliances in residential and commercial buildings are such that the complete set of power consumers, P_i , for every building is never known, or even if they are measured and known at the beginning of the process, appliances are replaced on a regular basis, and some tend to have a seasonal usage. This means that if one were to learn and model the appliances in some arbitrary training phase, the underlying condition can change significantly to the point that these models are no longer a valid representation of P_i . Furthermore, even in a hypothetical case that the perfect knowledge of all the possible appliances and their apparent power are known, there remains the very real possibility that one appliance can be described as the sum of some other. To quote Hart:

A more subtle problem is that the nature of the solution that 3.2 provides can be very inappropriate to the problem as a small change in the measured $P(t)$ would often be analyzed as a big change in the switch process, $a(t)$, with a number of appliances turning on or off simultaneously in such a way that the net change in approximates the observed change as well as possible.

Considering the current levels of data granularity provided by the smart meters, it is virtually impossible to find an optimal solution to 3.2 such that it satisfies an exact match to the ground truth. Nevertheless, there are ways to introduce certain regularization terms with good physical intuitions that would make solving a subset of the problem a realistic challenge. Furthermore, whenever exact inference is not intractable, approximation methods with lower bounds can be introduced for a faster inference that is accurate enough. The rest of this chapter discusses the standard NILM, explains inference in hidden Markov models, which has been the most popular

choice for unsupervised NILM, and finally explores the shortcomings in supervised training and inference with HMMs.

3.1 Setting the Standards for NILM

With the above considerations, we have identified the following rules that need to be followed by any successful NILM algorithm (partially inspired by [24]):

- **Non-intrusive:** The algorithm can use only a single point of measurement, and there is no additional information provided from other sensors or any retrofitting efforts required.
- **Unsupervised:** Training using meta-data is prohibited. This fact is dictated by the first assumption as the collection of meta-data involves using side information.
- **Compatibility with smart meter data:** The algorithm must be compatible with lower frequency datasets (Sampled at 1 minute and above). While a modern smart meter initially reads data at a higher frequency, the commercial models all report and store data at 1 minute and above. It is, therefore, crucial that the algorithm is tailored to such data granularity for commercial success.
- **Computational tractability:** Convergence should happen in a *reasonable* time. The definition of reasonable can, of course, be changed based on application. If, for example, real-time processing via a local microcontroller is desired for NILM, the algorithm should account for the limited computational capabilities of such controllers. On the other hand, offline scenarios would have more room for complex algorithms trained on larger portions of data collected from one or several buildings.
- **Robustness to unobserved/varying loads:** The algorithm must be able to handle unseen or un-modeled loads that are initially not present in the training

phase. If, for example, an appliance is added at a later stage in the process or a variation occurs in power demand due to seasonality, the NILM should either be calibrated to account for these changes or decouple such effects from the disaggregation process.

- **Accurate appliance power estimation:** The algorithm must be capable of tracking the power consumption of individual appliances with the same level of granularity as the main meter. This task is considered to be the ultimate goal of a NILM process.

Now that we have defined a set of rules and standards for NILM, we explore whether or not the current algorithms, specifically those based on HMM, abide by them. To this end, we first explain learning and inference in HMMs.

3.2 Hidden Markov Models

In the paper [48] published following the original NILM paper, Hart introduced the foundations of modeling NILM as a hidden Markov model. Ever since then, researchers have often used HMMs in modeling appliance power usage, assuming steady state power levels reminiscent of the finite state machine model proposed in the seminal paper of Hart. In modeling loads using HMMs, researchers often apply learning and inference advances in speech recognition to NILM. This is a reasonable approach since both problems can be defined under the bigger umbrella of structural learning problems. It is also with this view that some of the available literature has treated NILM as a handy example in demonstrating the applications of certain theoretical achievements in state inference of HMMs (e.g., [41, 40]).

A hidden Markov model is a Markov model, in that each variable only depends on the previous preceding variable (in time.) The term hidden refers to the variables that are not directly observed and inferring their state is of desire. Figure 3.1 depicts the structure of an HMM.

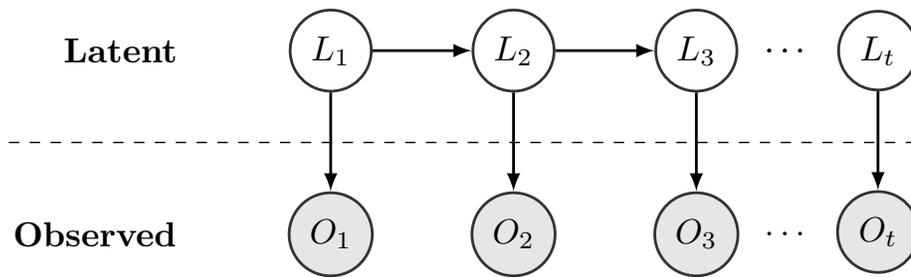


Figure 3.1: Graphical representation of a hidden Markov model.

A Hidden can be characterized using the following variables and parameters:

- Hidden (latent) variables: the finite set $L = \{l_1, l_2, \dots, l_m\}$.
- Observed variables: the finite set $O = \{o_1, o_2, \dots, o_n\}$.
- Transition matrix: $\mathbf{A} = \{a_{ij}, 1 \leq i, j \leq N\}$ representing the probability of moving between states from one time slice to another, $a_{ij} = P(L_t = j \mid L_{t-1} = i)$.
- Emission matrix: $\mathbf{B} = \{b_j(k)\}$ are assumed to be Gaussian and are the probability of emission for observation when the system is at a certain state, $b_j(k) = p(o_t = k \mid L_t = j)$.
- Initial state probability distribution: the probability of states at time $t = 1$, $\pi_i = P(L_1 = i)$.

The model parameters can be represented as $\Theta = (\pi, A, B)$. Our main goal is to find the most probable sequence of states given the model parameters Θ and observation sequence O . This should itself be broken down to two problems: 1. learning model parameters and 2. Inference using the learnt parameters.

Learning the model parameters that best describes the model can be defined by maximizing the likelihood $P(O \mid \Theta)$:

$$P(O | \Theta) = \sum_S P(O, L | \Theta) \quad (3.3)$$

where

$$P(O, L | \Theta) = P(O | L, \Theta)P(L | \Theta) \quad (3.4)$$

Given the states, observations are independent from one another and joint probability can be written as:

$$P(O | \Theta) = \pi_{l_1} P(o_1 | l_1) \sum \prod_{t=2}^T a_{ij} P(o_t | l_t) \quad (3.5)$$

This problem can be solved using the Baum-Welch algorithm [49], which is an Expectation-Maximization (EM) algorithm. We discuss the state of the art of this process and how it relates to the problem of supervision in the next section. The problem of learning can be broken down further into two problems. One is to find $P(O | \Theta)$ with the model parameters. Given this likelihood, find the most probable sequence for states L that explains the observations O . This requires the calculation of $P(O | L, \Theta)$ over all the possible states. If one is to do the math, there are $2Tm^n$ possibilities to be considered for this likelihood. However, there is a faster solution to the problem of exact inference using Dynamic programming. Viterbi [23] is one such algorithm. By considering the most likely path to a state and discarding the less likely path in a transition from time t to $t + 1$, the calculation is reduced to $T * m^2$. Again this would be ineffective as the number of the states grows. This fact has resulted in a number of approximation methods used for inference in HMMs.

Gibbs sampling [50] has been shown to be an effective approach in approximating

complex densities. It works with an acceptance-rejection mechanism. Starting with an initial guess of distribution parameters, a random walk of samples in parameter space is generated with the following criteria: if the sample increases the likelihood, it is always accepted, and if not, it is accepted with some degree of randomness. It has been shown that these chains of random walks can eventually converge to the underlying distribution with enough number of iterations [51], but the chains can potentially grow to be in the range of tens of thousands of iterations and, therefore, while the Monte Carlo methods are fairly simple to implement, they bring some computational cost along with them.

A faster alternative is Variational Bayes methods [52]. They work by finding a lower bound of the marginal likelihood by minimizing a probability distance measure (usually KL divergence) for a tractable auxiliary distribution. This process involves an assumption of some level factorization in the structure of HMM.

3.2.1 Factorial hidden Markov Model

Figure 3.2 depicts the structure of a factorial hidden Markov model (FHMM) as proposed by [53]. This particular HMM maps the observations to a hidden state such that a single multinomial random variable is represented by a collection of state variables from other HMMs. For each observation in Equation 3.2, the most probable corresponding hidden states can be calculated as [53]:

$$P(P_t|S_t) = \det(C)^{-1/2} (2\pi)^{-D/2} \exp \left\{ -\frac{1}{2} (P_t - \mu_t)' C^{-1} (P_t - \mu_t) \right\} \quad (3.6)$$

where

$$\mu_t = \sum_1^M W^{(m)} S_t^{(m)} \quad (3.7)$$

and

$$\hat{a}_t = \arg \max_a P(P_t|S_t) \quad (3.8)$$

Here, W is a matrix with binary elements explaining the contributions of states to μ_t . D is the length of observation P_t which itself is a Gaussian random vector. K is the number of possible modes for each state variable. Since there are K modes and M different state variables, there are K^M possibilities to be considered for each observation and since the mean of each observation sequence (μ_t) will depend on all the possible states, calculating the posterior probabilities in 3.5 becomes intractable for FHMM. Approximation methods for such models with good lower bounds have been introduced in several works, including [40, 53]. For a detailed review of these approximation methods, including variational Bayes methods, the readers are directed to Jordan [52].

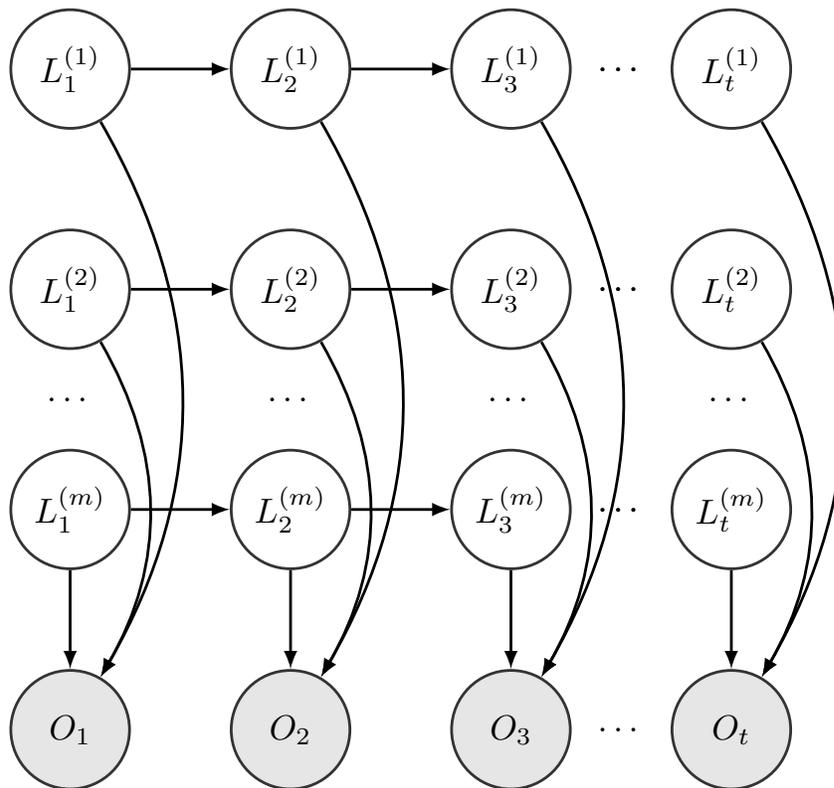


Figure 3.2: Graphical representation of a factorial hidden Markov model.

3.2.2 Challenges of NILM with hidden Markov models

Unsupervised NILM research has, by and large, been dominated by deployment of temporal graphical models in a probabilistic framework ([39], [40]). This has been partially driven by the fact that HMMs are very useful models for unsupervised practices and can also account for modeling of more complex types of behavior through their numerous extensions.

Table 3.1 shows the time complexity for the inference of some of the state of the art NILM algorithms using HMMs. It is clear that the computational cost quickly becomes intractable for large values of M . The use of approximate methods provide more tractable inference methods, an example of such methods is Markov Chain Monte Carlo algorithms or structural mean-field methods [53]. These methods are

highly susceptible to local optima in the likelihood function, which can significantly degrade the performance of the approximate method. Two good examples of such cases are inferring the power of a small load in the presence of some noise from significantly bigger loads or a case in which the presence of multiple loads is not distinguishable from a load that equals the sum of them.

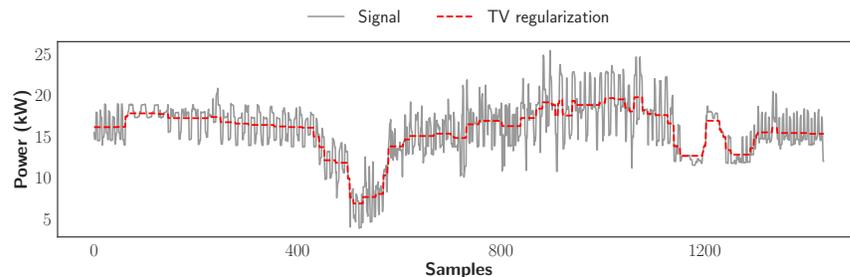
It is, therefore, necessary to introduce some regulatory constraints to reduce the computational cost and avoid the many local optima. For example, as pointed in the original work of Hart, it is a fair assumption to say that in small enough intervals, appliances very rarely change their states once and very rarely more than once. This can be interpreted as another way to say that the signal is piece-wise constant. To this end, Kolter and Jaakkola [40] impose a condition upon the posterior distribution such that at most, one HMM changes state at any given time. They also introduce a difference HMM (DHMM) to ensure that step changes are credited to the corresponding appliance power changes. To address the robustness of the algorithm to previously unseen data, Kolter introduces a generic mixture component to take on piece-wise constant values, which captures the nature of un-modeled loads. The prior for this signal $z_t \in \mathbb{R}^n$ is defined as:

$$P(z_{1:T}) = \frac{1}{Z(\lambda, T)} \exp \left\{ -\lambda \sum_{t=1}^{T-1} \|z_{t+1} - z_{t-1}\| \right\} \quad (3.9)$$

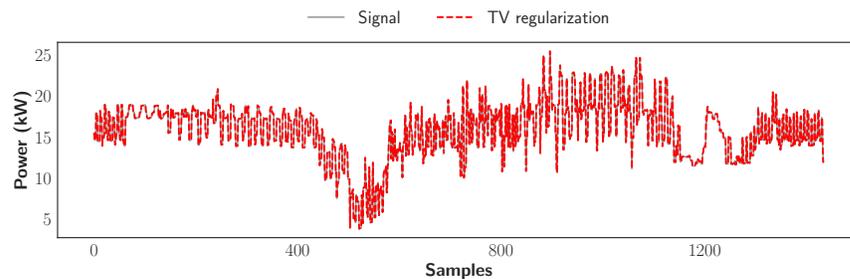
Table 3.1: Time complexity for a single observation sequence of length T for M hidden Markov models, each consisting of K states.

