Non-Intrusive Disaggregation of Advanced Metering Infrastructure Signals for Demand-Side Management

A Dissertation
Presented to
The Academic Faculty

by

Christy Green

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
G. W. Woodruff School of Mechanical Engineering

Georgia Institute of Technology
December 2022

Non-Intrusive Disaggregation of Advanced Metering Infrastructure Signals for Demand-Side Management

Approved by:

Dr. Srinivas Garimella, Advisor
G.W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology

Dr. David Anderson
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Berdinus Bras
G.W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology

Dr. Satish Kumar
G.W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology

Dr. Daniel Molzahn
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: August 11, 2022

To my family

ACKNOWLEDGEMENTS

I would like to express gratitude to Dr. Srinivas Garimella, my advisor, for his encouragement and guidance throughout this process. He has challenged me to produce high-quality work and advocated on my behalf on numerous occasions. I am thankful for the opportunity he has provided me to be a part of the Sustainable Thermal Systems Laboratory. I would also like to thank Dr. David Anderson, Dr. Berdinus Bras, Dr. Satish Kumar, and Dr. Daniel Molzahn for serving on my committee. They have provided valuable feedback and guidance.

I would like to thank Oak Ridge National Laboratory for funding this project and everyone at ORNL and Southern Company for providing technical assistance along the way. I would also like to thank the current and former STSL members, particularly Dr. Bachir El Fil, Dr. Khoudor Keniar, Dr. Girish Kini, Dr. Sriram Chandrasekaran, and Matt Hughes for their assistance and camaraderie during this process. I cannot imagine accomplishing this without them.

Last, but not in any way least, I would like to thank my parents, sister, and friends for their support. They have been a constant source of encouragement and motivation and have kept me focused on my goal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

Symbols

A	adjacency matrix
D	diagonal matrix
G	Graph
L	graph Laplacian
P	Power
q	graph classifier threshold value
s	graph signal
$s*$	graph signal classifier
Thr	Threshold
V	graph node set

Greek Symbols

ρ	scaling factor
μ	mean
σ	standard deviation

Acronyms

AMI	advanced metering infrastructure
ARIMA	autoregressive integrated moving average
DR	demand response
DSM	demand-side management
ERWH	electric resistance water heater
GSP	graph signal processing

HPWH	heat pump water heater
HVAC	heating, ventilation, air conditioning
IDU	indoor unit
LSTM	long short-term memory
MAE	mean absolute error
MAPE	mean absolute percent error
MAAPE	mean arctangent absolute percent error
NILM	non-intrusive load monitoring
ODU	outdoor unit
RMSE	root mean square error
SVR	support vector regression

SUMMARY

As intermittent renewable energy generation resources become more prevalent, innovative ways to manage the electric grid are sought. In the past, much of the grid balancing effort has been focused on the supply side or on demand-side management of large commercial or industrial electricity customers. Today, with the increase in enabling technologies such as Internet-connected appliances, home energy management systems, and advanced metering infrastructure (AMI) smart meters, residential demand-side management is also a possibility. For a utility to assess the potential capacity of residential demand-side flexibility, power data from controllable appliances from a large sample of houses is required. These data may be collected by installing time- and cost-intensive monitoring equipment at every site, or, alternatively, by disaggregating the signals communicated to the utility by AMI meters.

In this study, non-intrusive load monitoring algorithms are used to disaggregate low-resolution real power signals from AMI smart meters. Disaggregation results using both supervised and unsupervised versions of a graph signal processing (GSP) -based algorithm are presented. The effects of varying key parameters in each GSP algorithm, including scaling factor, sequence, and classifier threshold are also presented, and limitations of the algorithm based on energy use patterns are discussed. F_{M} values greater than 0.8 were achieved for the electric resistance water heater and electric vehicle charger using the unsupervised GSP algorithm. The disaggregated signals are then used to develop energy forecasting models for predicting the load of controllable appliances over a given demand response period. ARIMA, SVR, and LSTM forecasting methods were evaluated

and compared to a baseline model developed using the mean hourly power draw values. The minimum MAAPE was achieved for the water heater, with an approximate range of $10 < MAAPE < 20$. The total energy flexibility of each appliance and the associated uncertainty of the combined disaggregation and forecast are characterized to assess the feasibility of this approach for demand-side management applications. The framework presented in this study may be used to characterize the ability of signals to be disaggregated from a larger dataset of AMI data, based on the whole-house signal characteristics. This analysis can aid grid managers in assessing the viability of selected devices, such as the water heater, for demand response activities.

CHAPTER 1. INTRODUCTION

Greenhouse gas emissions continue to be a concern worldwide, due to their impact on global climate change (EPA, 2022). The electricity sector contributes to 25% of greenhouse gas emissions in the U.S., second only to the transportation sector (Figure 1.1). Approximately 60% of electricity produced in the U.S. is generated using fossil fuel resources (EIA, 2020). Increased renewable energy generation has contributed to a shift away from fossil fuels and renewable energy resources are increasing every year; however, the pace of their expansion must increase substantially to have the desired effect on emissions reduction. Additionally, the introduction of these generation sources, particularly when installed in a non-dispatchable, decentralized manner, has a destabilizing effect of the electric grid. To address this issue, electric utilities may balance the grid by controlling the supply of electricity distributed to customers, or by managing the demand

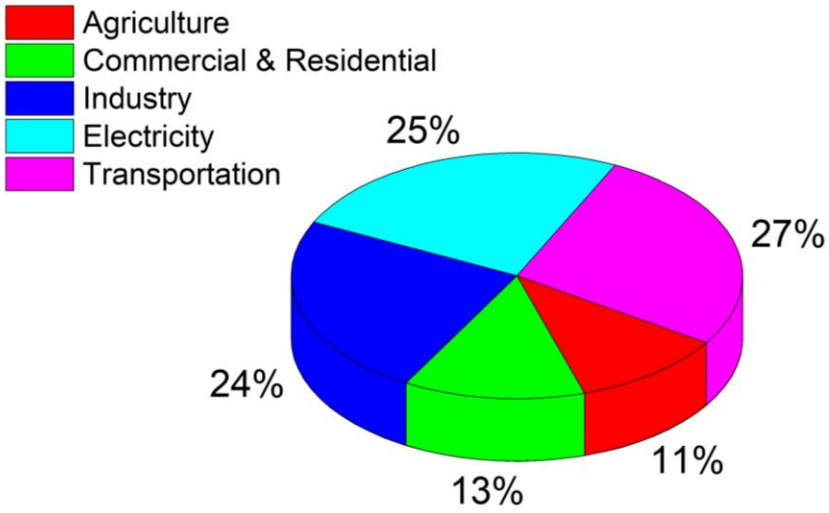

Figure 1.1: Greenhouse gas emissions by sector (EPA, 2022)

side of the electric grid through means such as demand response programs. In this study, a method of estimating the demand response potential of retail residential electricity customers is presented. This method uses advanced metering infrastructure (AMI) smart meter data and load disaggregation algorithms to detect appliances that may be controlled for demand-side management purposes.

1.1 Motivation

Demand-side management is an important component of the grid balancing strategy, particularly in response to the increase in electricity generation using renewable energy resources. Renewable energy technologies that are dependent on wind and solar are non-dispatchable, meaning that they cannot be turned on or off to meet changes in electricity demand. Moreover, the electricity generated by these resources fluctuates over the course of the day and is dependent on local weather patterns. Figure 1.2 presents electricity generation and total electricity demand over a day in January (EIA, 2022).

Solar generation on this day mostly occurred between 9:00AM and 7:00PM, while wind generation increased throughout the day. Electricity demand has two peaks during the day, one in the morning around 9:00AM and another in the evening around 7:00PM. This pattern is typical of winter days. A trough in the demand curve occurs during midday, coinciding with the peak solar electricity generation period. The combined effect of a solar peak and demand trough have the potential to result in greater difficulty balancing the grid, due to the steeper ramp in energy demand going into the evening peak. This effect will become more pronounced as solar generation capacity increases. Currently, these fluctuations in demand must be managed using mostly fossil fuel resources. Figure 1.3

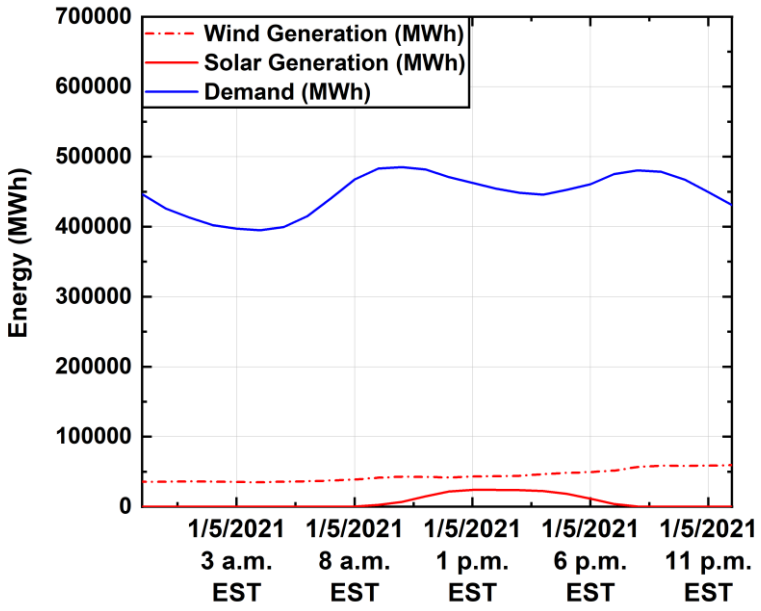

Figure 1.2: U.S. electricity generation and demand for a January day (EIA, 2022)

presents electricity generation by source in the United States over a day in January (EIA, 2022). Nuclear power generates approximately 19% of the electricity in the United States and forms the base generation capacity for the power grid. Hydro power accounts for 6.3% and is more dispatchable than other renewable energy resources, although its generating capacity has some variability dependent on water reservoir levels. Coal and natural gas power plants are the greatest generators of U.S. electricity, at 22% and 38%, respectively. It is most often the coal and natural gas power plants that must respond to sharp increases in electricity demand. Reducing the fluctuation in electricity demand by using demand shifting methods and energy storage reduces dependence on fossil fuel generation.