Model	Time Complexity
Exact Inference FHMM	$O(TMK^{M+1})$
Exact Inference for Super-State HMM[11]	$O(TK^{2M})$
Completely factorized variational inference [53]	$O(TMK^2)$
Approximate Inference Gibbs Sampling [54, 55, 41]	$O(TMK)$

It is important to note that while this can be considered a good smoothing constraint on datasets with 1 Hz and above frequencies, it would, however, not have the same desired effect on once per minute sampled data. Here, λ should be large enough so that the model does not assign all the output to the state variables but small enough so that $z_{1:T}$ is not a constant-valued). Figure 3.3 demonstrates an example of using such a constraint on once per minute data. It is clear that one way or another, the regularization is going to result in under or over smoothing without any significant gain in the goal of modeling the unobserved loads.



(a)



(b)

Figure 3.3: An example of TV Regularization with (a) $\lambda = 1000$ (over smoothing) and (b) $\lambda = 0.1$ (under smoothing).

For a better illustration, let us consider all the sum of un-observed loads in any given window of length T as a non-stationary signal of the same length:

$$x_{1:T} = \{x_1, x_2, \dots, x_t\}$$

where

$$x_t = x_{t-1} + \epsilon_t \quad \text{for } t \in 1, 2, \dots, T \quad (3.10)$$

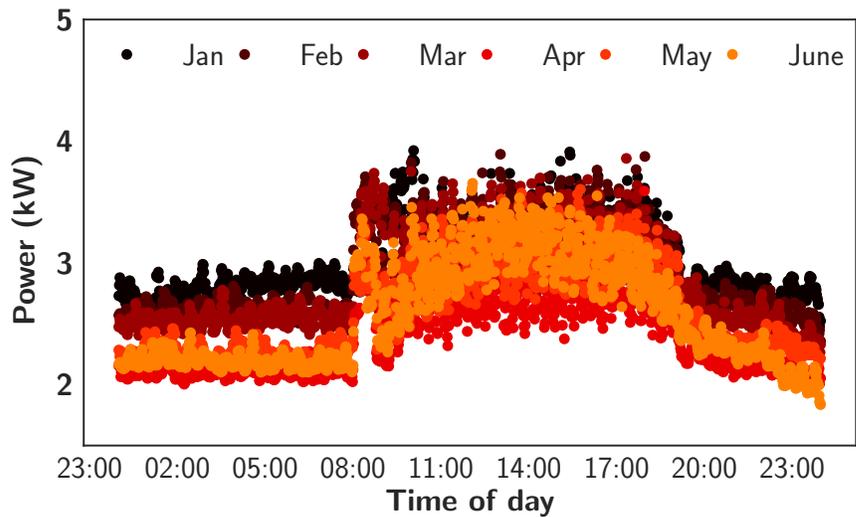
Where ϵ is an uncorrelated white-noise process (for all times t). From here on in this document, we refer to this signal as the baseline. Figure 3.4 is a depiction of such a baseline for a typical summer day in a commercial building. We believe that, for the most part, studies in NILM have neglected the presence of such a baseline. This is partially due to the syntactic nature of available datasets for NILM. For example, the proposed algorithm in Kolter [40] reports near-perfect performance of the proposed NILM algorithm on synthetic data while the same can not be said of the tests carried out on the REDD dataset.

Parson [24], on the other hand, suggests a detrending via a min extraction process. While this might be suited to some residential cases, it is visually visible that such an approach would not generalize to other datasets and domains of buildings.¹

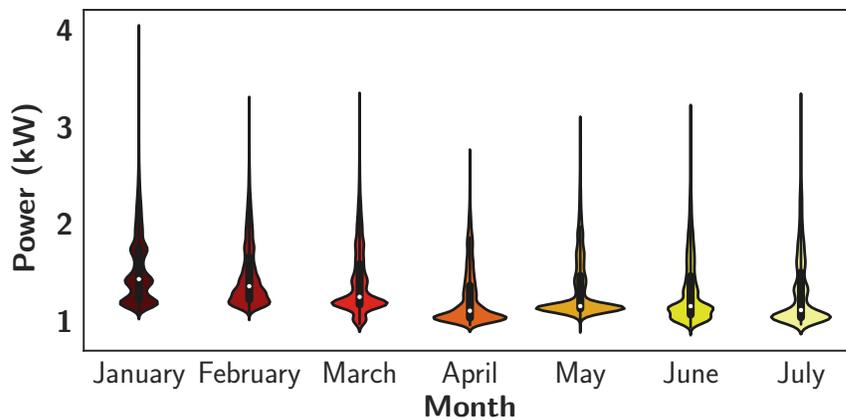
In one particular study that acknowledges the need for consideration of such phenomena, Makonin [11] proposes that all the un-metered data be treated as a modeled HMM. This could introduce further computational cost to the algorithm with no immediate gain in performance in cases where the upper boundary of the number of the state is not defined correctly. It is also worth pointing out that the un-metered load statistical characteristics, which the HMMs happens to be based on, can drastically change over time with the introduction of new loads in a building or simply by the seasonal nature of certain operations. Figure 3.4 depicts such a scenario on a sample

¹We choose a similar idea but generalize it to a robust background filtering algorithm in the next Chapter.

data set from a small size commercial building.



(a)



(b)

Figure 3.4: A statistical demonstration of non-stationary baseline.

Finally, even if all of the above issues are to be addressed, the assumption simply implies a need for detailed monitoring of the circuits in some training period. This fact alone violates the unsupervised nature of the NILM algorithm.

3.2.3 On the issue of supervision

In most of the cases above, along with other research carried out using HMM as the basis of disaggregation, including the more recent examples [11, 56], prior models for

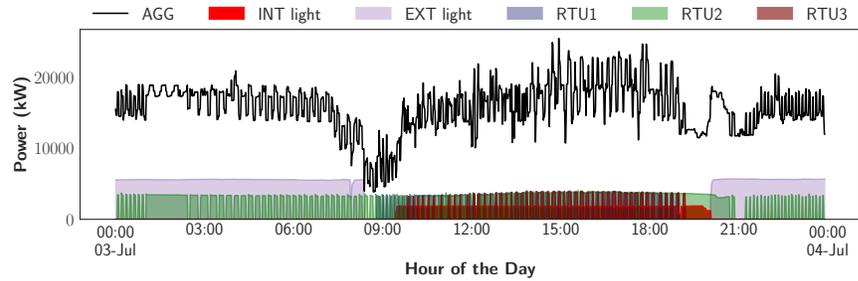
each appliance are required to be learned from data. Learning these models would, at the very least, demands a knowledge of the number and types of appliances present in each household and in some extreme cases, detailed sub-metered data for each appliance [11]. A comprehensive comparison of some of the more recent NILM approaches in [57] concludes that to achieve reasonable performance in NILM algorithm, some level of supervision is required. This process violates the unsupervised requirement for NILM.

To address this issue, Parson [24] proposes an approach that tunes a prior appliance model to a period in the aggregate data that only contains a single load. This fact is achieved by calculating the likelihood that a chunk of observed aggregate data is best explained by a single load.

In general, as pointed by [57] it is rare to find such periods in which only a single appliance is running. For example Figure 3.5 depicts one such case for a medium-size commercial building.

Others have suggested a clustering approach based on a power signature. These signatures can come from active and reactive power [1, 58]. As for clustering, in one successful example, Baranski [59] proposes that events be filtered using a thresholding mechanism, and the extracted events clustered using fuzzy clustering. Events in the same cluster are then assumed to be generated from an individual appliance power profile; however, the optimal number of clusters in this scenario is not identified. It is also not clear how the clusters can be appropriately tagged to represent their corresponding power consumers.

In the next chapter, we propose an algorithm that is capable of unsupervised modeling of loads through a novel edge clustering mechanism.



(a)

Figure 3.5: An example of aggregate power consumption, along with individual load constituents, for a typical weekday in a commercial building.

3.3 Summary

In this chapter, we discussed the required standards for NILM. We gave a review of hidden Markov models, including learning and inference using such models, and discussed the rich history of HMMs and its variations applied to the problem of nonintrusive load monitoring. We further discussed challenges in NILM, namely computational complexity, supervision, and sensitivity to un-observed/varying data. What follows in the next chapter is the general proposal in this thesis in addressing the challenges described in this chapter.

CHAPTER 4: A ROBUST NILM ALGORITHM

In the last chapter, we outlined the limitations of different HMM-based approaches for nonintrusive load monitoring. In this chapter, we propose a novel algorithm capable of doing *robust NILM*, by which we refer to a methodology that is capable of monitoring the power demand of major consumers of energy for a large portfolio of buildings. To this end, we break down the proposed approach into three major steps:

- **Background filtering:** In which we extract the background of the signal using an edge-preserving algorithm to dampen the effect of unknown/undesired loads on the inference stage. This step can be broken down into a change of mean detection and detrending.¹
- **Edge analysis and load modeling:** in which we detect the edges above a certain threshold from the aggregate data along with their corresponding time stamps and cluster the pairs of positive and negative edges based on their absolute size and order of appearance. From these clusters, we model the hidden states assuming a Gaussian distribution.
- **State inference and power estimation:** in which we disaggregate the signal by generating all the valid combinations of Gaussian clusters and infer the states by training a naïve Gaussian Bayes classifier. We estimate the power of individual loads from these inferred states and their corresponding means.

The general framework for the proposed NILM is depicted in Figure 4.1:

¹It is important to note that the change of mean detection step is mainly applicable to the dataset that we have been working with, i.e., medium-sized commercial buildings.

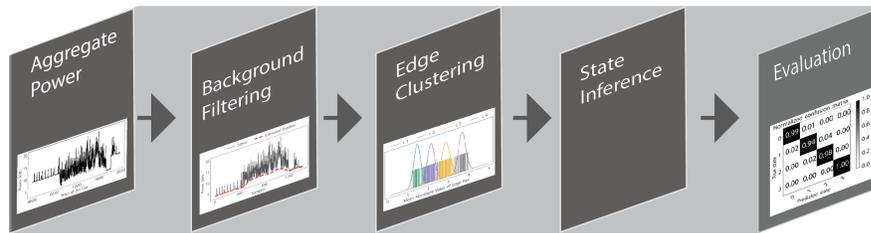


Figure 4.1: The diagram of the proposed algorithm.

Before we get to a detailed explanation of each step, it is important to note that since we are dealing with 1-minute data, we neglect the importance of making the signal piecewise constant, which has been discussed in several other NILM publications, including [40]. We firmly believe that while this is an essential step in datasets sampled at a higher frequency ($1Hz$ and above), it has virtually no effect on the performance of the proposed NILM, and it only presents a burden on the computational cost of the algorithm.

4.1 Background filtering

We divide the process of background filtering into two steps: 1. Model based change of mean detection and 2. Baseline removal. The first stage of this process is more tailored toward the commercial datasets, specifically those collected from building with large exterior lights. The major reason behind framing the process in this manner is the effect of dominant edges produced by the distinct power consumption pattern of these exterior lights in the performance of the baseline estimation algorithm (The latter algorithm guarantees edge preservation). A lack of such appliances in a building does not affect the generality of our approach, as no change of mean would be detected in the first step, and the signal is passed to the next stage untouched.

4.1.1 Model based change of mean detection

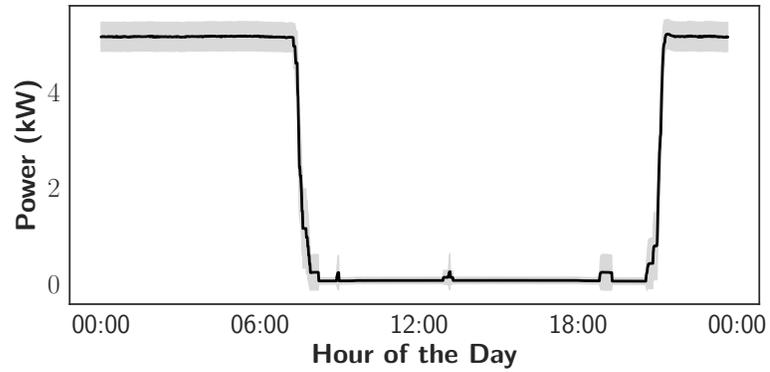
It can be assumed that by and large, the majority of commercial buildings have completely different modes of operation during night time and day time. This phenomenon is caused by two major factors: The presence of exterior lights (traditionally halogen lights) during the night hours and the 8-hour workday and all the activities that it generates by keeping the building occupied in a certain level that is not reached otherwise (this period also happen to have a high correlation with daylight). Figure 4.2 displays the exterior light uncertainty during May 2014 in a mid-sized commercial building along with an example of aggregate power during a workday in May for the same building. It is immediately noticeable that the lights are one of the largest constituents of power consumption in the building. It is also evident from the figure that they can be assumed as two-state loads. Therefore, tracking the state of the exterior light can be formulated as conducting a search to find points at which the mean of signal changes most significantly.²

Let x be a signal, of length n , sampled at equal intervals. We define $1 < i < n$ to be a change point in x , if it divides p into two segments with different and independent probability distributions. These probability distributions are to be interpreted given some model M . We accept a change point if the new probability is higher than that of a single model. This can be written as:

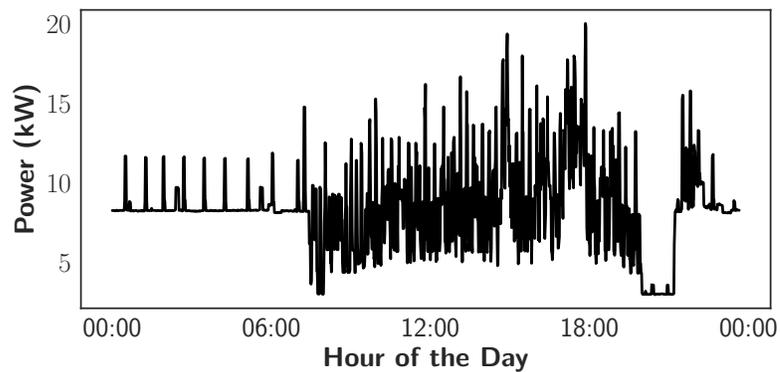
$$\log (P(M_1|x_{1:i})) + \log (P(M_2|x_{i+1:n})) - \log (P(M|x_{1:n})) > 0 \quad (4.1)$$

Where $P(M|p)$ is the likelihood of the model given the data. From here on we

²It can be rightly argued that there might be several change points in a signal that do not necessarily correspond to the exterior lights. In such cases, and for robustness to such similar scenarios, one can always conduct the search for the first and last most significant change points in case there are more than two of them present in the signal.