1.2 Advanced metering infrastructure (AMI)

As presented in Figure 1.4, advanced metering infrastructure (AMI) meters are rapidly replacing manual electric meters across the United States, with approximately 56%

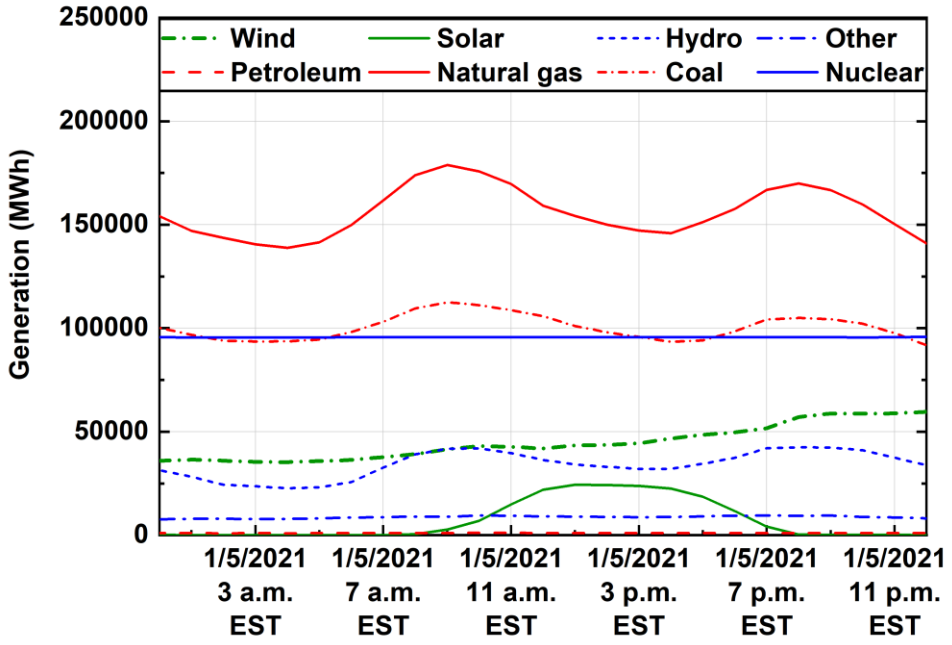

Figure 1.3: U.S. electricity generation by source for a January day (EIA, 2022)

of homes currently equipped with AMI meters (EIA, 2019). AMI meters allow power consumption data to be collected by the electric utility remotely over smaller intervals, typically every fifteen minutes or hour, rather than manually reading the total energy consumption from an analog meter once per billing cycle. In addition, AMI meters provide two-way communication between the meter and the electric utility. Using this two-way communication, the utility may not only monitor power quality and energy consumption remotely, but may also send information to the meter (DOE, 2016).

The total number of AMI meters deployed in the United States in 2018 for each sector is provided in Table 1.1. The residential sector has the greatest number of AMI installations, with residential AMI meters accounting for approximately 88% of the total number of AMI meters deployed. Currently, AMI data is used for remote energy consumption monitoring, communicating prices to consumers, monitoring outages, and

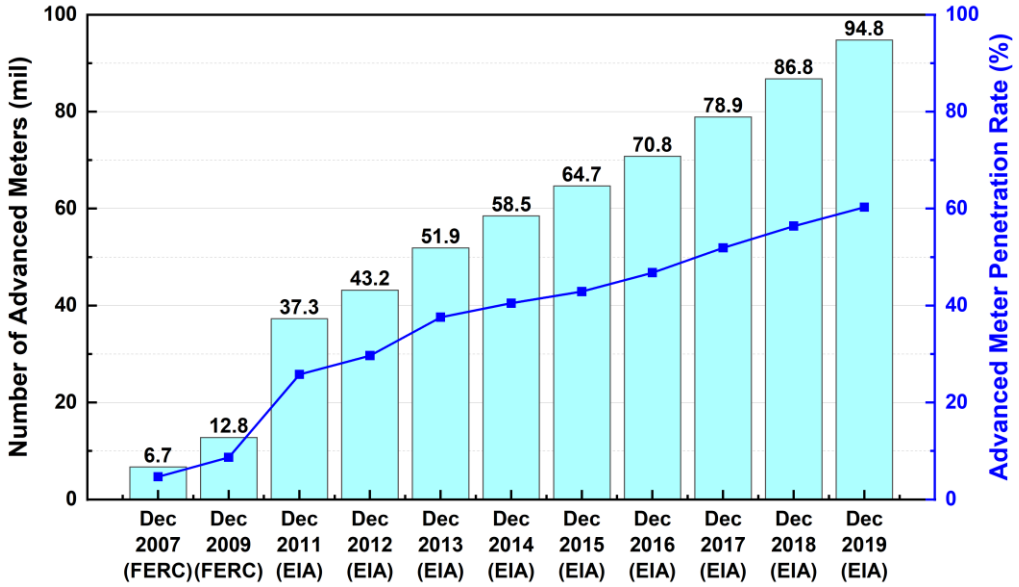

Figure 1.4: Annual change in number of AMI meters in U.S. (EIA, 2019)

monitoring power quality. Advances in data analytics are expected to allow for additional benefits from the large volume of available AMI data, including the addition of demand-side energy management features (DOE, 2016).

1.3 Demand-side management

As AMI meters with two-way communication, internet-connected appliances, and home energy management systems become more prevalent, more customers may enroll in demand-side energy management programs facilitated by electric utilities. Forecasting the

Table 1.1: Number of AMI installations by sector (DOE, 2016)

Number of AMI Installations				
Residential	**Commercial**	**Industrial**	**Transportation**	**Total**
76,498,388	9,932,993	411,287	1,489	86,844,157

energy consumption of individual appliances using AMI data will allow regional evaluation of demand-side management potential. Table 1.2 provides the potential peak demand energy savings for retail demand response programs for each retail sector (FERC, 2018).

Tools such as the Pacific Northwest National Laboratory Virtual Battery Assessment Tool exist but are based on survey data and may not be verified without extensive field monitoring. Development of a robust non-intrusive load monitoring (NILM) algorithm for low-resolution AMI data will allow for a demand response potential assessment of greater accuracy, by identifying specific devices within each house and monitoring their actual energy use, rather than basing such assessments on surveys. The development of high-accuracy AMI NILM algorithms could aid in verifying the results of tools such as the Virtual Battery Assessment Tool or be incorporated into these tools for improved accuracy.

1.4 Research needs and objectives

NILM research, until recently, has focused on the development of algorithms for devices installed locally on the electrical mains of a house. Such devices have the ability to monitor power at a high resolution and are not under the low-rate sampling constraints

Table 1.2: Potential peak demand savings from retail demand response programs by sector (FERC, 2018)

Potential Peak Demand Savings [MW]				
Residential	**Commercial**	**Industrial**	**Transportation**	**Total**
10,518	11,053	14,339	14	35,924

imposed by AMI metering infrastructure. Due to the high sampling rate, more features of the associated signal are capable of being used and devices with greater power draw are easily disaggregated. These algorithms primarily have difficulty disaggregating low-power devices such as televisions and lighting, and thus, many studies have focused on using algorithms targeting these devices and datasets that contain primarily signals of these devices. Recently, interest in NILM using very low-resolution AMI data has increased, as interest in assessing and deploying DR solutions has increased. NILM algorithms for the very low-resolution data sampled using AMI meters are emerging, and this study will assess the applicability of such algorithms to datasets that are more representative of single-family houses in the U.S.

The primary objective of this study is to understand the capabilities and limitations of NILM algorithms and energy forecasting methods when using very low-resolution data supplied by AMI meters. The resulting energy use estimates may be used to characterize demand response. To achieve this objective, the following tasks are conducted:

- A comparison is made among low-resolution NILM algorithms for AMI load disaggregation. The algorithms are benchmarked using publicly available datasets. Results include an assessment of algorithm performance on the publicly available Reference Energy Disaggregation Dataset (REDD) (Kolter and Johnson, 2011b) and Pecan Street dataset, data measured at the Oak Ridge National Laboratory Yarnell Station test facility, and proprietary data collected at a field testing location in Birmingham, Alabama.

- Characteristics of the whole-house signals and the appliance signals are analyzed in the context of NILM accuracy. Limitations on accuracy of these signal characteristics are discussed.
- Methods of forecasting using low-resolution data are compared, including autoregressive integrated moving average (ARIMA), support vector machines (SVM), and long short-term memory (LSTM) networks. The use of low-resolution data simulates data that may be available from applying NILM to AMI disaggregation. The effect of signal characteristics on forecastability is explored.

1.5 Dissertation outline

The dissertation is organized as follows:

- **Chapter 2** reviews the literature addressing non-intrusive load monitoring algorithms and energy forecasting methods. The state of the art in both supervised and unsupervised low-resolution NILM algorithms, as well as literature related to residential energy forecasting, is discussed.
- **Chapter 3** presents the application and comparison of a supervised and unsupervised graph signal processing algorithm. Discussion of the advantages and disadvantages is provided, and the effects of key parameters on disaggregation accuracy are presented.
- **Chapter 4** presents a comparison of energy forecasting methods, including ARIMA, SVR, and LSTM models. The models are compared to a baseline model of mean hourly values. The effect of energy use patterns on disaggregation accuracy are presented.

- **Chapter 5** discusses conclusions of the work and recommendations for future work, including extension of the study to a larger dataset of AMI data and the potential use of synthetic data for further investigation on the limitations of energy forecasting methods.

CHAPTER 2. LITERATURE REVIEW

Demand-side management programs encourage utility customers to shift load to accommodate the peaks and troughs associated with the mismatch between intermittent renewable generation and daily energy consumption patterns. In recent years, there has been an increase in the number of technologies that help to enable automated demand-side management, including Internet-connected smart appliances, home energy management systems (HEMS), advanced load forecasting algorithms, and the increased market penetration of advanced metering infrastructure (AMI) smart meters. Automated demand-side management allows customer devices to automatically respond to demand management activities. For example, a home energy management system may receive pricing signals from the utility that allow the automated curtailment of the energy consumption of select appliances during peak demand periods. In addition to localized home energy management technologies, initiatives such as OpenADR (Piette et al., 2009) and the CTA-2045 (EPRI, 2019) communication standard are increasing demand-side management capabilities at the grid scale by standardizing communications protocols. Initiatives such as the Pacific Northwest National Laboratory Virtual Battery Assessment Tool attempt to characterize the amount of energy available at a regional level for demand response activities (Somani et al., 2018). However, such tools use appliances, operational schedules, and efficiencies typical of a region rather than assessing the loads that are actually present.

Appliances typically considered for demand response include those with high energy consumption and inherent thermal energy storage or high energy consumption appliances

that would not inconvenience the end user if a change was made to their operational schedule. Callaway and Hiskens (2011) designate these loads as “fully responsive” and “nondisruptive”. Fully responsive loads enable system-level control across multiple time scales and include devices such as thermostatically controlled appliances and electric vehicles. Nondisruptive loads are those that have no perceived effect of end-use performance when controlled. The appliances with the greatest energy consumption in U.S. homes are HVAC systems and water heaters (EIA, 2015), and these loads are also considered fully responsive and nondisruptive when controlled within a specified temperature deadband. These HVAC systems and water heaters are inherently capable of providing thermal energy storage. Water heaters are capable of storing sensible heat within the water reservoir, while HVAC systems have sensible and latent thermal energy storage that is governed by the thermodynamic characteristics of the house. The HVAC system and water heater are the primary targets of this study, but other nondisruptive appliances that may be rescheduled without inconveniencing the customer are also evaluated, including dishwashers and clothes dryers.

The disaggregated appliance signals are initially used to characterize demand-side flexibility and the associated uncertainty for demand response event periods using historic data. After developing the disaggregation algorithm, a forecasting algorithm is developed to forecast the amount of energy available for demand response in a given look-ahead period using the disaggregated appliance signals. The disaggregation and forecasting algorithms are useful as stand-alone algorithms, in addition to their combined usefulness in predicting loads from disaggregated signals.

2.1 Very low-rate non-intrusive load disaggregation

To use AMI data for demand-side flexibility assessment, the signals of controllable appliances must be accurately disaggregated from the electric mains signal measured by the AMI meter. AMI data are typically sampled in 15-minute to 1-hour intervals. While much research has been conducted on NILM in the past few decades, little attention has been given to the problem of very low-rate NILM. Most NILM devices were developed to address disaggregation of high-rate data, with sampling rates on the order of kHz or MHz. Low-rate NILM is considered to address disaggregation for sampling rates of 1 to 60 seconds. Very-low rate NILM addresses disaggregation of signals sampled in the 15-minute to 1-hour range. With the increase in the number of AMI meters and interest in residential demand-side management, more research has been conducted in applying NILM algorithms to very-low rate NILM. Low sampling rates produce signals that are challenging to disaggregate, with many key signal features being lost in the downsampled signal or disguised by the baseload. There are approaches that address very low-resolution disaggregation, including the use of dynamic time warping (Elafoudi et al., 2014), factorial hidden Markov models (Kolter and Jaakkola, 2012), variations of k-nearest neighbors (Zhao et al., 2018b), and graph signal processing (Zhao et al., 2018a).

2.1.1 Supervised algorithms

Supervised NILM algorithms use labeled appliance power measurements to train a model. This approach is not ideal for an AMI-based application, due to the lack of labeled appliance data across large sets of residences. These algorithms may be classified as steady state or transient, with the steady-state algorithms using features like changes in real power

(Liao et al., 2014), reactive power (Hart, 1992), current, or voltage, and transient algorithms using transient signatures of the same features. Transient algorithms require high-frequency sampling on the order of kHz or MHz, while steady-state algorithms are suited to low-frequency sampling data, such as the < 1 Hz data available through AMI meters (He et al., 2018). Steady-state supervised NILM algorithms will be reviewed due to their suitability for low-resolution sampling frequency applications.

Supervised algorithms use sets of labeled measured data to train a model that is used to disaggregate the same set of appliances from the aggregate load in a later time period. Basu et al. (2015) used a time series distance-based multilabel classification approach to disaggregate residential loads. The water heater was disaggregated with the highest accuracy among all appliances tested at 10-minute and 1-hour sampling rates, which the authors attribute to the high level of energy consumption for water heating. The proposed algorithms performed better than a standard hidden Markov model applied to the same data, with F-measure scores ranging from 0.89 – 0.99 for the proposed algorithms. Zhao et al. (2018b) used a supervised approach with K-Nearest Neighbors (KNN) to disaggregate low-resolution smart meter data. The resulting algorithm could disaggregate up to 62% of the daily energy consumption using 15-minute and 60-minute sampling frequencies. Stankovic et al. (2016) used a decision tree method for disaggregating appliances from the main residential load signal. He et al. (2018) used a supervised version of graph signal processing (GSP) and compared the accuracy with that of a hidden Markov model (HMM) and dynamic time warping (DT) algorithm. The GSP algorithm disaggregated power signals with greater accuracy that either the HMM or DT algorithms.

2.1.2 Unsupervised algorithms

Unsupervised NILM algorithms do not require prior knowledge of appliance load signatures and operational states to disaggregate appliance loads. Most unsupervised algorithms use a version of Hidden Markov Models (HMMs), in which the signal is considered piecewise constant (Rahimpour et al., 2015). Zhao et al. (2016) compare a new unsupervised low-resolution NILM algorithm, graph signal processing (GSP), with several unsupervised and supervised HMM-based algorithms. The unsupervised graph signal processing algorithm outperformed the factorial hidden Markov model (FHMM) and HMM of Johnson and Willsky (2012), but did not perform better than their hidden state Markov model (HSMM). This HSMM was supervised, unlike the unsupervised GSP algorithm. The GSP algorithm also displayed greater accuracy than the two FHMM models of Aiad and Lee (2016).

2.1.3 Graph signal processing-based algorithms

Graph signal processing (GSP) – based approaches have been used extensively, with both supervised and unsupervised versions described in the literature. Supervised GSP is trained on labelled samples of individual appliance data, while unsupervised GSP uses filtering and clustering to disaggregate and label loads with no prior knowledge of particular load shapes. A semi-supervised approach to GSP may also be taken, in which a limited set of labelled data is used to disaggregate and classify a larger set of samples. He et al. (2018) compared the accuracy of a supervised version of graph signal processing (GSP) with that of a hidden Markov model (HMM) and dynamic time warping (DT) algorithm. The GSP algorithm disaggregated power signals with greater accuracy than

either the HMM or DT algorithms. Supervised GSP has also been demonstrated to outperform unsupervised GSP algorithms (Zhao et al., 2016) when tested on data from houses in the REDD data set (Kolter and Johnson, 2011a). Supervised or semi-supervised approaches may outperform unsupervised approaches because these algorithms have the advantage of informing the disaggregation algorithm with samples of complex operational cycles that may not be successfully accounted for in unsupervised approaches. Additionally, supervised approaches provide an opportunity to customize the algorithm to a set of appliance types, allowing for greater accuracy in disaggregation.

Appliances disaggregated in these studies include most appliance types available in a residence, from low-power, consistently operating appliances such as refrigerators to high-power, intermittent devices like microwave ovens. One appliance that has not been included in these disaggregation studies is the water heater. The high accuracy demonstrated by the GSP algorithm in disaggregating appliances with fairly consistent operational patterns and a range of power draws, such as the refrigerator and air conditioner, indicates that the water heater load may be a good candidate for disaggregation using this algorithm. Additionally, continuously variable loads and loads with more than two states of operation do not appear frequently in the literature.

2.2 Load forecasting

Load forecasting is not a new concept in electric grid management but has conventionally been applied to the problem of predicting demand in aggregate, rather than the demand of an individual device type. As more building energy data become available, however, research into the forecasting of individual appliances has increased. Zhao and

Magoulès (2012) categorize energy forecasting methods into five categories: 1) engineering methods, 2) statistical methods, 3) neural networks, 4) support vector machines, and 5) grey models. Each of these five categories is discussed in this section.

2.2.1 Engineering methods

Engineering methods are forecasting methods using complex models governed by the thermodynamics and energy consumption behavior of occupants on the whole building level, or for individual appliances. These methods require extensive knowledge of the building construction characteristics, building systems, and occupant behavior at each building site; therefore, this study will not consider forecasting methods involving engineering models.