(a)



(b)

Figure 4.2: An example of the Exterior light uncertainty along with the aggregate power in a mid-size commercial building.

simply refer to a two-segment model as M_t , one segment model as M_o and data as D . The above equation can be rewritten as:

$$\frac{P(M_t|D)}{P(M_o|D)} = \frac{P(D|M_t)P(M_t)}{P(D|M_o)P(M_o)} \quad (4.2)$$

Where $P(D|M)$, the marginal likelihood for the model. This likelihood is calculated with an integration over the model parameters θ :

$$P(D|M) = \int P(D|\theta, M)P(\theta|M)d\theta \quad (4.3)$$

Assuming a Gaussian process, the data should follow a Gaussian distribution of the form $N(\mu, \sigma)$ and we can calculate the marginal likelihood as:

$$p(D|M) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2\right) \quad (4.4)$$

where $D = \{x_1, x_2, \dots, x_n\}$. Since we can claim to have no knowledge of model structure, the priors of the models $P(M)$ can be assumed to be equal, and with this assumption, the task of finding the changepoint in the segment becomes:

$$\arg \max_i \log odds = \ln \frac{P(D|M_t)}{P(D|M_o)} \quad \text{if } \log odds > thr \quad (4.5)$$

where i is the changepoint. With the assumption that the underlying distributions are Gaussian, the above equation can be further simplified as:

$$\arg \max_i |\bar{x}(i) - \mu| \quad \text{if } |\bar{x}(i) - \mu| > thr \frac{\sigma}{\sqrt{n}} \quad (4.6)$$

where

$$\bar{x}(i) = \frac{1}{n} \sum_{k=n(i-1)+1}^{i \times n} x_i \quad (4.7)$$

It is important to note that if the value of the threshold is chosen to be too small, this can result in too many undesired changepoints, and if it is too large, we can easily miss the changepoints. In practice and particularly in this example, we use prior information to assist the speed and accuracy of our algorithm. First, we fix a maximum number of changepoints (4 in this case). Second, we favor longer distances between the changepoints, i.e., two changepoint that are closer than a certain distance to each other are not considered simultaneously. Third, the length of each partition should be more than a certain threshold. So far, we have talked about how to decide on accepting or rejecting a changepoint. We propose a divide and conquer algorithm to cut the partitions to the smallest possible predefined partitions of a certain length and merge them according to the equation 4.6 until no further merging is possible. If merging is not possible, we search for a change point in the neighborhood such that it maximizes equation 4.6. Once the partitioning is done, we prune the partitions according to our prior conditions. It is important to note that this approach doesn't guarantee optimal number of change points, but we can guarantee that it would always work in the case of finding the first and last one if there exist such changepoints.

Finally, the author would like to openly acknowledge the fact that while more sophisticated changepoint algorithms exist [60, 61], this fast and straightforward approach is justified by the fact that it guarantees that the task at hand is carried out with relative simplicity.

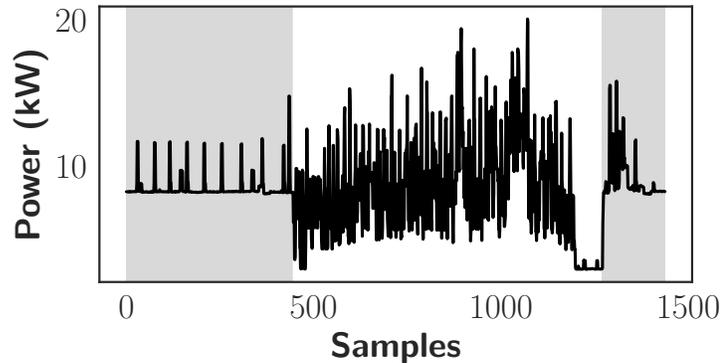


Figure 4.3: An example of the output of the change of mean detection algorithm, highlighted by the grey area.

4.1.2 Baseline removal

In the previous chapter, we discussed the importance of removing the baseline for the sake of the robustness of the approach. To clarify, what we mean by baseline in any given signal is the *gradual* variation in the local mean of the signal within a small neighborhood. Gradual here has a specific meaning in a low-frequency signal: It refers to variation due to *small* steps in succeeding points in a signal. From this definition, it is clear that what we are trying to achieve by this baseline elimination is merely removing the small unknown/miscellaneous loads that, if present in the signal, will significantly affect the task of inference in disaggregation. Let P be a signal, of length t , sampled at equal intervals. B , a signal of the same length, is the baseline to P if it has the following properties: 1) B is smooth and 2) B is faithful to P . The smoothness (differentiability) guaranties the edge preservation in the process while faith fullness ensures that we are following the trend of P . The problem of finding the baseline can, therefore, be defined as minimizing the following term:

$$\sum_i w_i (P_i - z_i)^2 + \lambda \sum_i \Delta^2 z_i^2 \quad (4.8)$$

where, by introducing the first term, we are guaranteeing that we follow the original signal, and with the second term, we ensure the edge preservation by adding $\Delta^2 z_i^2 = (z_i - z_{i-1}) - (z_{i-1} - z_{i-2})$ and setting λ as a tuning parameter. w is a vector of weights for generalization. The choice of w should be such that it ensures the estimated baseline not only follows the trend of signal but also lies on the base of it. One such choice can be defined as [62]:

$$w_i = k \text{ for } p_i > z_i \text{ and } w_i = 1 - k \text{ otherwise} \quad (4.9)$$

where $0 < k < 1$. Solving for this minimization's leads to the following system of equations:

$$(W + \lambda D'D)z = Wy \quad (4.10)$$

where $W = \text{diag}(w)$ and D a difference matrix: $Dz = \Delta^2 z$. Eilers et al. [62] suggest a linear time convex solution based on Cholesky factorization, leveraging the sparsity of the system of equations for fast implementation. We have implemented the algorithm using Scipy package in Python.

One final note to make here is that it is important to note that λ and p should be tuned properly and according to application. In our case, trying the algorithm on several different datasets, we found out that $10^2 < \lambda < 10^3$ and $0.01 < p < 0.1$ works well for NILM datasets, but the readers are encouraged to do further investigation on this.

Figure 4.4 demonstrates an example of background filtering described in this sec-

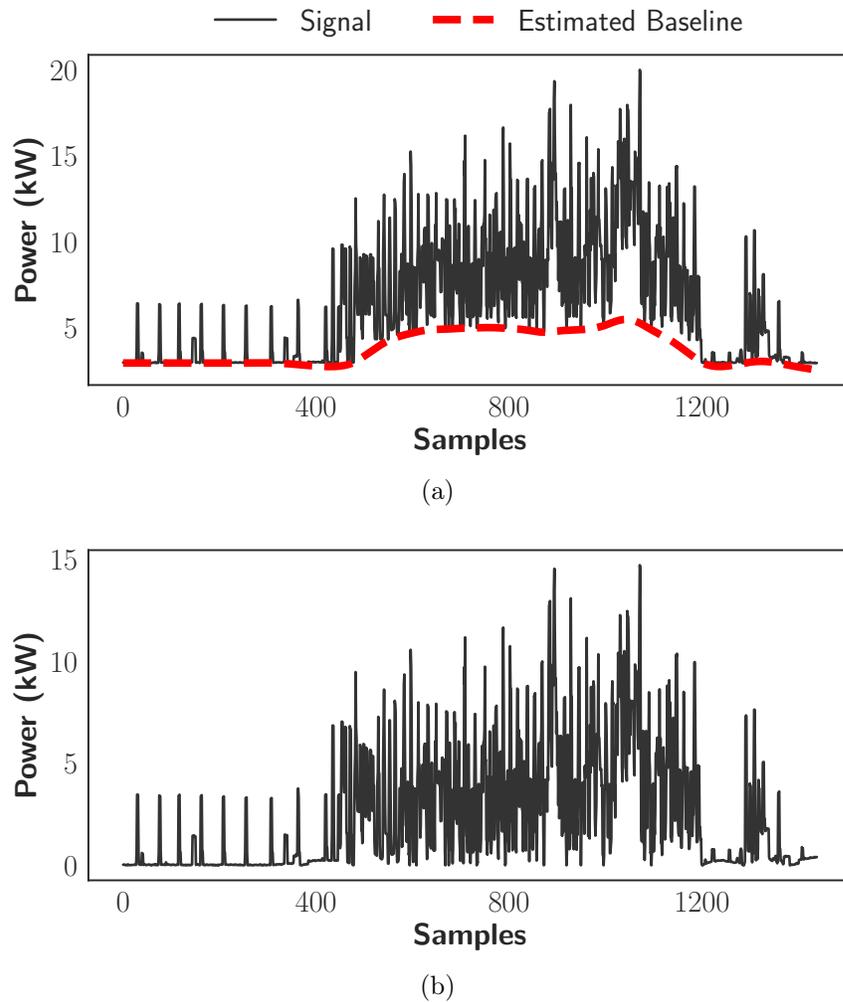


Figure 4.4: An example of (a) baseline estimation along with the (b) detrended signal.

tion.

4.2 Edge analysis and load modeling

As discussed in the previous chapters, we consider the use of smart meter data for training NILM disaggregator to be in violation of an unsupervised algorithm. While acknowledging the interesting work of Parson [24] in tuning the prior appliance models to a period in the aggregate data that only contains a single loads in small residential houses, we have found the task of finding such periods to be virtually impossible for more complex buildings (complex here refers to buildings with a lot of concurrent load activities). Instead, we focus on an approach based on an analysis of the edges

in signal and paring them into their corresponding source (loads). To do so, we first extract all the signatures with an absolute value above a threshold along with their respective time stamps. This threshold should be large enough to avoid unnecessary edges and small enough not to miss any important information. In practice, this depends on the frequency of sampling, level of noise and the desired loads to be disaggregated.³ From this set, using a novel approach, we compute all the possible pairs of positive/negative edges within a small threshold ϵ of each other that do not have a time stamp difference of more than a few hours with one general rule: the chronological order constitutes that a negative edge should always follow a positive one. It is important to note that in this step, we do not care about the faithfulness of the edges to their corresponding loads but we rather care about the faithfulness of the edge magnitudes to an underlying distribution. The result of this process is a set of pairs $(u, d) \in \mathbb{R} \times \mathbb{R}$ where u represents all the extracted on edges and d represent the potential corresponding off edges.

To optimally cluster this set, we map each point in this $2D$ set to $y_i \in \mathbb{R}$ by averaging the absolute value of the pairs and use CKmeans [63] clustering. CKmeans is specifically suited to this problem as it provides optimal, fast solution to the problem of 1D clustering. A brief explanation of the algorithm is as follows:

Let $Y = \{y_1, y_2, \dots, y_n\} \in \mathbb{R}$ be the set of extracted edges that we intend to cluster. The optimal clustering of this set for k clusters is equal to assigning each element in the set to a cluster such that the sum of squares to the mean of each cluster is minimized. If y_j is the smallest element cluster m in an optimal solution, then all the other points must also be optimal for $m+1$ clusters or otherwise, there is a better solution. We evaluate this using function D , within cluster sum of squares. By taking this as the optimal substructure property, we can establish the following

³For example in a case where the only desired load is HVAC with a corresponding edge of 3400 watts, it is self-evident that extract 100 watt edges would serve no purpose as it can simply be considered as the noise level of the mentioned load.

Bellman equation:

$$D[i, m] = \min_{m \leq j \leq i} \left\{ D[j, m] + d(y_j, y_i) \right\}$$

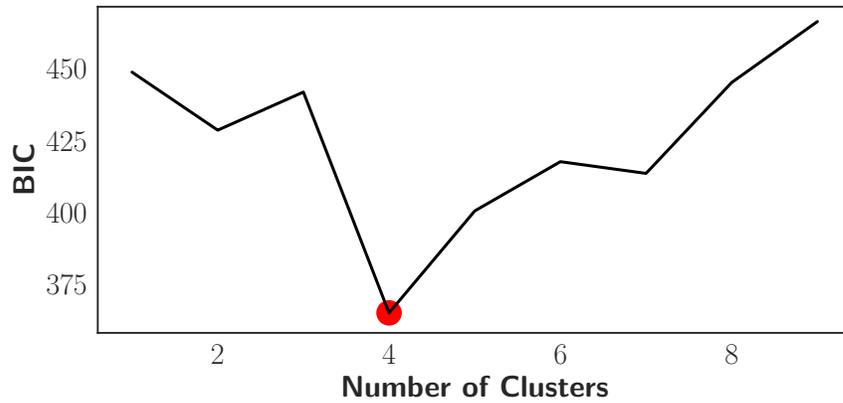
$$1 \leq i \leq n, 1 \leq m \leq k \quad (4.11)$$

where $d(y_j, y_i) = \sum_{k=j}^i \left(y_j - \frac{1}{N} \sum_{l=j}^i y_l \right)^2$. For further information on this algorithm, the readers are directed to [63].

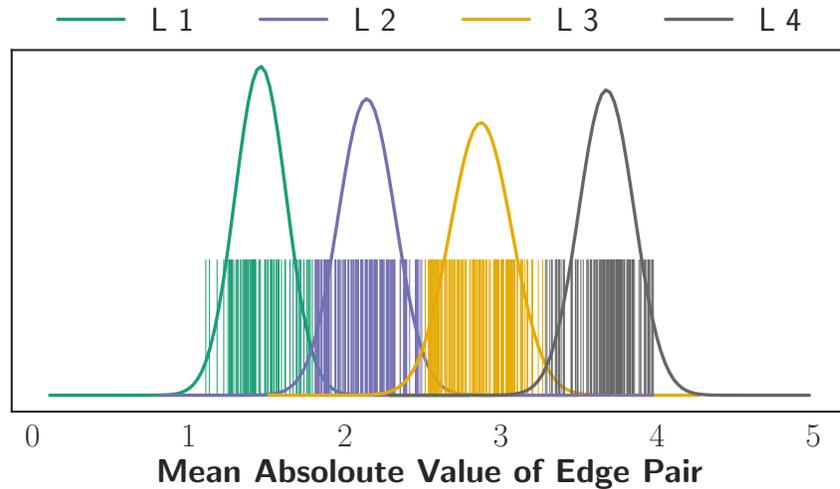
Finally, it is important to note that, like any other clustering problem, the optimal number of clusters should be addressed properly. We use Bayesian information criterion (BIC) for this purpose:

$$BIC(K) = \sum_{p=1}^N \left(\frac{y_p - c_k}{\sigma_\epsilon} \right)^2 + 2K \log N \quad (4.12)$$

where σ_ϵ is the intraclass variance, N is the members count of a cluster and c_k represents the cluster centroids. We can think of K as a regularization parameter in minimizing BIC . The top picture in Figure 4.5 is an example of different values calculated against the number of clusters for edges extracted during weekday. Based on this figure, we choose four as the optimal number of clusters. The bottom figure demonstrates the output of optimal clustering along with a fitted Normal curve for each load.



(a)



(b)

Figure 4.5: An example of (a) Choosing the optimum number of clusters based on BIC and (b) the corresponding CKmeans clustering results.

To model the loads, we assume that each load follows binary states (loads with more than two states can be modeled as a combination of several two-state loads). Figure 4.6 shows an empirical distribution of power demand for a few loads in a building. As seen in the figures, the power demand of appliances roughly follow a Gaussian distribution $N(\mu, \sigma)$ (This is also supported by the literature [55]). Given this, we estimate the parameters of the Gaussian distributions given the observations

in each cluster using a MAP estimation.⁴

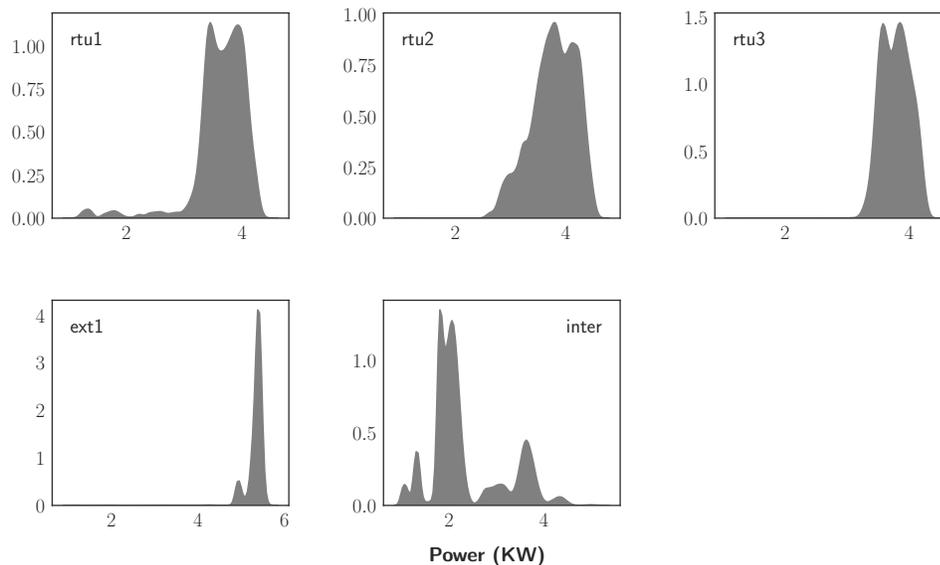


Figure 4.6: Empirical ON Distributions of Loads in an example building.

4.3 State inference and power estimation

The state inference problem has to do with figuring out the state of appliances (hidden variables) such that it explains the aggregate power demand (observations):

$$y_t = \sum_1^M W^{(m)} s_t^{(m)} \quad (4.13)$$

where W is a matrix with binary entries explaining the contributions states to y_t . Given the models described in the previous section, the problem can be formulated as explaining the mean of observations as a combination of Gaussian mixture models (univariate normal densities). Let us consider $S_t^m = \{s_t^1, s_t^2 \dots s_t^M\}$ as all the possible M different states at time t . The inference task for some observed sequence of data $Y = \{y_1, y_2, \dots y_n\}$ can be written as:

⁴closed form solutions for mean and variance in 1D.

$$p(s|y, \theta) = \frac{p(y, s|\theta)}{p(y)} = \frac{p(y|s)p(s)}{\int_{-\infty}^{\infty} p(y|s)p(s)d\theta} \quad (4.14)$$

where θ is the model parameters. Computing the integral (marginal likelihood) in this equation is generally very hard. To find the set of states that best explain an observation, a Viterbi path can be calculated as the optimal model parameters for which the likelihood equation is maximized. In other words:

$$\hat{\theta} = \arg \max_{\theta} P(s_i(\theta)|y) \quad (4.15)$$

In the case of a hidden Markov model, finding the Viterbi path requires an estimation of all the model parameters such as emission probabilities, transition probabilities, etc. The exact inference is intractable, and there are several approximate methods, e.g., Monte Carlo and Variational Bayes (VB), that can give good estimates of Viterbi path[52, 53, 64]. Alternatively, a dynamic programming approach for exact inference can be tried in cases where the set of model parameters is small [23]. All of the above, however, would require an EM training of HMMs on data, a procedure that violates the prohibiting condition of using metadata for training.

Alternatively, given the assumption of independent Gaussian mixtures with known parameters (μ, σ) , the integral can be analytically derived. First, the goal is defined as finding the set of weights W such that the following mean square term is minimized:

$$E[|y - \sum_{j=1}^M W_j s_j|^2] \quad (4.16)$$

where j represents every single Gaussian mixture. As discussed before, each Gaussian is considered as a uniform distribution representing a single ON state. If there exists a load with multiple ON states, it can be presented by a combination of the Gaussians. Considering the Gaussian mixtures, the probability that an observation belongs to a model can be defined using the Bayes rules as:

$$P(s_i|y) = \frac{\alpha_i N(y ; \mu_i, \sigma_i)}{\sum_{k=1}^M \alpha_k N(y ; \mu_k, \sigma_k)} \quad (4.17)$$

where the Normal distribution N is defined as:

$$N(y ; \mu_i, \sigma_i) = ((2\pi)^{1/2})\sigma^{-1/2} \exp(-(y - \mu)^2/2\sigma^2) \quad (4.18)$$

by substituting Equation 4.18 into Equation 4.17, it can be written as:

$$P(s_i|y) = \frac{\alpha_i \sigma_i^{-1/2} \exp(-(y - \mu_i)^2/2\sigma_i^2)}{\sum_{k=1}^M \alpha_k \sigma_k^{-1/2} \exp(-(y - \mu_k)^2/2\sigma_k^2)} \quad (4.19)$$

The Viterbi path can be calculated as:

$$\hat{\theta} = \arg \max_{\theta} P(s_i(\theta)|y) \quad (4.20)$$

$$\text{subject to } \sum_m \pi_m = 1 \quad (4.21)$$

where $\theta = (\mu, \sigma, \pi)$ and π is the probability of each cluster given the full data. This model is effectively a naïve gaussian Bayes classifier[65]. Figure 4.7 depicts a diagram of how such a model can be trained. Using some known probabilistic models with categorical tags, we generate a set of samples with the corresponding categories. Assuming the independence for parameters of θ we take the derivative with respect to each one to estimate the maximum likelihood in Equation 4.20:

$$\frac{\delta P(s_i(\theta)|y)}{\delta(\mu_i)} = 0, \quad \frac{\delta P(s_i(\theta)|y)}{\delta(\pi_i)} = 0, \quad \frac{\delta P(s_i(\theta)|y)}{\delta(\sigma_i)} = 0 \quad (4.22)$$

which results in the following set of equations:

$$\begin{aligned} \hat{\pi}_i &= \frac{n_i}{n_y} \\ \hat{\mu}_i &= \frac{1}{\sum_j \delta(S_j = s_i)} \sum_j y_j \delta(S_j = s_i) \\ \hat{\sigma}_i^2 &= \frac{1}{\sum_j \delta(S_j = s_i)} \sum_j (y_j - \hat{\mu}_i)^2 \delta(S_j = s_i) \end{aligned} \quad (4.23)$$

where n_i is the number of times class i is observed, n_y is the length of all the data. j is the index of training example and $\delta(True) = 1$ and 0 otherwise.

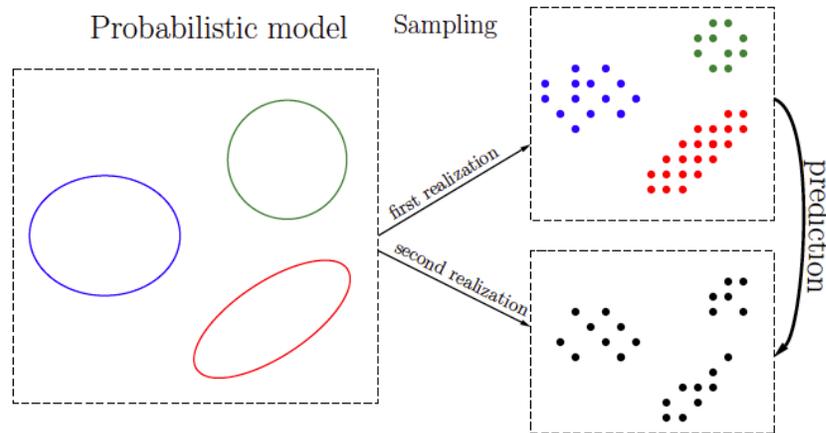


Figure 4.7: Training a naïve gaussian Bayes classifier.

In training the classifier, we have to account for all the possible load combinations that are otherwise not explained by a single cluster. Below we give a detailed explanation for how to calculate a set of all the valid combinations.

4.3.1 Combination of Gaussian

For any given set $L = \{l_1, l_2, \dots, l_n\}$, there are $2^n - 1$ possible (non-empty) subsets. We seek to calculate the sum of all the possible subsets. We propose a simple process for calculating the sum of all subsets using binary numbers. Consider the binary number B with n digits where each digit corresponds to one element of the set L . The $2^n - 1$ combinations can then be expressed by a scalar product of all the possible binary numbers from $00\dots1$ to $11\dots1$ with L :

$$sum_i = \langle B_i, L \rangle \quad (4.24)$$

for a simple example $L = \{l_1, l_2, \dots, l_3\}$, the unique combinations of the set can be given as:

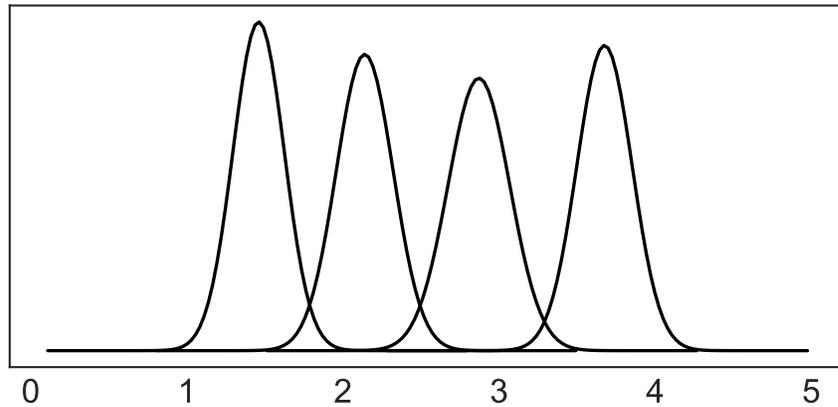
i	binary	sum
1	001	l_3
2	010	l_2
3	011	$l_2 + l_3$
4	100	l_1
5	101	$l_1 + l_3$
6	110	$l_1 + l_2$
7	111	$l_1 + l_2 + l_3$

Considering all the possible combinations, we create the new set $L' = \{l'_1, l'_2, \dots, l'_{2^n-1}\}$, where each new element is a sum of one or more independent Gaussian distributions derived as:

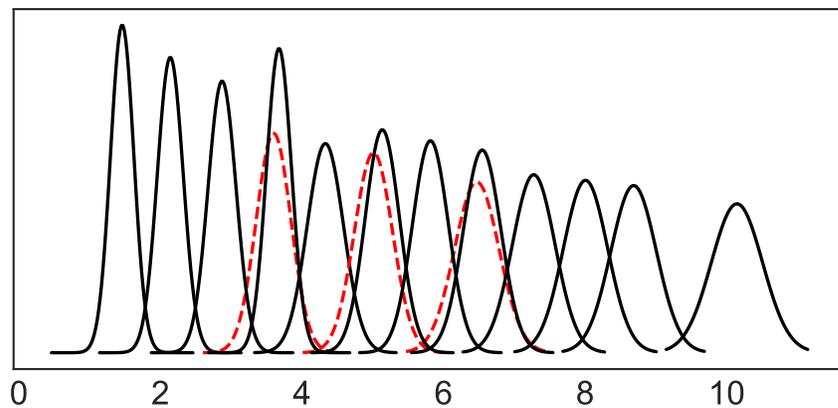
$$N(\mu_1, \sigma_1) + N(\mu_2, \sigma_2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \quad (4.25)$$

There is one more caveat to the process. It is unavoidable that the mean of some of the new derived distributions would land in the proximity of each other, not that dissimilar to the original NILM problem of two or more steady states of small loads summing up to a similar size of a bigger load. It is evident that the result of inference can be significantly affected under such a scenario. Therefore, we propose to prune the new set with the following condition: if $\mu'_i - \mu'_j < \epsilon$, we remove l'_i if $\sum_{k=1}^n b_{ki} > \sum_{k=1}^n b_{kj}$ or else keep it and remove l'_j where b_{ks} are the binary digits of the corresponding binary numbers. The threshold ϵ effect on the performance of the algorithm should be considered carefully when pruning the set. Figure 4.8 depicts

the process in action where starting from clusters in (a), we arrive at the new pruned set in (b).



(a)



(b)

Figure 4.8: An example of (a) initial clusters (b) pruned sum of clusters (red are filtered).

Figure 4.9 depicts a graphical representation of inference using a mixture of Gaussians with known parameters. It is of high importance to note that in the observation sequence, we only consider the *large* change points in the time series. This significantly reduces the computational cost without any apparent effect of the performance of the algorithm (this fact is also reported by [41]). Once the hidden states are inferred, the power can be estimated as:

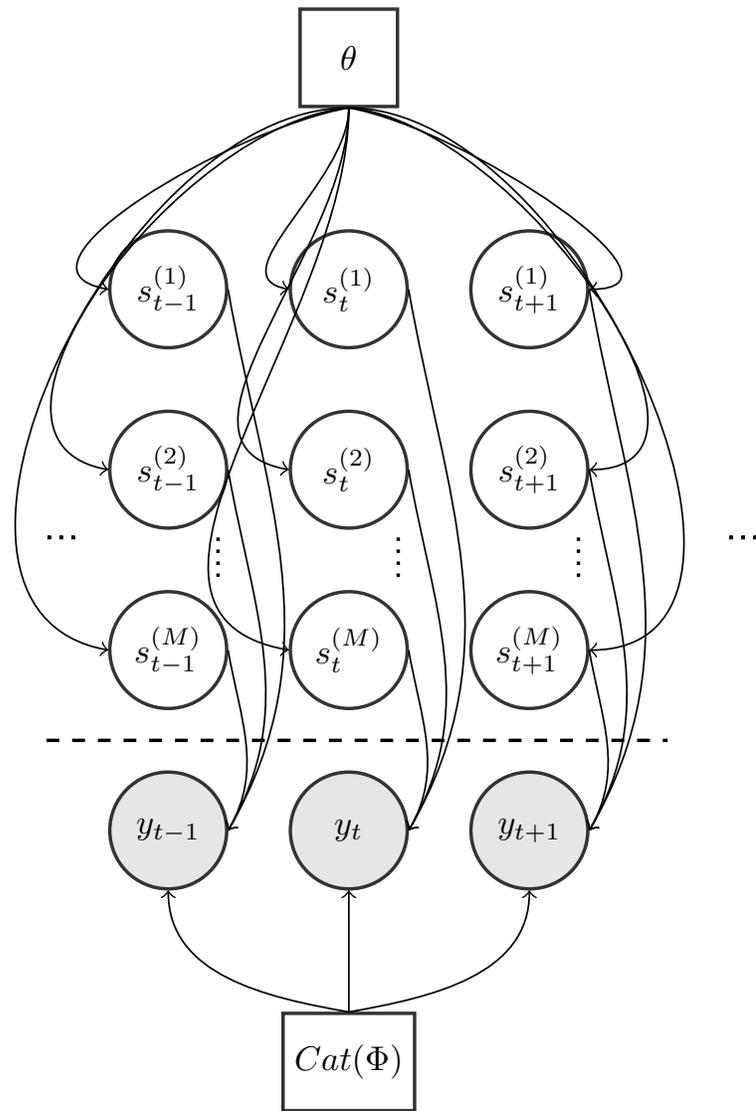


Figure 4.9: Graphical representation of Gaussian mixtures with known parameters.

$$\hat{y}_t = \sum_1^M W^{(m)} \theta_t^{(m)} \quad (4.26)$$

where $\theta_t^{(m)}$ are the means of the original clusters and W is the state matrix with

binary entries.