2.2.2 Statistical methods

Statistical methods are regression models. These models may be autoregressive, in which they regress upon previous values of the time series, or multivariate regression models that correlate energy consumption with relevant site data such as local weather information. In the past, simple regression models, such as Autoregressive Moving Average (ARMA) or Autoregressive Integrated Moving Average (ARIMA) were used to forecast energy consumption for a region. These types of models are still used today, often in modified versions that improve modeling accuracy. Abdel-Aal and Al-Garni (1997) investigated the use of ARIMA models to predict a year of energy consumption based on data from the preceding five years. The ARIMA model was found to have greater accuracy than regression and abductive network machine learning methods. Abosedra et al. (2011) compared ARIMA models to Ordinary Least Square (OLS) and exponential smoothing,

finding that the ARIMA models provided the greatest accuracy. Twanabasu and Bremdal (2013) compared ARIMA to Artificial Neural Networks (ANN) and Support Vector Machines (SVM). While the ANN outperformed the ARIMA model, the ARIMA model was preferred due to its transparency. Abledu (2013) used Seasonal ARIMA (SARIMA) models to forecast one year of energy consumption using ten years of training data. Using the SARIMA model reduced the sample size needed to train an accurate model. The SARIMA method adds seasonal effects to the ARIMA model. Vagropoulos et al. (2016) compared a SARIMA model with ANN-based models and a SARIMA model with an exogenous factor (SARIMAX) to forecast photovoltaic (PV) power generation. Model performance was compared for winter, spring summer, and fall. The accuracies of the models exhibited seasonal dependencies, with each of the models having higher accuracies than the others in at least one season.

2.2.3 Neural networks

Neural network-based models have been studied extensively for use in time series forecasting. Network architectures that have some memory capability, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks are often used. The RNN is capable of remembering short-term time dependencies, while the LSTM network is capable of storing information regarding previous states for longer periods of time. Kim et al. (2018) use an LSTM network for month-ahead energy forecasting and report an absolute percent error for their prediction of 82.5%. Yan et al. (2019) compare an LSTM network to two hybrid LSTM algorithms. The hybrid LSTMs include convolutional neural network (CNN) LSTM and stationary wavelet transform (SWT) LSTM. The SWT-LSTM resulted in the highest accuracy. Deb et al. (2017) reviewed

current building energy forecasting methods and found that hybrid neural network models produced the most accurate forecast results.

2.2.4 Support vector machines

Support vector machines (SVM) are commonly used in classification machine learning problems, but may also be used for regression modeling. Dong et al. (2005) used SVM to forecast energy consumption of "landlord" energy consumption in Singapore, which is the energy consumption in a multi-family building not consumed by the tenants. Jain et al. (2014) also used SVM to forecast the energy consumption of multi-family buildings and examined the effect of data granularity on model accuracy. Xuemei et al. (2009) used least squares support vector machine and others, including Hong (2009) and Oğcu et al. (2012), have used hybrid models that combine SVM with other modeling methods to forecast energy consumption of large buildings. Kaytez et al. (2015) compared SVM to neural networks and regression models and found SVM to have the highest accuracy among the three methods.

2.2.5 Grey models

Grey models combine the engineering models with a statistical, neural network, or SVM model when the physics of a system are partially known. Because information regarding building construction or other parameters that may inform a physics-based model will be known in practice, grey models are excluded from consideration in this study.

2.3 Demand-side management resource assessment

Electric demand-side management programs seek to increase the stability and reliability of the electric grid by altering the electrical demand curve. Demand-side management may involve energy efficiency (EE), distributed generation (DG), and demand response (DR) activities. EE upgrades may be made by replacing equipment within a building or retrofitting the insulation and other materials used in construction of the building. DG includes on-site electricity generation, such as photovoltaic arrays, microturbines, and wind turbines, as well as any associated energy storage devices such as batteries. DR activities typically involve controlling one or more appliances in the home to facilitate electrical load shifting.

Using DR, electric load may be curtailed or shifted using peak shifting and clipping, conservation, load building, valley filling, and load flexibility. The effect of each of these approaches on the electricity demand curve is shown in Figure 2.1. Residential appliances may be used as a form of energy storage, or "virtual battery" to help achieve these goals,

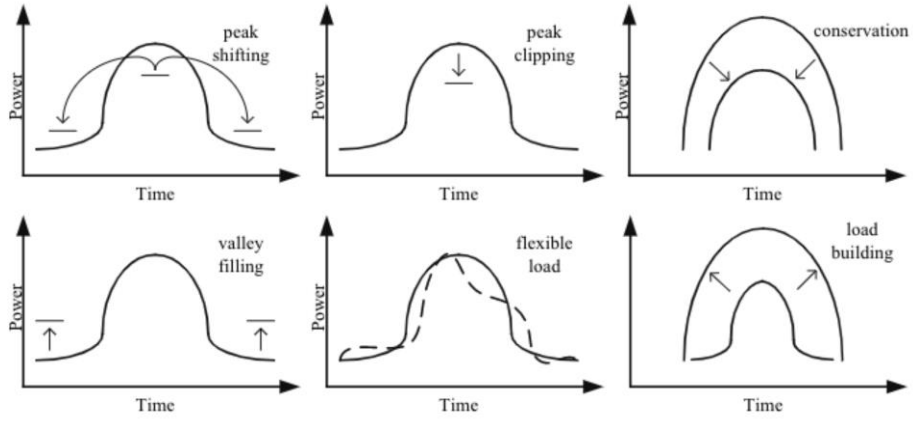

Figure 2.1: Electricity load shaping methods (Wu and Xia, 2017)

particularly that of peak shifting. Water heaters and HVAC systems have been demonstrated to provide significant load shifting capabilities (Green and Garimella, 2021). Malik et al. (2019) found that HVAC alone could account for a 9% reduction in peak demand, in some circumstances.

Pacific Northwest National Laboratory (PNNL) developed the Virtual Battery Assessment Tool (VBAT) for characterizing the flexibility of demand-side resources (Hao et al., 2018). This tool allows a user to select an appliance type and assess the potential for demand response load shifting across a region using that appliance. In this approach, appliance mixes across each region are assumed based on survey data, rather than empirically determined using AMI data. In this study, we investigate the possibility of AMI data disaggregation for improving estimates of demand response flexibility.

CHAPTER 3. NON-INTRUSIVE LOAD MONITORING

In this chapter, the performance of supervised and unsupervised GSP algorithms are analyzed. The unsupervised algorithm is particularly well-suited for use with AMI data, as it does not require a labeled data set collected on-site. However, supervised GSP-based algorithms have been demonstrated to exhibit slightly better performance in the literature, and could be applied in a semi-supervised application in the context of AMI.

3.1 Supervised GSP-based NILM

In this section, we apply an algorithm that has been benchmarked against other common algorithms using the REDD data set (He et al., 2018) to a data set including signals of an HVAC system and two water heaters collected at Oak Ridge National Laboratory in Knoxville, Tennessee (Fugate, 2019). Benchmarking of this algorithm using the REDD data set, as presented in the literature, demonstrated this algorithm to be superior in performance to other commonly used algorithms. Extending evaluation of the algorithm to water heaters and HVAC systems provides insight into the potential of the algorithm for disaggregating these types of loads. Additionally, variable parameters appearing in the algorithm are varied iteratively to demonstrate their effect on algorithm accuracy, and a potential optimization sequence for selecting these parameters is presented.

Supervised or semi-supervised approaches may outperform unsupervised approaches because these algorithms have the advantage of informing the disaggregation algorithm with samples of complex operational cycles that may not be successfully accounted for in unsupervised approaches. Additionally, supervised approaches provide an

opportunity to customize the algorithm to a set of appliance types, allowing for greater accuracy in disaggregation.

3.1.1 Methodology

In this section, the supervised GSP algorithm is described, noting the parameters relevant to this study that will be varied. The data set used in the study is then described, along with a description of the appliance load types present in the data set. The evaluation process and metrics used in the evaluation are then discussed.

3.1.1.1 Graph signal processing algorithm

Graph signal processing extends conventional Fourier analysis to discrete topologies described by graphs. The supervised graph signal processing algorithm was developed following the methodology described by He et al. (2018). In this approach, a graph $G = (V, A)$ is developed from a set of nodes V and adjacency matrix A. The set of nodes V correspond to the change in power ΔP between each time interval. The adjacency matrix is defined using a Gaussian kernel weighting function in which $V(i)$ and $V(j)$ correspond to power values in the series of nodes V, and ρ is a scaling factor (Equation (3.1)).

$$A_{i,j} = \exp\left(-\frac{(V(i)-V(j))^2}{\rho^2}\right) \tag{3.1}$$

A discrete signal s is indexed by the nodes of the graph G and is determined by classifying each node according to Equation (3.2), in which the values of Thr_m are equal to half the mean power draw of the appliance during operation.

$$s_m = \begin{cases} +1, & \left|\Delta P_i^m\right| \geq Thr_m \text{ and } i \leq n \\ -1, & \left|\Delta P_i^m\right| < Thr_m \text{ and } i \leq n \\ 0 & i > n \end{cases} \quad (3.2)$$

The total Lipschitz smoothness of the graph signals is defined using Equation (3.3), which has an approximate solution equal to $s^T Ls$. L is the graph Laplacian, defined as $L = D - A$, where D is the diagonal matrix that for $k = 1,\ldots,N$, $D_{k,k} = \sum_{n=1}^{N} A_{n,k}$.

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{i,j}\left(s_i - s_j\right)^2 \quad (3.3)$$

The disaggregation problem is posed as the minimization of $\left\|s^{m^T} Ls^m\right\|_2^2$ as in (Sandryhaila and Moura, 2013). This minimization problem has an approximate closed form solution (Equation (3.4)).

$$s^{m*} = L^{\#}_{n+1:N,n+1:N}\left(-s^{m^T}_{1:n}\right)L^T_{1:n,n+1:N} \quad (3.4)$$

This is known as the total graph classifier. Once s^{m*} is determined, the values of this vector are compared to a threshold value q to determine whether the appliance has changed state. If $s_i^{m^*} > q$, the appliance has undergone a state change and the value of $\left|\Delta P_i^{m^*}\right|$ is set to the mean value of P_i^m which is determined using the training sample. This load is then removed from the aggregate signal and the process is repeated for subsequent appliances. Just as in any other disaggregation algorithm, the accuracy of GSP depends partly on the intermittency of the load being disaggregated and the magnitude of power draw compared to the unmonitored or unlabeled loads in the house. In the case of intermittent loads, varying the time horizon of the training sample and disaggregation period may allow the inclusion of enough operational events for reasonably accurate disaggregation. Signals that

are lost in the house baseload, however, may be impossible to disaggregate. Additionally, the multiple operational states of some appliances (e.g., the operation of the heat pump and electric resistive elements in the hybrid water heater) may increase the difficulty of disaggregation. In supervised approaches, these states may be split into appliance load sub-components to improve disaggregation accuracy. In addition to these considerations, customizing the scaling factor, graph classifier threshold value, and disaggregation sequence to the load type that is being disaggregated may improve the accuracy of the algorithm. In this study, the effect of variations in these parameters and disaggregation sequences is investigated. A variety of load types is considered to investigate the impact of load type on disaggregation accuracy in the presence of other load combinations operating in real-world conditions.

3.1.1.2 Data collection

The data set used in this study was collected from a house in Knoxville, Tennessee operated by Oak Ridge National Laboratory for residential building energy analysis (Fugate, 2019). The data set consists of one-minute power measurements of an air-source heat pump water heater (HPWH), electric resistance water heater (ERWH), and an HVAC system. The HPWH is a hybrid water heater that may heat water using either an air-source heat pump, two electric resistive elements, or a combination of both. The electric resistive water heater heats water using two electric resistance elements. Water is drawn from the water heating units based on a water draw schedule developed in the Building America House Simulation Protocol (Hendron and Engebrecht, 2010) to simulate water consumption of an occupied house. The outdoor (HVAC ODU) and indoor (HVAC IDU) components of the HVAC system are monitored separately. The outdoor unit consists of a

variable speed compressor and fan, while the indoor unit contains a variable speed fan and supplementary resistive heating elements.

Data was downsampled from 1-minute to 15-minute intervals by averaging the power draw across the 15-minute interval (Equation (3.5)). Downsampling is performed to simulate data collected using AMI smart meters, which typically sample data at 15-minute to 1-hour intervals in the United States (Hansen et al., 2017). One disadvantage of this approach may be that taking the average over constant intervals does not reflect the sampling rate error present in AMI smart meters. These meters may have an error of up to +/- 8 seconds when sampling in the real world (Zhao et al., 2020); however, this only amounts to less than +/- 1% error in the sample time when considering the 15-minute interval.

The average power draw, when multiplied by a constant, provides the energy consumed during the interval, which is representative of the information supplied by an AMI smart meter (Equation (3.5)).

$$P = \frac{\sum_{n}^{N} P(n)}{N} \tag{3.5}$$

3.1.1.3 Load types

Residential loads may be divided into four categories, based on their operational characteristics: binary, finite state, continuously variable, and always on. Binary loads are either on or off, with the power draw displaying little to no variance when the appliance is on. Finite state appliances have two or more operational states, during which the magnitude and duration of power draw may vary in discrete steps. Continuously variable appliances have power draw magnitudes that vary continuously throughout the appliance operational

cycle. In addition to load type, each appliance has varying use intermittency. For example, devices such as an HVAC unit may be active much more often than a device such as a coffee maker. Loads often also operate simultaneously or overlap, resulting in aggregate load shapes that may be mistaken for the incorrect appliance. The characteristics of loads present in the house, along with the occupant usage schedules, impact the ability of the supervised graph signal processing algorithm to accurately disaggregate a load.

In this study, appliances representing a variety of load types are disaggregated from an aggregate signal. The air-source heat pump water heater (Figure 3.1(a)) has two finite states, one representing the operational cycle of the heat pump and one representing the operation of electric resistive elements within the tank. The electric resistive water heater (ERWH) (Figure 3.1(b)) is a binary load that may be either on or off with a constant power draw during operation. The operational cycle of the ERWH is only minutes long; thus, the magnitude of power draw for the downsampled signal varies considerably from one cycle to another. The indoor component of the HVAC system (Figure 3.1(c)) has two states, the first represents the operation of the fan and other miscellaneous small loads such as control equipment and is always on and continuously variable. The second state represents the operation of the electric resistive supplementary heating elements and is a binary load. The outdoor component of the HVAC unit (Figure 3.1(d)) is a continuously variable load that is always on. It should be noted that while both the indoor and outdoor units may be classified as "always on," the main components of the systems (compressor and fans) cycle on and off in response to space-conditioning demand.

The power draw of these appliances may vary by several orders of magnitude, with the lowest power draw around 10 W. This level of power draw is most likely attributed to

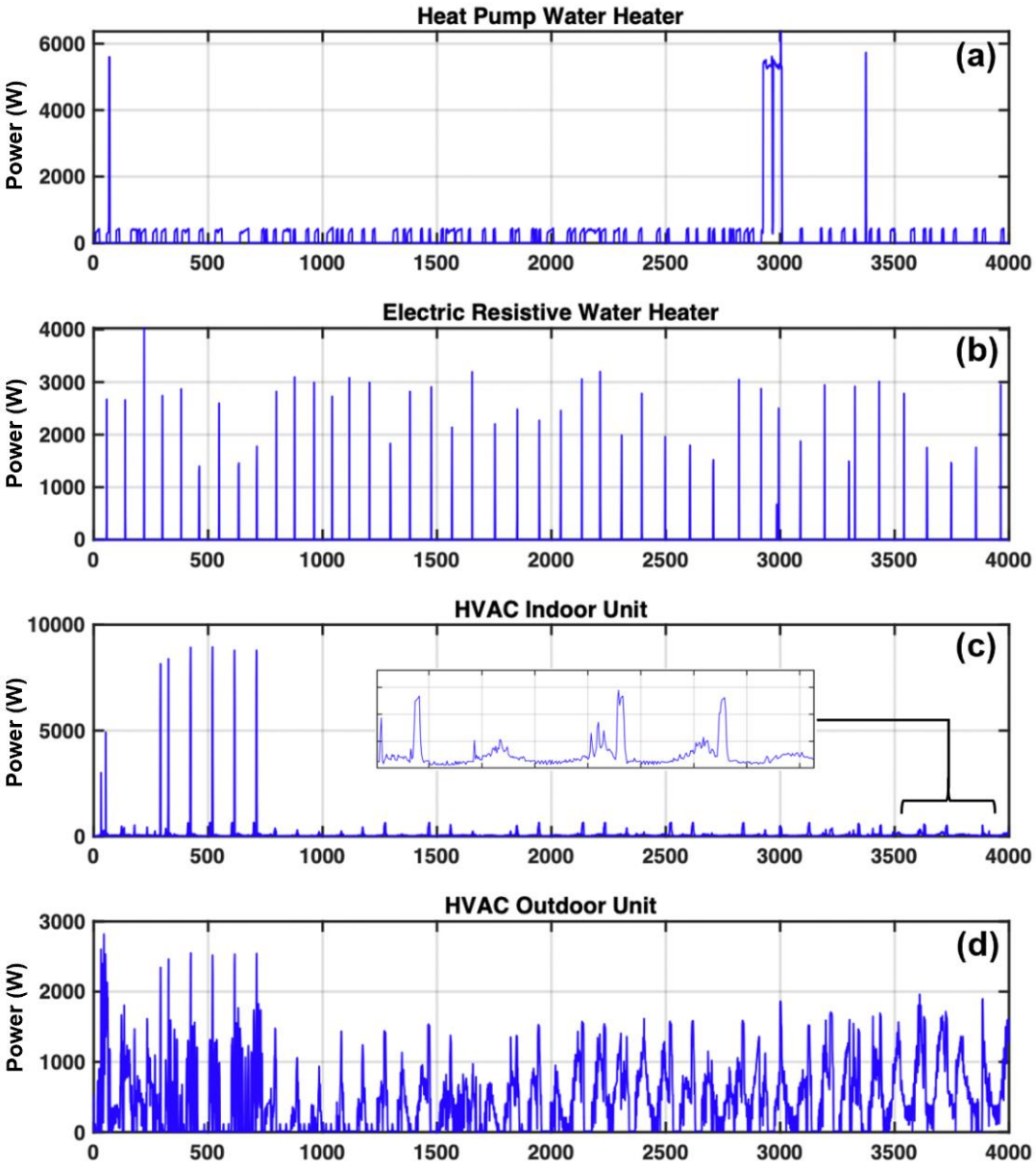

Figure 3.1: Load types present in data set

control equipment and parasitic loads; however, because the individual components are not monitored, this cannot be confirmed. Thus, the entire signal is included for disaggregation and these loads appear as "always on". The disaggregation capability of supervised GSP on these loads is evaluated with the goal of understanding the algorithm parameters best suited to individual load types for improved disaggregation accuracy.

3.1.1.4 Parameter variation and sequencing

The effects of varying scaling factor, classifier threshold values, and disaggregation sequence on algorithm accuracy were investigated. Two weeks of data sampled at 15-minute intervals were used for the supervised appliance signals and applied to the disaggregation of a two-week-long aggregate signal. The signal was comprised of the four loads described in Section 3.3. The aggregate signal was constructed from the individual appliance signal components, rather than using the whole-house aggregate signal, to eliminate baseload. Baseload was excluded to eliminate the effect of its variation on the disaggregation outcomes. In practice, baseload may be estimated by manually disabling as many appliances as possible and measuring the resulting load (baseload) for some period of time, or by assuming that appliances used during a certain time of day, such as early morning hours, may be attributed to baseload (Zhao et al., 2020). This measured baseload value may then be removed from the aggregate signal for the entire disaggregation period. In this study, we assume an ideal estimation and removal of baseload, to remove any interference of variations in baseload on algorithm performance. The intention of this approach is to highlight the effect of parameter variation, rather than demonstrate the effectiveness of the algorithm, which was previously established in He et al. (2018).