4.4 Summary

In this chapter, we presented a sequential process to infer and disaggregate individual load activity from a single aggregate point of measurement. This process starts by first filtering the background. We then model individual loads from their corresponding edges as Gaussian mixtures. Using the models, we present a fast approximation cable of handling the inference task using all the viable combinations of Gaussian mixtures. The major contributions presented in this chapter are put through empirical evaluation in the next chapter.

CHAPTER 5: PERFORMANCE ANALYSIS ON SEVERAL PUBLIC AND PRIVATE DATASETS

This chapter is an in-depth look at the results of deploying the proposed algorithms described in the previous chapter to the problem of energy disaggregation. To this end, a comparison is made for the performance of the proposed NILM against that of the state of the art methodologies discussed in Chapters 2 and 3. We define the grounds of comparison in detail in the first section, and the terms defined here are used in the evaluation process. A performance comparison of different algorithms on the REDD [8] dataset is presented in the next section. This is specifically of value since most of the available publications of NILM have used this dataset for experimental evaluation. In the next section, we run tests on AMPds [9] dataset, a publicly available dataset collected minutely from a residential building for a whole year. Finally, and most notably, we make a performance comparison between the proposed NILM and SparseNILM [11] for data collected from a couple of mid-size Commercial buildings. To the best of our knowledge, there is no comparable public dataset collected from commercial buildings available at the moment. Therefore a thorough description of data accusation, and the general characteristics of the dataset is provided alongside the results.

5.1 Evaluation Metrics

There has been a range of accuracy metrics used for evaluation purposes in NILM. This is partly due to different datasets and application areas with NILM.

Recently there has been an effort to make a unified platform for disaggregation, one with a standard metadata format and accuracy metrics. This resulted in the

development of NILMTK [66], a toolkit designed to help researchers evaluate the accuracy of NILM algorithms. Although not completely accepted as a standard for the field, it has provided a good guideline for some of the accuracy metrics that can be used for NILM evaluation. Here we mention a few useful accuracy metrics that are relevant to the comparisons along with the reasoning behind it.

Like any other learning task, the popular metrics precision, recall, accuracy, etc., can be adopted for NILM evaluation, but they need to be defined properly in context, or else they can result in misleading interpretations. Consider the following example: often in NILM publication, accuracy has often defined as $\text{Acc.} = \frac{\text{correctly predicted instance}}{\text{total instance}}$. Consider a scenario in which a load is only active for an aggregate time of an hour during a 24 hour period (This is actually a very common phenomenon with most residential loads like dishwasher, washing machine, dryer, etc.). If one predicts the load to be off for the whole day, this will result in an accuracy of 95.83%, which is a very skewed representation of accuracy! We can easily address this issue by reporting recall and precision.

Kolter [40] reports precision and recall at the circuit level. He defines these terms in a rather peculiar way: recall measures what portion of a given circuit's energy is correctly classified, while precision measures, of the energy assigned to a circuit, how much truly belonged to that circuit. The reason he suggests reporting circuits rather than actual load activity is the fact that his dataset (REDD) does not contain the data for the individually submetered loads. Since we did not have such a problem with our own collected dataset, and in general, such a scenario is not ideal, we chose to put the focus on reporting the states of the tracked loads rather than circuits. We also consider the fact that a load can take more than two states. With that in mind, we define precision and recall for a multi-class scenario as:

Confusion Matrix

Actual	Class₀	C_{00}	...	C_{0M}
	⋮	⋮	⋮	⋮
	Class_M	C_{M0}	...	C_{MM}
		Class₀	...	Class_M

Prediction

Figure 5.1: Confusion matrix.

$$\text{Precision}_i = \frac{C_{ii}}{\sum_j C_{ij}} \quad (5.1)$$

$$\text{Recall}_i = \frac{C_{ii}}{\sum_j C_{ij}} \quad (5.2)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5.3)$$

Where C is the confusion matrix, the table depicting the correct and confused classified states for a classifier, as shown in Figure 5.1.

As an example, in the hypothetical example described above, we would have a 0% recall. F1 score represents the harmonic mean of precision and recall and would also be 0%.

Finally, the ideal NILM should be able to track the power, and not only the state of each individual load. One popular way for judging the performance of the disag-

gregator is to calculate the normalized disaggregation error as:

$$\text{Acc.} = 1 - \frac{\sum_{t=1}^T \sum_{i=1}^M |\hat{y}_t^{(i)} - y_t^{(i)}|}{2 \sum_{t=1}^T \bar{y}_t} \quad (5.4)$$

where \bar{y}_t is the observed total power consumption, $\hat{y}_t^{(i)}$ and $y_t^{(i)}$ are the estimated and ground-truth power of appliance i at time t .¹

5.2 Performance of NILM on Residential Datasets

Of the datasets discussed in Chapter 2, we consider two of them that have sufficient metadata for validation and are also publicly available AMPds [9] and REDD [8]. These two datasets have different sampling frequency and measurement types. In favor of consistency, we have downsampled to once per minute for apparent power, which we believe is a reasonable frequency offered by the commercial smart meters available in the market. It is important to note that while our approach should be considered as semi-supervised, we have kept the level of supervision limited to tagging the results of our clustering algorithm. It is, therefore, important to distinguish a major difference between our approach and other works in that all of them require training of a model based on sub-metered data.

5.2.1 Disaggregating REDD

The Reference Energy Disaggregation Data set (REDD) [8] is a public dataset which monitors six residential homes at high frequency in Boston for a few month periods each.² The low-frequency version of the dataset has once per second data for the mains and once per 3 seconds of data for the appliances. We downsampled

¹While the averaging with observed total power consumption can be questionable; we adopt this metric as it has been very popular in other available literature and it provides a good basis for comparison with other algorithms whose codes are not readily available online.

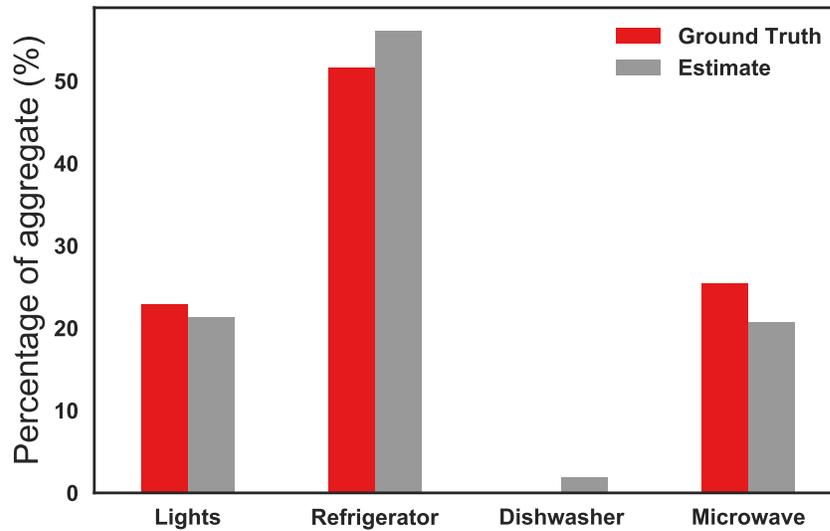
²This varies significantly for each home.

Table 5.1: Performance comparison between state of the art algorithms and the proposed NILM, performed on REDD dataset.

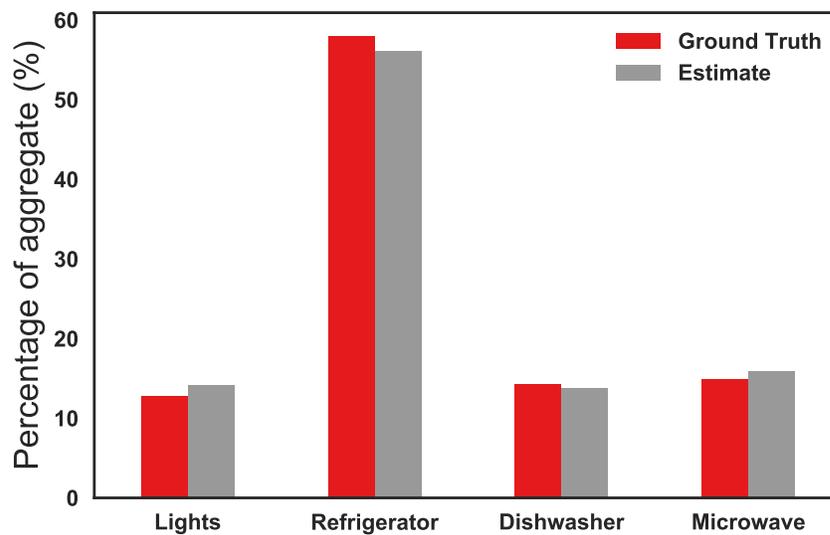
House	Kolter [8]	Johnson [41]	Makonin [11]	NILM
House 1	46.6%	82.1%	99.3%	97.0%
House 2	50.8%	84.8%	99.0%	93.0%
House 3	33.3%	81.5%	97.5%	96.0%
House 6	55.7%	77.7%	99.7%	92.0%
Average	46.6	81.5	98.9	94.5

the dataset to once per minute data using the Fourier method to avoid aliasing. As suggested by Johnson [41] we chose the top 5 power-drawing devices (refrigerator, lighting, dishwasher, microwave, furnace) across houses 1,2,3 and 6. In that paper, Johnson mentions using 18 24-hour segments for their experimental results. Since these periods were not specifically identified, we carried the test on all the available data from the houses mentioned above. We also had to address some syncing problems, also mentioned by [11]. It is important to note that not all the mentioned appliances are present in all the datasets. Nevertheless, to make comparable results, we stick to the proposed structure and choose equation 5.4 as our accuracy metric for the same sake. Table 5.1 shows the final results. On the first look, it is clear that we were able to improve on Kolter’s [40] and Johnson’s [41] results significantly and come close to matching the results of Makonin [11]. The author would like to mention the fact that having had no success in reproducing the results of [11] based on his published code, we chose to copy the numbers directly from the tables provided in the paper.

Figure 5.5 depicts the breakdown of the result of disaggregation for two days versus that of the ground truth data. While it is not possible to make direct comparisons (the exact dates are unknown), it looks likely that there is an improvement in results compared to the figures published in the three other papers.



(a)



(b)

Figure 5.2: Estimation accuracy results for 2 different test days compared to their respective ground truth.

5.2.2 Disaggregating AMPds

Almanac of Minutely Power dataset (AMPds) [9] is collected from a single residential house in Vancouver, BC, using 21 sub-meters for an entire year. The frequency of data acquisition is once per minute, and unlike REDD, it is consistent at the mains

and appliance level. We followed the publisher’s guideline in tracking the following deferrable³ loads in the dataset: the clothes dryer, the dishwasher, the heat pump, and the kitchen wall oven.

On top of accuracy, we report recall and precision for each load. Table 5.1 Table 5.2 shows the results of our proposed approach versus that of the SparseNILM. As evident by the numbers in the table, the proposed NILM outperforms SparseNILM considerably. On a more detailed level, it looks like the performance of proposed NILM is less impressive for Oven compared to the other loads. This is fueled by the fact that oven is a multi-stage load, and there is significant variation in its power demand.

It is important to note that upon closer look and after many trials with the SparseNILM algorithm, we were unsuccessful in reproducing the results published in Makonin full thesis [67]. We suspect that it might be the case that the authors have chosen to use up to 8 states for the training of their HMMs and have also used a considerable amount of data (several months) in the training phase. In such a mode, there is a good chance that running the algorithm would improve the results. Having said that, our trials on a personal computer were all unsuccessful. It is our belief that given the computational complexity of exact inference using Viterbi, it is virtually impossible to solve such a problem in a sensible time, even in the case of solving a very sparse system.

³Deferrable appliances are have operation schedules that can be moved during a day based on the feedback provided by NILM.

Table 5.2: Performance comparison between SparseNILM and the proposed NILM, performed on AMDPs dataset.

Load	SparseNILM[11]		NILM	
	Precision/Recall	F1	Precision/Recall	F1
Dishwasher	98.2% / 60.6%	74.9%	92.3% / 97.1%	94.6%
Heat Pump	87.9%/94.6%	91.1%	99.7% / 97.4%	98.5%
Dryer	98.3% / 97.8%	98.0%	99.3% / 98.4%	98.8%
Oven	46.3% / 54.5%	50.1%	94.4% / 86.1%	90.1%
Average	84.4% / 75.2%	78.5%	96.4% / 94.8%	95.5%

5.3 Disaggregating Commercial Buildings

The temporal pattern of energy consumption in small retail facilities is quite different from that of residential households. Figure 5.3, for instance, shows the major usage by category at one 3,000ft² retail site over several months. As shown, the aggregate load has the following major categories:

- **Heating, Ventilation, and Air Conditioning (HVAC):** These are the primary space-conditioning systems and typically consist of several packaged rooftop units (RTUs).
- **Interior Lighting:** This category includes all lights within the building.
- **Exterior Lighting:** This includes all parking-lot lights and signage.
- **Plug load:** This includes items such as computers, printers, refrigerators, etc.
- **Miscellaneous Process Loads:** This includes miscellaneous systems such as water heaters, water fountains, and bathroom exhaust fans.

Figure 5.3 serves to frame the goals for a NILM deployed in a retail building. Without an extensive network of sub-meters, the best-case monitoring system would

be one that tracks daily, weekly or monthly energy consumption for an entire building. Such information allows energy managers to compare various sites to determine which ones consume the most energy. This information is not granular enough, however, to provide any particular reasons why some sites consume more. With a NILM, on the other hand, one could track the energy consumption by major category. For decision-makers to begin pursuing energy-efficiency projects or maintenance initiatives, data such as the chart shown in Figure. 5.3 provides tremendous granularity.

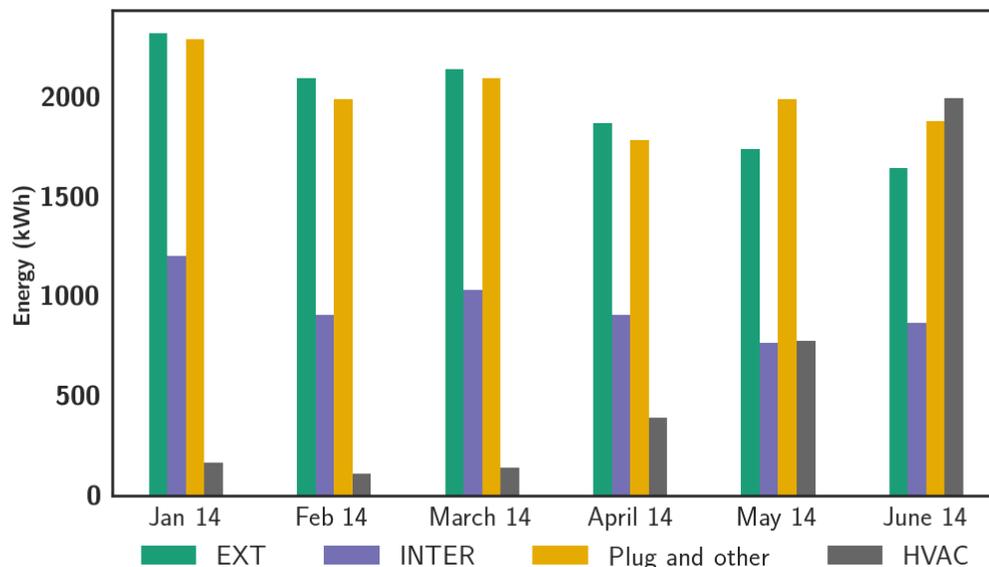


Figure 5.3: Major monthly consumers in a retail bank in South Carolina. Note that HVAC electricity consumption is low in the winter because the facility relies on gas heat.

One approach to creating a monthly chart such as Figure. 5.3 would be to first detect some of the major consumers in the aggregate power signal. For the purpose of this thesis, we formally define the task of NILM in these types of building as tracking the consumption of the HVACs and exterior lights.

To understand the motivation, consider Figure 5.4, which shows an example daily waveform. There are clearly two distinct periods. During non-business hours (i.e.,

from about 8:30 PM until 8:00 AM the next day), the power level is relatively large and constant. This high-level results from the combination of the building's baseload and its exterior lights. During business hours, on the other hand, the load varies continuously as the air conditioners cycle and staff work with customers. The first major goal of a NILM would be to detect the operation of the exterior lights. This is relatively straightforward since these lights should only be energized from approximately dusk until dawn. Assuming the lights are operating properly, they will cause large step changes at relatively predictable times, as shown in Figure 5.4. In addition, one can also extract the energy consumed by the HVAC equipment. In a retail store, this typically means that one must detect and track the operation of several RTUs throughout the day. The remaining load categories include interior lights and plug and process loads. Interior lights may be possible to track, but plug and process loads are often very difficult. In the author's experience, these loads are often left energized at night, and thus they can be viewed as one continuously variable element. Such an approximation is not unreasonable. Consider, for instance, that a bank may contain many computers, and knowing the consumption of any one is not nearly as important as knowing their combined impact. The approach taken in here is thus to disaggregate the exterior lighting and HVAC equipment, lumping everything else into one aggregate category. Such granularity may sound incomplete, but it provides a tremendously greater level of detail than is currently available. We take it as the task in hand and compare the results of proposed NILM to SparseNILM.

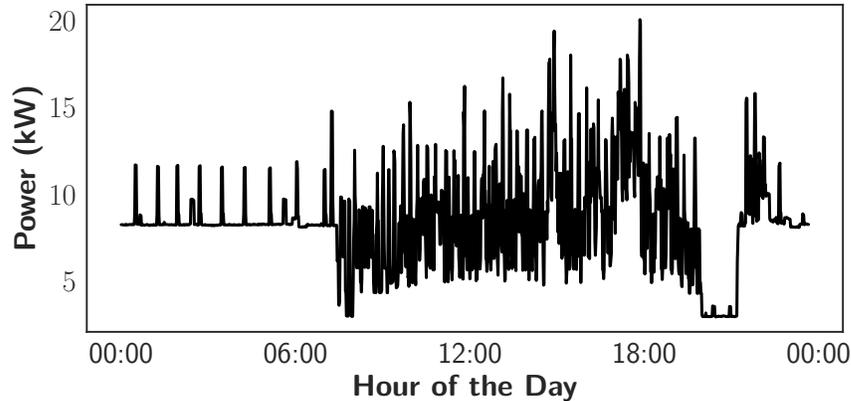


Figure 5.4: An example of a daily waveform in a commercial building.

We should mention that to the best of our knowledge, there is no comparable public dataset available of this kind at the moment. Therefore, we find it necessary to discuss the process of data acquisition and provide an overview of the dataset before discussing the results.

5.3.1 Data Acquisition

To support the work in this thesis and other activities, data was collected from a number of retail banks throughout the Charlotte, North Carolina, region. The primary dataset used in this thesis was collected from a building served from a 208/120V, 60Hz three-phase electrical service. The facility has two main breaker panels with a collection of three-phase, two-phase, and single-phase loads. The top image in Fig. Figure 5.5 shows the two breaker cabinets with the front panels removed. As shown, individual clamp-on current transformers (CTs) were installed on each of the circuits in the building. The analog output current from each of these CTs was then passed to a set of custom-designed data-acquisition systems. The final arrangement appears as shown in the bottom of Fig. Figure 5.5. Each CT feeds into a sensor box that interfaces to a custom Linux machine. These sensor boxes are shown on the floor underneath the breaker cabinets.

Figure 5.6 shows the basic hardware arrangement in the custom data-acquisition

system. As shown, each CT was terminated in a 100Ω resistor to convert the output current to a voltage. Thirteen individual CT outputs were directly interfaced to a 14-channel, 16-bit datalogger made by Labjack. The remaining channel was interfaced to the output of an LV-20 voltage sensor from LEM. Five individual dataloggers were then directly interfaced to a PC via a 5-port gigabit ethernet switch. Several such switches were interfaced to the same computer.

Data was recorded from each individual channel at a rate of 3kHz. A custom software program allowed data to be continuously streamed from several different dataloggers simultaneously. Raw data from each channel was directly recorded and stored into an hourly data file. In addition, lower frequency power signals were also derived for each channel in real-time. This process occurred in several steps. First, data from each channel was placed in a small buffer and resampled according to the process outlined in [68]. This approach allowed the data to be resampled so that each period contained an exact integer number of points per period. Each current was then multiplied by the corresponding voltage and averaged over one-period. The corresponding power value was recalculated every half-period, leading to an output signal recorded at 120Hz. This power data was also archived into a file for each channel on an hourly basis. Using the resampling process, it was possible to derive the other two voltage signals in the three-phase system. This leads to some small error, but the team was able to verify it to be less than 1 by comparing the total power to that obtained from a building meter.



(a)



(b)

Figure 5.5: NILM data acquisition.

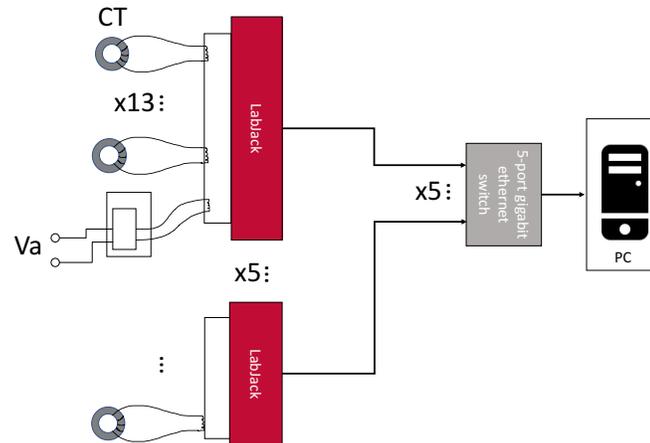


Figure 5.6: Basic hardware arrangement in the custom data-acquisition system.

The data-acquisition system was extremely extensive and was designed assuming that significant data might be needed for various NILM applications. Ultimately, each hourly file was downsampled to provide the once-per-minute values used in this work.

5.3.2 Experimental results on the bank datasets

To better illustrate the performance of the algorithm, we first evaluate the corner cases, by which we are referring to the period for which there are a lot of concurrent load activity in the data. One such case happens during the summer, for which there is a lot of cooling activity. In the two bank buildings here, the 3 RTUs often have overlapping operation periods. This is depicted in Figure 5.7 for B1 in July 2014. In this figure, the upper diagonal represents the scatter plots, the lower diagonal is the bivariate kernel density estimates, and the diagonal itself represents a kernel density estimate of the power distribution. Here, each RTU has two operating modes: solo fan which has an operating demand of $\tilde{700}$ watts and fan+compressor which has a varying demand of 3500-4200 watts.⁴ It can be argued that disaggregating the concurrent cooling activities in the presence of a varying baseline is one of the main challenges

⁴This variation is caused by excessive heat periods in summer.

that ultimately decides the success or failure of any NILM algorithm deployed in commercial buildings, especially in warmer climates.

The proposed disaggregation process, as described in chapter 4, is as follows: To estimate the state of exterior lights, we use a change of mean detection algorithm; this process is done iteratively for each day, and exterior light is extracted from the aggregate. The loads are modeled based on edge detection and clustering. The only supervised step of the algorithm is to associate the clusters with the operating power of the appliances (more specifically, a desired subset of appliances), e.g., multiple rooftop units. The disaggregation is carried out by inferring the Viterbi path for observed change points in the time series using a combination of Gaussians.

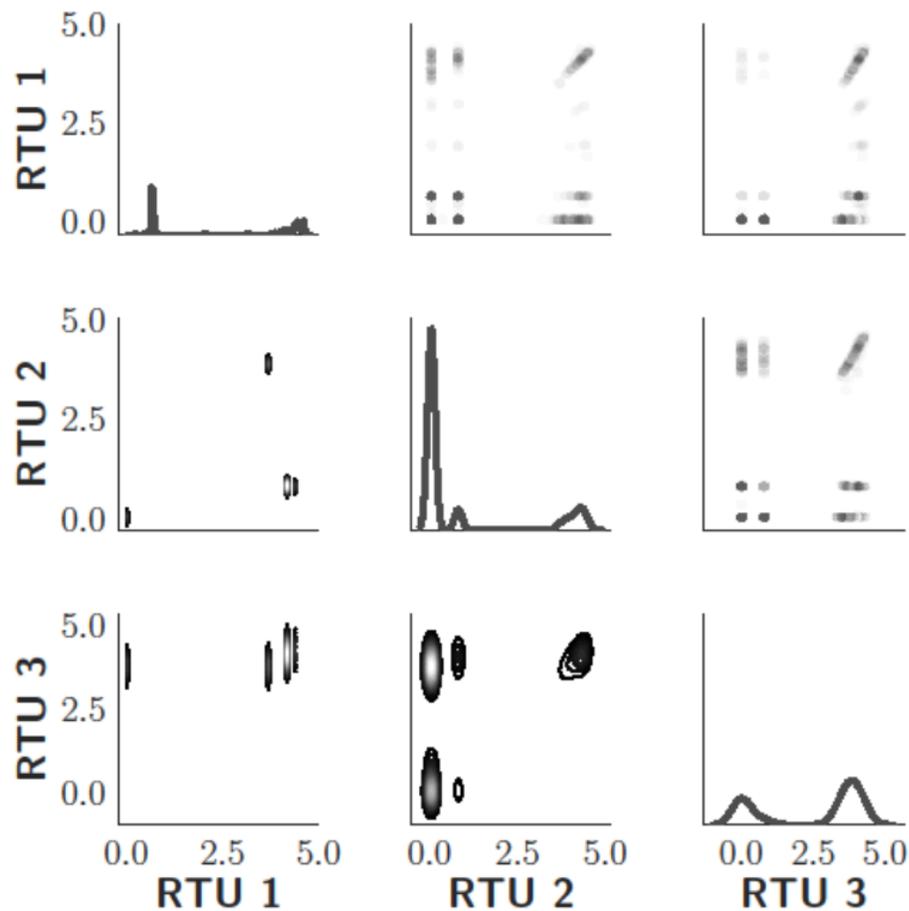


Figure 5.7: Pairwise distributions of three RTU units in a mid-size bank building during the first week of July.

The SparsNILM algorithm builds a super-state HMM trained on individually metered loads and assumes the remaining un-metered portion as a separate load.⁵ These are modeled as probability mass function, PMF (a bimodal ON/OFF probability assignment for each mode of load) that are then combined and used in a Sparse Viterbi algorithm to do exact inference. We went through the efforts of submetering every individual circuit in two mid-size bank buildings, which we refer to as B1 and B2. We collected sub-metered data from 81 and 104 separate channels for each building, respectively, at 3kHz frequency and down-sampled to once per minute data. Using this data, the PMFs for top individual energy consumers of the buildings (i.e., exterior light, interior light, and RTUs) and the remaining subtracted signal were learned. It is important to note that the super-state HMMs are formed using at most two states for each load (which is a reasonable assumption for the exterior lights and RTUs).

Figure 5.8 shows the results of the state tracking of exterior lights and RTUs for each algorithm for the month of June (high cooling activity), depicted through the confusion matrix for full transparency. (a) and (c) are for the aggregate operational state of 3 identical RTUs while (b) and (d) depict the results for the state prediction of the exterior lights. It is evident that the SparseNILM algorithm is significantly under predicting the cooling activity in the building. This can be explained by the presence varying baseline, which is also being modeled as a two state load here. In most of the misclassification, the algorithm is wrongly confusing the observation of multiple RTUs for that of a combination of exterior light and the varying baseline. This also simultaneously results in an overestimation of exterior light power. While the results clearly display the superiority of the proposed approach, it is important to note that we believe the performance of SparseNILM algorithm can improve if we increase the number of allowed states in modeling the PMFs. However, and like it was the case in the previous test, that would significantly increase the time

⁵Details on why this can not be considered a reasonable assumption were discussed in Chapter 3.

complexity of the algorithm to the point that convergence becomes unrealistic. It is also worth noting that the training phase of SparseNILM can significantly affect the performance of the algorithm. For example, if the training happens in January, there is very limited RTU activity in that month (and only in the shape of fan activity) as the building is being conditioned primarily using gas-fueled appliances. As a result, tests during a period in June would miss all the HVAC activities as PMFs were not learned accurately in training using January data. This is even without considering the effect of a completely different baseline in the building in January and June, which would further complicate the matter and reduce accuracy.