To understand the relationship between load type and scaling factor, the scaling factor was varied from 20 to 200 in increments of 20. All loads were disaggregated simultaneously to control for the potential effects of disaggregation sequencing on algorithm accuracy. Graph signal classifier threshold values were also held constant. This analysis provided insight on the effect of scaling factor for each load type when in the presence of other loads.

After observing the factors that contribute to the choice of a set of optimal scaling factors, disaggregation sequences were evaluated to determine their effect on disaggregation accuracy. Six sequences were investigated. Table 3.1 presents three of the sequences, the other three sequences are the same as those presented but in reverse order (least to greatest).

Table 3.1: Disaggregation sequencing

Disaggregation Sequence	Greatest % Total Energy	Greatest Max Power	Greatest Operational Time
1	HVAC ODU	HVAC IDU	HVAC IDU
2	HVAC IDU	ERWH	HVAC ODU
3	ERWH	HVAC ODU	HPWH
4	HPWH	HPWH	ERWH

Graph signal classifier threshold values were then evaluated using the two values, 0 and 0.5, presented in the literature, as well as an intermediate value of 0.25 for the electric resistance water heater, to determine the effect of the threshold value on disaggregation accuracy for each appliance type.

3.1.1.5 Metrics

F-measure and percent error in total energy assigned were used to determine the accuracy of the algorithm. F-measure (Equation (3.6)) is commonly used as a measure of accuracy in binary classification problems and is frequently used in the NILM literature (Batra et al., 2014; Stankovic et al., 2014; Zhao et al., 2015; Zhao et al., 2020). F-measure is the harmonic mean of precision and recall.

$$F-measure = 2\frac{PR \cdot RE}{PR + RE} \tag{3.6}$$

Precision represents the ratio of times an appliance was correctly predicted to be on (true positive results) to the total number of times it was predicted to be on, the sum of both true and false positive results (Equation (3.7)). *TP* is the number of true positives, which indicates the number of times the appliance was predicted to be on and was actually on. *FP* is the number of false positives, which is the number of times the appliance was predicted to be on but was actually off.

$$PR = \frac{TP}{TP + FP} \tag{3.7}$$

Recall is the ratio of true positives to the total number of times the appliance was actually operational, represented by the sum of true positives and false negatives (*FN*) (Equation (3.8)). False negatives are instances in which the algorithm predicted the load to be off when it was actually on.

$$RE = \frac{TP}{TP + FN} \tag{3.8}$$

The highest possible value of F-measure is 1.0, which indicates perfect precision and recall. The lowest possible value is zero, which will occur if either precision or recall have a value of zero. Williams (2021) demonstrated the effect of the ratio of positive to negative events on the F-measure. It was shown that the precision-recall curve varies significantly with changes in this ratio, making it difficult to compare problems with different class ratios. This effect is important to consider when comparing F-measure values for appliance disaggregation, as class ratios vary significantly from one appliance to another. F-measure should only be compared against F-measure values of the same signal, to ensure a consistent class ratio, and this a consistent basis for comparison.

The percent error of total energy assigned was calculated using Equation (3.9). This metric was used in lieu of the "error in total energy assigned" metric presented in (Batra et al., 2014), as the use of percent error rather than absolute magnitude error allows for direct comparison among appliances that may have very different power draw magnitudes. For example, an error in total energy assigned of +/- 100 W for an appliance with a mean power draw of 400 W would appear to be superior to an error of +/- 500 W in an appliance with a mean power draw of 5000 W. However, when considering these errors as percentages, we see that this is a +/- 25% error for the 400 W appliance and a +/- 10% error for the 5000 W appliance. Thus, the larger absolute error does not indicate superior performance.

$$TPE = \frac{\sum_n^N \hat{P}(n) - \sum_n^N P(n)}{\sum_n^N P(n)} \tag{3.9}$$

3.1.1.6 Algorithm Comparison

The algorithm used in this study was previously presented in He et al. (2018), where the algorithm was benchmarked against hidden Markov model and dynamic time warping based disaggregation algorithms using the REDD data set. In Zhao et al. (2018a), an unsupervised version of the algorithm was also demonstrated. Figure 3.2 presents a comparison of the performance of each of these algorithms across a set of appliances from the REDD data set, compiled from the results of the application of these algorithms found in the literature. Supervised GSP performed significantly better than other algorithms in disaggregating the signals of devices with highly intermittent power draw, such as the stove and the dishwasher.

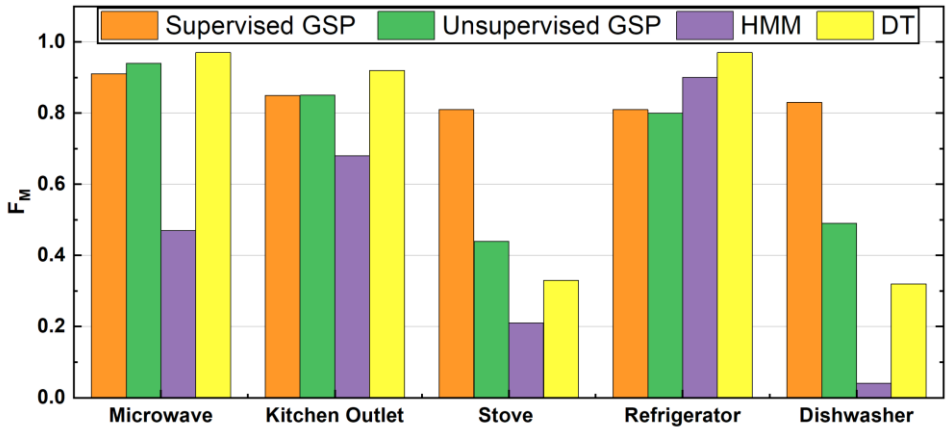

Figure 3.2: Supervised GSP benchmarking

3.1.1.7 Experimental execution time

The graph signal processing algorithm was written using MATLAB 2021a (The MathWorks, 2021). Each appliance disaggregation with two weeks (1344) training sample and four weeks (2688) testing sample required approximately 24 seconds on a MacBook Pro with 2.4 GHz Intel i9 processor and 32GB 2400 MHz DDR4 RAM.

3.1.2 Results and Discussion

In this section, the following results are presented: (1) effect of scaling factor, (2) effect of load disaggregation sequencing, (3) effect of graph signal classifier threshold values, and (4) recommendations for metrics to evaluate the accuracy of disaggregation.

3.1.2.1 Scaling factor

The scaling factor of the Gaussian kernel weighting function used to develop the adjacency matrix was varied to understand its sensitivity to the load type being disaggregated. In this scenario, all loads were disaggregated concurrently from the same mains signal to isolate the effect of scaling factor on disaggregation accuracy. A range of scaling factors from 10

to 200 was applied to the GSP algorithm and the resulting accuracy, defined by F-measure, was evaluated. He et al. (2018) used an iterative approach to determine the optimal scaling factor in the original development of the algorithm presented in this study. Larger scaling factors result in more non-zero values in the adjacency matrix, which could promote higher accuracy for appliances with higher operational frequency. Figure 3.3(a) presents the cumulative F-measure across appliances for each of the scaling factors tested. In terms of cumulative F-measure, a scaling factor of 100 was the best choice for this data set. However, individual appliances may benefit from a customized scaling factor. For example, in Figure 3.3(b), it can be seen that the electric resistive water heater disaggregation exhibits the best performance with a scaling factor of 20. A lower scaling factor results in fewer highly weighted edges in the associated appliance adjacency matrix, which helps to disaggregate loads that may have high-magnitude power draws with highly intermittent operational frequency. In the case of the electric resistive water heater, the

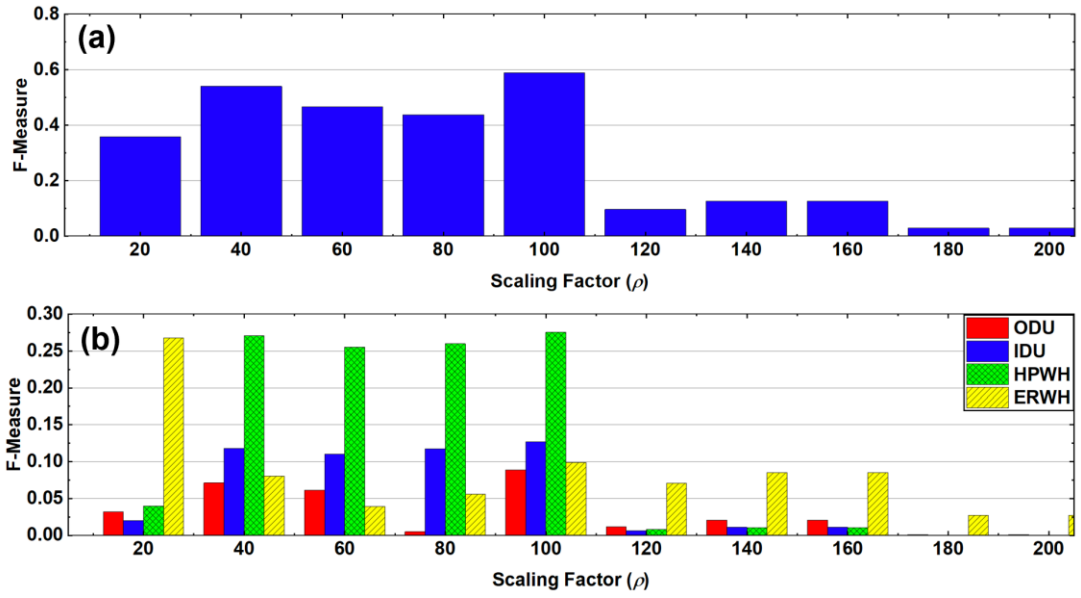

Figure 3.3: Disaggregation accuracy at various scaling factor values, (a) summed F-measure, (b) F-measure by load type

power draw magnitude is similar to that of several other loads, including the compressor of the outdoor HVAC unit, the electric heating elements in the indoor HVAC unit, and the electric heating elements in the heat pump water heater. This similarity, combined with infrequent operation, results in a signal that is difficult to disaggregate from the mains signal. By reducing the scaling factor, the number of non-zero events included in the adjacency matrix is reduced, which results in a graph that is more representative of the intermittent operation of the ERWH. Conversely, appliances with more frequent operational cycles perform best when disaggregated with a scaling factor between 40 and 100.