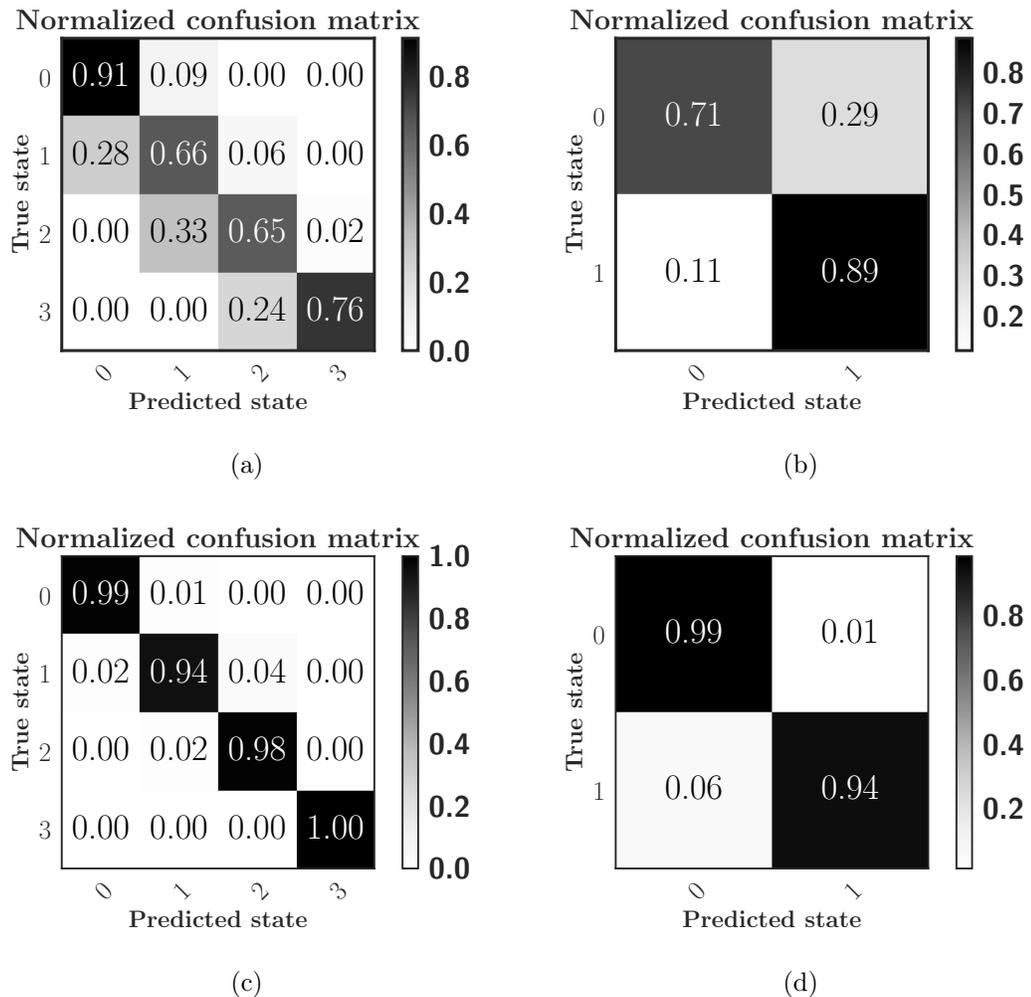


Figure 5.8: Comparing the performance of SparseNILM (Top) vs proposed NILM (Bottom) on one month of data.

Table 5.3 reports the result of disaggregation on the dataset for 6 months (Jan-June) and two months (June-July) for building B1 and B2, respectively. The superiority of our approach is even further highlighted once the extensive training period, consisting of several months of data (May-June), required for training SparseNILM is taken into account. The poor results of SparseNILM on B2 are due to high variability in un-metered load, something that also affects the results of our algorithm slightly. Even with this in mind, the robustness of the proposed approach is put to the test properly with this second building, and we believe that the results are very promising and a huge step in the right direction towards a more robust NILM algorithm.

Table 5.3: Performance comparison between SparseNILM and the proposed NILM, performed on mid-size commercial building dataset.

Build.	Load	Acc.		Precision		Recall		F1 score	
		SparseNILM	NILM	SparseNILM	NILM	SparseNILM	NILM	SparseNILM	NILM
B1	RTU	69.91%	92.18%	74.50%	97.16%	75.85%	97.17%	75.13%	97.16%
	EXT	76.13%	95.06%	80.00%	96.72%	81.0%	96.68%	80.49%	96.69%
B2	RTU	71.06%	90.76%	62.08%	88.95%	60.94%	87.79%	61.50%	88.36%
	EXT	70.10%	95.62%	76.15%	96.32%	72.99%	96.32%	74.53%	96.32%
Average		71.80%	93.40%	73.18%	94.78%	72.67%	94.49%	72.91%	94.63%

5.3.3 An example use case for HVAC monitoring using NILM

Figure 5.9 provides an example of HVAC energy consumption for different hours of the day for B1 as described in previous section (The building has gas heating). The plot is partially inspired by [69]. It is apparent from this figure that during unoccupied hours (0-8 and 20-23), unless the outside temperature is greater than 75, there is very little cooling by the air conditioner apart from some fan action. Occupied hours, on the other hand, follow the general rule of heat transfer in that there is almost a linear relationship between the increasing temperatures and the HVAC demand. Monitoring this relationship can provide significant diagnostic insight. Furthermore, comparing such graphs to those of similar regional buildings provides great feedback on operational schedules and efficiency of units among buildings, and the potential benefits become even greater as the size of the cluster increases. Also, from a utility perspective, this information is not only crucial in deriving their policies but is also essential in developing predictive models for demand control.

In the plots provided in this figure, we have demonstrated the strength of our proposed algorithm in producing accurate sub-hourly performance curves for a RTU unit, based on the disaggregated activities from the aggregate power signal, as described in the previous section.

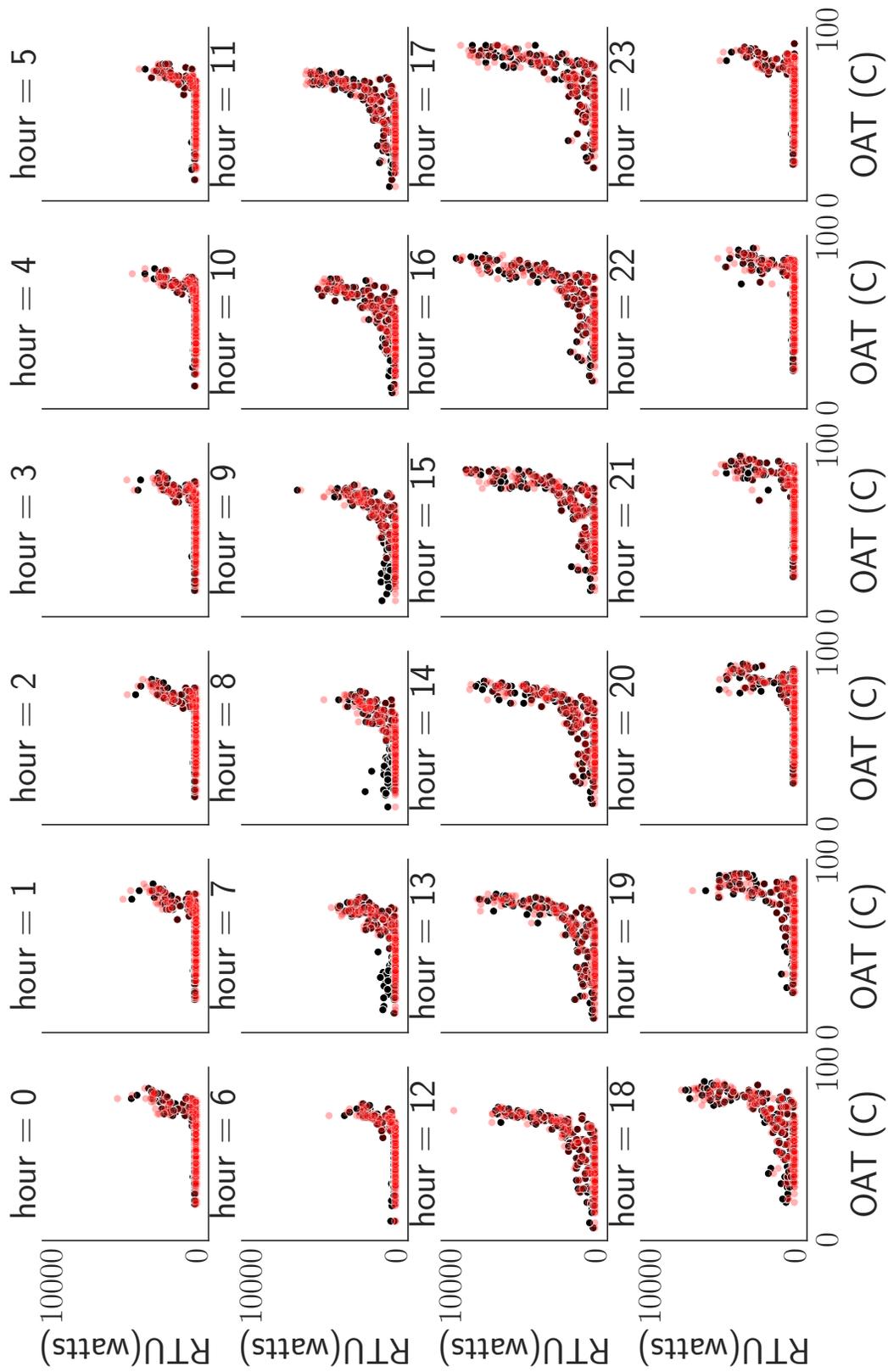


Figure 5.9: Sub-hourly load data vs. outside air temperature (one plot for each hour). Red dots are estimates, and black dots are ground truth.

5.4 Summary

In this chapter, we put the performance of the proposed NILM algorithm to test across multiple datasets. First, we defined the terms of comparison, some consistent with the available literature for the benefit of direct comparison, and others proposed by this research for further transparency. We first evaluated our algorithm on available public residential datasets REDD and AMPDs for which there are available results in the literature to compare to and showed that we improve upon these results. We then described the process of data acquisition for creating a dataset with corresponding appliance level submetered data for commercial buildings. This dataset is unique in that it is very different in nature to the available residential datasets. We evaluated our algorithm against SparseNILM using the commercial buildings datasets and showed that our proposed algorithm is capable of handling one of the more difficult tasks in disaggregation, namely disaggregating concurrent appliance activities in the presence of a varying baseline.

CHAPTER 6: SUMMARY, LIMITATIONS AND FUTURE DIRECTIONS

In this thesis, we proposed a new NILM methodology suited to disaggregating HVACs and exterior lighting in commercial buildings. The final chapter of this thesis summarizes the results, highlights the limitation of the proposed method, and provide ample suggestion for future work and possible directions.

6.1 Summary

We began this dissertation by highlighting the need to address the challenges facing the energy sector today. Specifically, we focused on challenges in the building sector, which is responsible for nearly 40% of the consumed energy in U.S. We focused on space cooling as a major driver of this consumption and highlighted the role of appliance level granular data in helping with the monitoring of cooling systems. It was in this context that we arrived at the conclusion that the urgency, cost, risks, and challenges of appliance level monitoring constitute the need to use the state of the art infrastructure, namely the smart meters. Therefore, we defined the goal of the thesis to be disaggregation of major loads in commercial buildings through non intrusive load monitoring.

In chapters 2 and 3 of this dissertation, a historical overview of the different types of monitoring and their various limitations was presented. In Chapter 2, the consumption patterns in residential and commercial buildings and the merits of monitoring in each sector were discussed. A brief description of the available public datasets highlighted the lack of publicly available data from commercial buildings and a need to collect data for these types of buildings. Finally, A detailed review of different NILM algorithms was provided in chapter 2 with more emphasize on non-intrusive

low-frequency methods using HMMs.

In Chapter 3, we highlighted the major shortcomings in the current NILM approaches. First, We discussed the problem of supervision and as most of the current approaches require monitored appliance level data for training the HMMs, and thus are incompatible with the current feed of metered data and would require significant retrofitting and or expert level involvement to launch the training phase. We previously showed that the financial incentives for such efforts are not clear, and the burden is most likely to fall on the consumer shoulder. Second, we argued that the computational complexity of some of the approaches would not be suited to edge computing and in some cases, makes inference an impossibly costly process. We also showed that the approximations does not always converge to a reasonable sub-optimal solution. Third, we showed that a smooth variation in the baseline of daily consumption could significantly dampen the accuracy of NILM algorithms relying on edge detection. Our argument once again highlighted a need for a more diverse portfolio of data as this problem is a direct result of a lack of such patterns in the limited available residential datasets.

In Chapter 4 of this dissertation, we proposed algorithms that are capable of doing a robust NILM. We defined the robustness in NILM as the capability to monitor the power of major consumers of energy in buildings for a large, diverse portfolio of different types. The problem was broken down into three steps; a smooth background filtering removes the baseline from the aggregate pattern and thus decreases the chances of mix-ups in the inference stage. Edge detection and clustering build generalized Gaussian models for individual appliances, and a combination of Gaussians is then used to for state inference. Finally, The algorithm estimates appliance power using the appliance states and means of operating power.

Finally, in Chapter 5, we carried out an in-depth performance analysis of the proposed algorithms versus several other studies that use HMMs. We first defined the

metrics for comparison in detail. Using these metrics, we compared the results of our algorithm to [41, 40, 11] on the REDD [8] dataset. We showed the superiority of our approach against two of those and were able to match the results of the third. We also compared our algorithm and SparseNILM [11] on another dataset collected from a single residential house for a whole year [9] and was able to show significant improvements in both Recall and Precision. Finally, we provided the details of creating a dataset with ample submetered data for commercial buildings and trained SparseNILM using this data. We then compared the results of SparseNILM vs. the proposed NILM on our dataset collected from two mid-size Commercial buildings. Our algorithm demonstrated a significant advantage against SparseNILM for disaggregating HVAC and exterior lights in commercial buildings. We then demonstrated an example use case for HVAC monitoring using the results of our algorithm.

As the final remarks of this summary, we believe that we have presented a novel algorithm capable of tracking the major loads in commercial buildings. We also demonstrated the robustness of this algorithm by testing it against various different datasets collected from different types of buildings.

6.2 Limitations

We openly acknowledge the fact that our proposed methodology is not capable of disaggregating all appliances in any given building. In fact, in the introduction of this thesis, we put the emphasis on HVAC and controlled lights and discussed the importance of tracking the state of such loads. It was with this consideration that we intentionally reduced the scope of the NILM problem to a subset of all available appliances in favor of increasing the accuracy of results for the target loads. We would like to discuss a number of potential cases for which our algorithms will likely not perform well.

One of the major assumptions that we have made in this thesis is that appliances can be modeled using binary states, or a combination of these binary states. This

is not necessarily true at all times. In fact, a good example of such cases is buildings equipped with variable refrigerant flow (VRF) systems. VRF systems use many evaporators at different capacities, with different configurations through control systems. Therefore their consumption patterns can vary continuously and significantly from one-time frame to another [64]. Disaggregating such a load would likely require either a better data granularity or the use of the corresponding control signal for VRF.

Another limitation has to do with the level of supervision required to train the algorithm. At the moment, the disaggregation requires manual labeling of disaggregated loads. This might not pose a significant challenge in a case where there is a data repository of purchased appliances inside commercial buildings, but as the author has experienced this multiple times, getting access to such data would create a whole lot of other bureaucratic challenges. We discuss the potential solutions to this problem in the next section.

The third limitation of the work is the ability of the algorithm to disaggregate between multiple appliances with similar steady state levels. Again this would require a finer resolution of data, possibly 1 second and better, as with once per minute data, there is a very limited potential to differentiate between such scenarios given the noise levels involved. We also discuss one potential way to address this in the next section.

6.3 Future directions

Any future work has to address the above limitation of the current NILM. Disaggregation for continuously variable loads with low-frequency data appears to be an almost impossible task at the moment. While there are ways to extend HMMs to account for continuous variations, for example through the introduction of nonlinear state estimators like extended Kalman filter [70], it is not that clear that data collected in one-minute intervals would be suited to that kind of approach. Also, this would definitely require the use of appliance level data which makes this approach

even more impractical. There is another way to be considered. At the moment, commercial buildings are being equipped with building monitoring systems that regularly report and store the control signals for such units at minute intervals. It is possible to leverage that information in edge classification and pair the edges relevant to specific appliances with their respective temporal patterns. This would specifically be handy in dealing with a VRF system. While this problem can be considered a different one to the original formulation of NILM, it still pursues the same goals of this study, which is disaggregating the cooling loads in the commercial building. Furthermore, it is important to mention that even if BAS was to be rolled out to all buildings tomorrow, we have seen enough evidence to suggest that NILM can still be very useful in validating the control signals as the systems can often go offline or report/record bad data.

As pointed out by Parson [24], gas and water smart meters are also being deployed, and there are several studies that try to map the problem of NILM to this new domain [71, 72, 73]. Considering the literature, we think that the current proposed approach would be able to handle the problems in that domain as well, with minimal alterations.

Finally, and ultimately, any successful research in NILM would have to be generalizable on a bigger scale. Going forward, we believe that if a large scale deployment of our proposed algorithms on a portfolio of commercial buildings is pursued, there is a real chance that this would lead to a realization of the untapped potentials of NILM, especially in the commercial sector.

REFERENCES

- [1] G. W. Hart, “Nonintrusive appliance load monitoring,” *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.
- [2] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, “Is disaggregation the holy grail of energy efficiency? the case of electricity,” *Energy Policy*, vol. 52, pp. 213–234, Jan. 2013.
- [3] U. G. G. A. M. Initiative, “Reducing us greenhouse gas emissions: How much at what cost?,” 2007.
- [4] G. T. Gardner and P. C. Stern, “The short list: The most effective actions us households can take to curb climate change,” *Environment: science and policy for sustainable development*, vol. 50, no. 5, pp. 12–25, 2008.
- [5] J. A. Laitner, K. Ehrhardt-Martinez, and V. McKinney, “Examining the scale of the behaviour energy efficiency continuum,” in *European Council for an Energy Efficient Economy*, 2009.
- [6] S. Katipamula and M. R. Brambley, “Methods for fault detection, diagnostics, and prognostics for building systems—a review, part I,” *HVAC&R Research*, vol. 11, no. 1, pp. 3–25, 2005.
- [7] “U.s. energy information administration (eia) independent statistics and analysis.” <https://www.eia.gov/tools/faqs/faq.php?id=86&t=1>. Accessed: 2019-09-01.
- [8] J. Z. Kolter and M. J. Johnson, “Redd: A public data set for energy disaggregation research,” in *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, San Diego, CA, vol. 25, pp. 59–62, 2011.
- [9] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajić, “Ampds: A public dataset for load disaggregation and eco-feedback research,” in *2013 IEEE Electrical Power & Energy Conference*, pp. 1–6, IEEE, 2013.
- [10] A. Filip, “Blued: A fully labeled public dataset for event-based non-intrusive load monitoring research,” in *2nd Workshop on Data Mining Applications in Sustainability (SustKDD)*, p. 2012, 2011.
- [11] S. Makonin, F. Popowich, I. V. Bajić, B. Gill, and L. Bartram, “Exploiting hmm sparsity to perform online real-time nonintrusive load monitoring,” *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2575–2585, 2015.
- [12] “Navigant research leaderboard report: Building energy management systems,” tech. rep., Navigant Research, 2016.

- [13] “Realizing the energy efficiency potential of small buildings,” tech. rep., National Trust for Historic Preservation Preservation Green Lab and New Buildings Institute, June 2013.
- [14] R. J. Mowris, A. Blankenship, and E. Jones, “Field measurements of air conditioners with and without txvs,” in *Proceedings of*, 2004.
- [15] H. Li and J.E.Braun, “Application of automated fault detection and diagnostics for rooftop air conditioners in California,” in *Proc. of the 2004 ACEEE Summer Study on Energy Efficiency in Buildings*, 2004.
- [16] V. Loftness, A. Aziz, B. Lasternas, and S. Peters, “Ducks, dollars, or kWh,” in *Energy Accounts: Architectural Representations of Energy, Climate, and the Future* (D. Willis, W. W. Braham, K. Muramoto, and D. A. Barber, eds.), Routledge, 2016.
- [17] Statistica, “Smart appliances - united states: Statista market forecast.” <https://www.statista.com/outlook/389/109/smart-appliances/united-states#market-users>.
- [18] K. Gillespie, P. Haves, R. Hitchcock, J. Deringer, and K. Kinney, “A specifications guide for performance monitoring systems,” *Lawrence Berkeley National Laboratory, Pacific Gas and Electric Company, Deringer Group, QuEST*, 2007.
- [19] N. Gayeski, S. Kleindienst, J. Gagne, S. Samouhos, R. Cruz, and B. Werntz, “Data and interfaces for advanced building operations and maintenance.,” *ASHRAE Transactions*, vol. 122, no. 2, 2016.
- [20] R. McWilliam and A. Purvis, “Potential uses of embedded rfid in appliance identification.,” *International appliance manufacturing.*, vol. 10, pp. 100–107, 2006.
- [21] A. Schoofs, A. Guerrieri, D. T. Delaney, G. M. O’Hare, and A. G. Ruzzelli, “Annot: Automated electricity data annotation using wireless sensor networks,” in *2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, pp. 1–9, IEEE, 2010.
- [22] Y. Kim, T. Schmid, Z. M. Charbiwala, and M. B. Srivastava, “Viridiscop: design and implementation of a fine grained power monitoring system for homes,” in *Proceedings of the 11th international conference on Ubiquitous computing*, pp. 245–254, ACM, 2009.
- [23] A. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.
- [24] O. Parson, *Unsupervised training methods for non-intrusive appliance load monitoring from smart meter data*. PhD thesis, University of Southampton, 2014.

- [25] S. R. Shaw, S. B. Leeb, L. K. Norford, and R. W. Cox, "Nonintrusive load monitoring and diagnostics in power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 7, pp. 1445–1454, 2008.
- [26] S. Mostafavi, B. Futrell, J. Troxler, and R. W. Cox, "Leveraging cloud computing to convert the non-intrusive load monitor into a powerful framework for grid-responsive buildings," in *2016 IEEE International Conference on Big Data (Big Data)*, pp. 3107–3114, IEEE, 2016.
- [27] I. Grace, D. Datta, and S. Tassou, "Comparison of hermetic scroll and reciprocating compressors operating under varying refrigerant charge and load," 2002.
- [28] C. R. Laughman, *Fault detection methods for vapor-compression air conditioners using electrical measurements*. PhD thesis, Massachusetts Institute of Technology, Department of Architecture, 2008.
- [29] 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, no. 3, pp. 566–574, 2005.
- [30] S. B. Leeb, S. R. Shaw, and J. L. Kirtley, "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," *IEEE Transactions on Power Delivery*, vol. 10, no. 3, pp. 1200–1210, 1995.
- [31] K. D. Lee, *Electric load information system based on non-intrusive power monitoring*. PhD thesis, Massachusetts Institute of Technology, 2003.
- [32] 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 (nominated for the best paper award)," in *International Conference on Ubiquitous Computing*, pp. 271–288, Springer, 2007.
- [33] H.-H. Chang, "Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses," *Energies*, vol. 5, no. 11, pp. 4569–4589, 2012.
- [34] L. K. Norford and S. B. Leeb, "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms," *Energy and Buildings*, vol. 24, no. 1, pp. 51–64, 1996.
- [35] H.-H. Chang, C.-L. Lin, and J.-K. Lee, "Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms," in *Computer supported cooperative work in design (cscwd), 2010 14th international conference on*, pp. 27–32, IEEE, 2010.
- [36] M. Figueiredo, A. De Almeida, and B. Ribeiro, "Home electrical signal disaggregation for Non-Intrusive Load Monitoring (nilm) systems," *Neurocomputing*, vol. 96, pp. 66–73, 2012.

- [37] 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,” in *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*, pp. 859–866, IEEE, 2011.
- [38] S. Mostafavi, J. Troxler, and R. W. Cox, “An unsupervised approach for disaggregating major loads in small commercial buildings,” in *2017 IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 1575–1581, IEEE, 2017.
- [39] O. Parson, S. Ghosh, M. Weal, and A. Rogers, “Non-intrusive load monitoring using prior models of general appliance types,” in *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- [40] J. Z. Kolter and T. Jaakkola, “Approximate inference in additive factorial hmms with application to energy disaggregation,” in *Artificial intelligence and statistics*, pp. 1472–1482, 2012.
- [41] M. J. Johnson and A. S. Willsky, “Bayesian nonparametric hidden semi-markov models,” *Journal of Machine Learning Research*, vol. 14, no. Feb, pp. 673–701, 2013.
- [42] J. Kelly and W. Knottenbelt, “Does disaggregated electricity feedback reduce domestic electricity consumption? a systematic review of the literature,” *arXiv preprint arXiv:1605.00962*, 2016.
- [43] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
- [44] A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. Chonov, M. Werner, and R. Steinmetz, “On the accuracy of appliance identification based on distributed load metering data,” in *2012 Sustainable Internet and ICT for Sustainability (SustainIT)*, pp. 1–9, IEEE, 2012.
- [45] “Pecan street inc. dataport.” <https://dataport.pecanstreet.org/>. Accessed: 2019-09-01.
- [46] O. Parson, G. Fisher, A. Hersey, N. Batra, J. Kelly, A. Singh, W. Knottenbelt, and A. Rogers, “Dataport and nilmtk: A building data set designed for non-intrusive load monitoring,” in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 210–214, IEEE, 2015.
- [47] N. Batra, O. Parson, M. Berges, A. Singh, and A. Rogers, “A comparison of non-intrusive load monitoring methods for commercial and residential buildings,” tech. rep.
- [48] G. W. Hart and A. T. Bouloutas, “Correcting dependent errors in sequences generated by finite-state processes,” *IEEE Transactions on information Theory*, vol. 39, no. 4, pp. 1249–1260, 1993.

- [49] L. R. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [50] W. R. Gilks and P. Wild, “Adaptive rejection sampling for gibbs sampling,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 41, no. 2, pp. 337–348, 1992.
- [51] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*. Chapman and Hall/CRC, 2013.
- [52] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, “An introduction to variational methods for graphical models,” *Machine learning*, vol. 37, no. 2, pp. 183–233, 1999.
- [53] Z. Ghahramani and M. I. Jordan, “Factorial hidden markov models,” in *Advances in Neural Information Processing Systems*, pp. 472–478, 1996.
- [54] J.-A. Monin Nylund, “Semi-markov modelling in a gibbssampling algorithm for nialm,” 2014.
- [55] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, “Unsupervised disaggregation of low frequency power measurements,” in *Proceedings of the 2011 SIAM international conference on data mining*, pp. 747–758, SIAM, 2011.
- [56] W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, “An extensible approach for non-intrusive load disaggregation with smart meter data,” *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3362–3372, 2016.
- [57] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, “The eco data set and the performance of non-intrusive load monitoring algorithms,” in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, pp. 80–89, ACM, 2014.
- [58] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, “Leveraging smart meter data to recognize home appliances,” in *2012 IEEE International Conference on Pervasive Computing and Communications*, pp. 190–197, IEEE, 2012.
- [59] M. Baranski and J. Voss, “Genetic algorithm for pattern detection in nialm systems,” in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 4, pp. 3462–3468, IEEE, 2004.
- [60] J. D. Scargle, “Bayesian blocks in two or more dimensions: Image segmentation and cluster analysis,” in *AIP Conference Proceedings*, vol. 617, pp. 163–173, AIP, 2002.
- [61] R. Killick, P. Fearnhead, and I. A. Eckley, “Optimal detection of changepoints with a linear computational cost,” *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.

- [62] P. H. Eilers and H. F. Boelens, “Baseline correction with asymmetric least squares smoothing,” *Leiden University Medical Centre Report*, vol. 1, no. 1, p. 5, 2005.
- [63] H. Wang and M. Song, “Ckmeans. 1d. dp: optimal k-means clustering in one dimension by dynamic programming,” *The R journal*, vol. 3, no. 2, p. 29, 2011.
- [64] W. Goetzler, “Variable refrigerant flow systems,” *Ashrae Journal*, vol. 49, no. 4, pp. 24–31, 2007.
- [65] I. Rish *et al.*, “An empirical study of the naive bayes classifier,” in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, pp. 41–46, 2001.
- [66] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, “Nilmtk: an open source toolkit for non-intrusive load monitoring,” in *Proceedings of the 5th international conference on Future energy systems*, pp. 265–276, ACM, 2014.
- [67] S. W. Makonin, *Real-time embedded low-frequency load disaggregation*. PhD thesis, Applied Sciences: School of Computing Science, 2014.
- [68] “A Kalman-filter spectral envelope preprocessor,”
- [69] R. G. Pratt, P. J. Balducci, C. Gerkenmeyer, S. Katipamula, M. C. Kintner-Meyer, T. F. Sanquist, K. P. Schneider, and T. J. Secret, “The smart grid: an estimation of the energy and co2 benefits,” tech. rep., Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2010.
- [70] L. Ljung, “Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems,” *IEEE Transactions on Automatic Control*, vol. 24, no. 1, pp. 36–50, 1979.
- [71] H. Dong, B. Wang, and C.-T. Lu, “Deep sparse coding based recursive disaggregation model for water conservation,” in *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.
- [72] B. Ellert, S. Makonin, and F. Popowich, “Appliance water disaggregation via non-intrusive load monitoring (nilm),” in *Smart City 360°*, pp. 455–467, Springer, 2016.
- [73] G. Cohn, S. Gupta, J. Froehlich, E. Larson, and S. N. Patel, “Gassense: Appliance-level, single-point sensing of gas activity in the home,” in *International Conference on Pervasive Computing*, pp. 265–282, Springer, 2010.