3.1.2.2 Disaggregation sequencing

Disaggregation sequencing can have an impact on accuracy. Typically, the baseload is removed from the mains signal first, followed by the appliances with the highest operational time. However, this approach does not always produce the highest accuracy for a targeted appliance. For DSM programs, controllable loads that may be adjusted locally by the consumer or remotely by a utility or other energy manager are of interest. Thus, the disaggregation strategy that maximizes the accuracy of the selected controllable loads is the most desirable. Six different load disaggregation sequences were considered: (1) highest percent total energy, (2) lowest percent total energy, (3) highest maximum power draw, (4) lowest maximum power draw, (5) highest total operational time, and (6) lowest total operational time. In each case, once a load is disaggregated, it is removed from the mains signal before subsequent loads are disaggregated. The accuracy of each appliance, in terms of F-measure, is presented in Figure 3.4 for the case in which all appliances are disaggregated concurrently (no sequencing). The electric resistance water heater had the

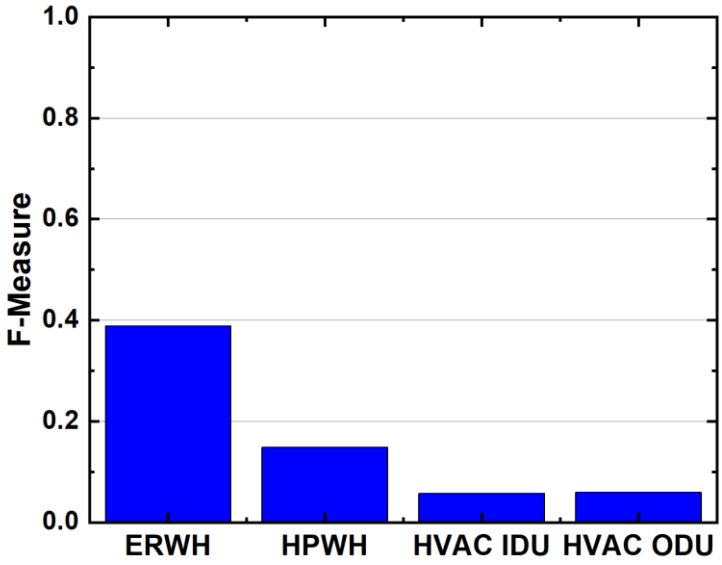

Figure 3.4: NILM accuracy when loads are disaggregated simultaneously, (ρ = 100, q = 0.5)

highest accuracy, which may be attributed to the fairly consistent operational cycle with high-magnitude power draws. The heat pump water heater had the second highest accuracy, due to frequent operation and longer operational cycles. The HVAC outdoor unit and HVAC indoor unit have similarly poor accuracies, with F-measure values less than 0.10.

Initially, the theoretical F-measure for each appliance when sequentially disaggregated was calculated by removing the measured signal from the aggregate signal before each sequential disaggregation, rather than removing the disaggregated signal separated using GSP. This simulates a situation in which all signals are perfectly disaggregated sequentially. Results for the theoretical sequential disaggregation are presented in Figure 3.5. The actual sequential disaggregation accuracy was then found by subtracting each disaggregated signal from the mains signal as it was disaggregated. The results of the sequential disaggregation are presented in Figure 3.6.

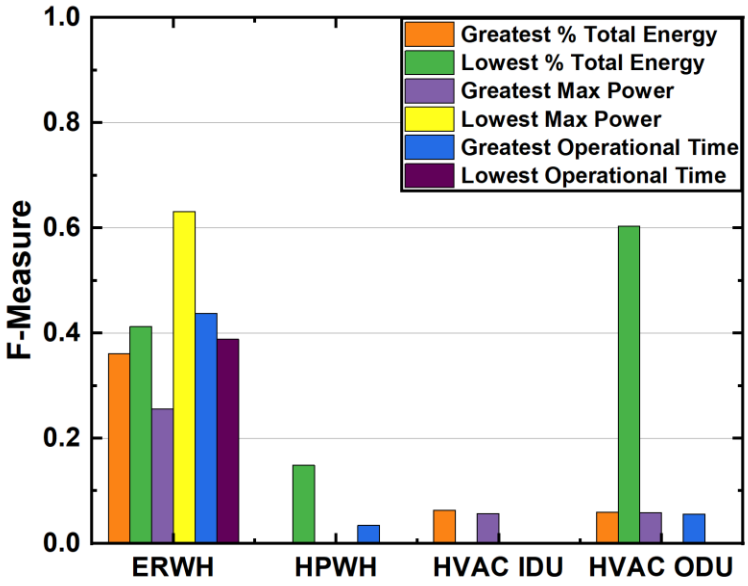

Figure 3.5: Theoretical NILM accuracy when loads are disaggregated sequentially

The theoretical sequential disaggregation accuracy shows that the electric resistance water heater demonstrates the most variability in accuracy according to disaggregation sequence, achieving a maximum F-measure of over 0.60 when disaggregation is performed in order of least maximum power draw. This is due to the ERWH ranking second in maximum power draw, meaning that in this sequence, the ERWH must only be disaggregated from the appliance with greatest maximum power draw - the HVAC indoor unit. Because there is no baseload included in this simulation, the last appliance to be disaggregated is, in theory, represented by the signal remaining after all other appliances are removed from the aggregate signal. However, in reality, error in each sequential disaggregation results in the final signal not being an exact representation of the desired final appliance signal. Therefore, disaggregation must be performed again. This final disaggregation is, in effect, a disaggregation of the appliance signal from itself. Interestingly, the ERWH disaggregation does not achieve the highest accuracy when being

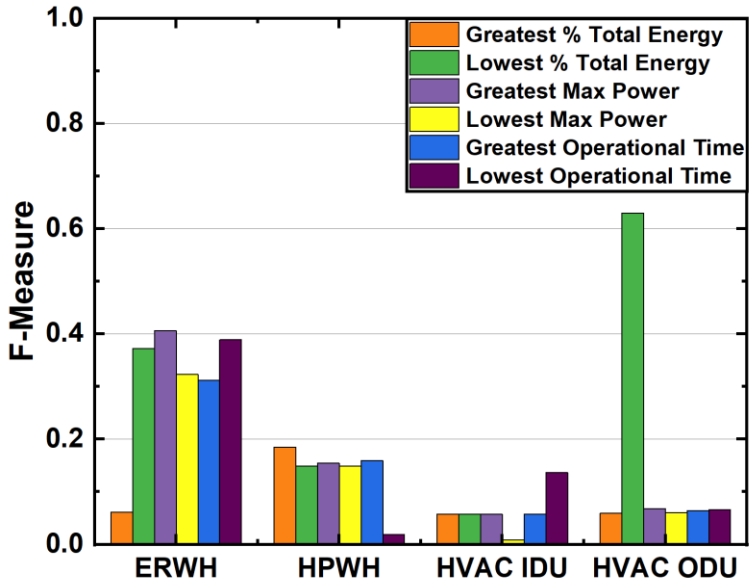

Figure 3.6: Actual NILM accuracy when loads are disaggregated sequentially

disaggregated only from itself, as is the case when disaggregating in order of greatest operational time to least. This may be because the presence of the HVAC IDU signal introduces more high-power-draw peaks to the graph that may account for differences in the training and testing period, resulting in a smoother graph and more correctly identified operational events.

The HVAC ODU also has much higher accuracy of $F_M = 0.62$, as compared to $F_M \approx 0.06$, only in the sequence in which appliances that consume the least amount of total energy are disaggregated first. In this sequence, the HVAC ODU is disaggregated last, meaning the signal is, theoretically, disaggregated only from itself. The ODU is continuously variable and almost always on, making it easy to confuse with other signals. Even when disaggregating this signal from itself, we see that the graph may not be completely smooth, due to variations in power draw that are influenced by local weather patterns as well as occupant behavior.

No accuracy improvement was expected for disaggregating the HVAC indoor unit in order of operational time in theory, but in practice, removing loads with the lowest operational time first results in an improvement in F-measure from $F_M = 0.06$ to $F_M = 0.14$. The HVAC indoor unit has the highest operational time of all the loads disaggregated. Because this HVAC indoor unit is a finite state appliance as a result of the fan and electric resistive element operation, the associated graph for this device has smoothness most similar to the graph of the aggregate signal when all loads are present, due to the increased incidence in power draw spikes, similar to those observed in the training sample of the HVAC indoor unit.

The heat pump water heater was expected to have poorer disaggregation performance in most sequencing scenarios but was fairly consistent in accuracy in practice. This is likely due to two factors: (1) inaccurate disaggregation of the HVAC IDU and ODU results in few true or false positive events, which has little effect on the aggregate signal, (2) the ERWH has few operational events; therefore, removing these also has a low effect on the aggregate signal.

3.1.2.3 Graph classifier threshold values

The threshold values of the graph signal classifier have been set at 0 and 0.5 in the literature (Sandryhaila and Moura, 2013; He et al., 2018), with the higher value of 0.5 chosen to ensure that only samples that are highly correlated with the training sample are classified together. Disaggregation of appliances with higher total operational time and more operational states correlates with better performance, according to F-measure, when a threshold value of 0 is used for graph classification (Figure 3.7).

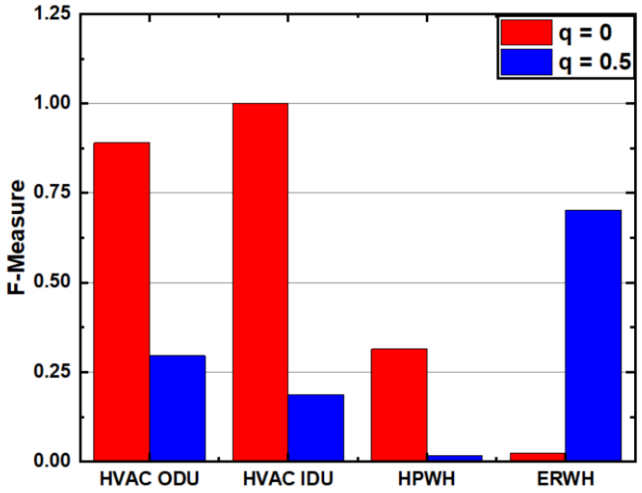

Figure 3.7: NILM accuracy at varying threshold values

Because the appliances have multiple states, requiring a greater degree of correlation between the testing and training sample excludes operational events that differ from the typical operational pattern. For example, in the HVAC indoor unit, the occasional operation of the electric resistive heating elements may not be included in either the training or testing sets, resulting in differences in the graphs that lead to the exclusion of operational events. Additionally, the variable nature of continuously variable appliances may result in graphs that exclude operational events. Reducing the threshold allows the algorithm to capture events for these appliances. Unfortunately, higher F-measures due to reduced threshold values do not always result in better performance from a DR perspective. For DR, the primary objective is to accurately estimate energy consumption within a given period. Reducing the threshold may result in greater event detection capability, but this does not always translate to better energy estimate accuracy (Table 3.2). It may be noted that for the value of $q = 0$, much greater percent errors occur. This is due to the difference in mean device power draw, which is substituted for "on" events during signal reconstruction, and the actual power draw. For loads with high variability between the

Table 3.2: Percent error at two graph classifier threshold values

Load	Percent Error (%) $q = 0$	$q = 0.5$
HVAC ODU	148%	65%
HVAC IDU	1267%	41%
HPWH	472%	99%
ERWH	8097%	-47%

training and testing intervals, the mean power value does not correctly represent the load and the inclusion of more operational events may result in less accuracy when estimating total energy consumption. Thus, it is critical to consider not only F_M, but also percent error, when evaluating the performance of the supervised GSP disaggregation algorithm. This is discussed further in Section 3.2.4.4.

In the case of binary loads like the electric resistance water heater, threshold values less than 0.5 reduce the F-measure as well as total energy accuracy. Reducing the threshold in this case allows for the inclusion of events that may belong to other appliances but resemble the operation of the ERWH, for example, the operation of the heat pump water heater electric resistance elements. The effect of threshold value on the ERWH disaggregation is presented in greater detail in Table 3.3. A variety of time horizon combinations are presented with disaggregation performed at threshold values of $q = 0.5$, 0.25, and 0. It is evident that the ERWH consistently achieves greater accuracy, as measured by F-measure at non-zero threshold values, with little variation between $q = 0.25$ and $q = 0.5$. A threshold value of $q = 0.25$ does outperform $q = 0.5$ for training windows of 4 weeks with testing windows of 3 weeks or greater. This may be due to greater variation in operation in the longer training signal, which would benefit from a more inclusive threshold value.

Table 3.3: Electric resistance water heater (ERWH) F-measure

ρ = 20, q = 0.5					
		Training Weeks			
		1	2	3	4
Testing Weeks	1	0.08	0.50	0.59	0.14
	2	0.34	0.55	0.48	0.28
	3	0.44	0.48	0.38	0.13
	4	0.44	0.39	0.20	0.11
ρ = 20, q = 0.25					
		Training Weeks			
		1	2	3	4
Testing Weeks	1	0.08	0.50	0.59	0.14
	2	0.34	0.55	0.48	0.28
	3	0.44	0.48	0.38	0.29
	4	0.44	0.39	0.36	0.24
ρ = 20, q = 0.0					
		Training Weeks			
		1	2	3	4
Testing Weeks	1	0.03	0.04	0.04	0.04
	2	0.03	0.04	0.04	0.04
	3	0.03	0.04	0.04	0.04
	4	0.04	0.04	0.04	0.03

3.1.2.4 Use of F-measure and mean power

When analyzing the performance of the disaggregation algorithm, F-measure is a commonly used metric that provides information about the number of times operational events are correctly identified by the algorithm.

It is important to note that substituting the mean power draw of an appliance for the points at which the appliance is classified as "on" introduces significant error in the total amount of energy consumed, even when the appliance disaggregation correctly identifies all operational events. If the mean of the power draw values during the training period is not an accurate estimate of the mean power draw during the disaggregation period, there may be times that an F-measure of 1.0 is obtained, indicating a perfect disaggregation of

the signal, but the substitution of mean power at all "on" points results in greater accuracy in terms of total percent energy error than a lower F-measure value. For example, the HVAC outdoor unit (Figure 3.8) may be perfectly disaggregated, but the continuously variable nature of its operation results in a load profile with a mean that may vary day-to-day. Thus, substituting the mean power value from the training sample for all appliance operational periods results in a high total percent error.

In addition, some appliances have multiple operational states, which may vary more frequently during either the testing or training periods. For example, the HVAC indoor unit (Figure 3.9) has two components, the fan, and the electric resistance elements of the supplemental heater. If the electric resistance elements operate during the training period but not the testing period, the mean power draw is much higher than the actual mean during the test period (e.g. $P = 900$ as opposed to $P = 100$), similar to the case of the continuously variable load. Differences in the mean power draw between the testing and training data sets may result in high percent error in the total energy consumption of an appliance across

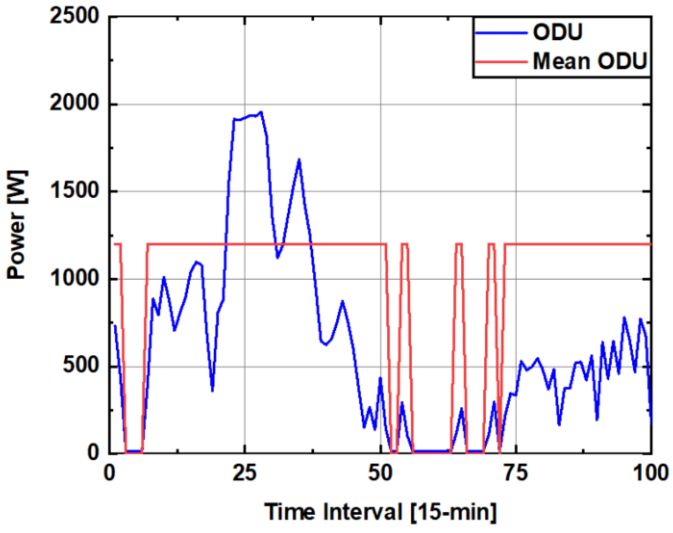

Figure 3.8: HVAC outdoor unit power

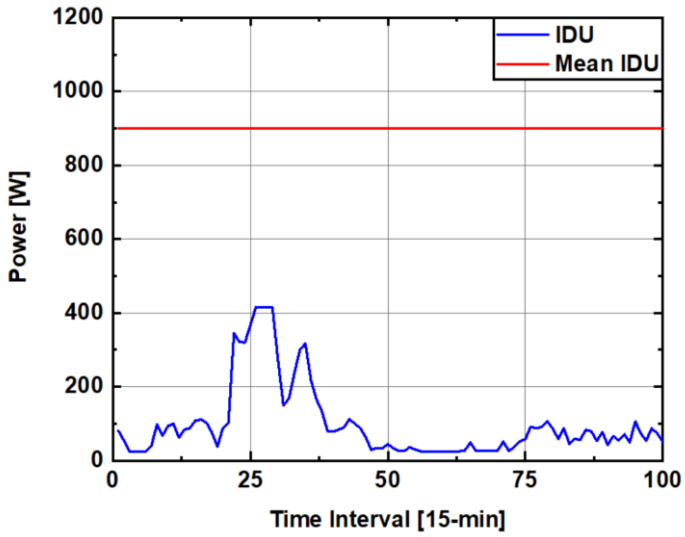

Figure 3.9: HVAC indoor unit power

the disaggregation horizon. Table 3.4 presents the total percent error for each of the four appliances disaggregated. When evaluating NILM algorithms for demand response applications, the primary concern is the correct estimation of the energy consumption of a controllable appliance over a given time horizon. This may not always be achieved by optimizing the algorithm for highest F-measure. In many cases, despite inaccuracy in event detection, greater accuracy for total energy consumption is achieved due to the use of the mean power value during all disaggregated operational events.

3.1.3 Supervised GSP algorithm conclusions

Table 3.4: Maximum possible accuracy using mean power value

Load	Mean Operational Power [W]	Percent Error (%)
HVAC ODU	960	99
HVAC IDU	1248	1267
HPWH	450	7
ERWH	2000	-69

The effects of varying scaling factor, classifier threshold value, and disaggregation sequencing order were investigated in this paper. Scaling factor may be adjusted to accommodate appliances with high magnitude power draw and brief, highly intermittent operation, such as electric resistive water heaters, to improve disaggregation accuracy. In addition, it was demonstrated that the disaggregation sequence has some effect on the ability of the algorithm to disaggregate appliances, particularly the electric resistance water heater, and ordering this disaggregation sequence to provide the greatest accuracy to appliances relevant to DSM may be performed to improve algorithm accuracy. The impact of graph classifier threshold values was also investigated, with results showing increased accuracy according to F-measure at lower threshold values for all appliances except the electric resistance water heater. Improvements in F-measure for these appliances, however, was not reflected in the total energy accuracy. Of all appliances disaggregated, the electric resistance water heater was the strongest candidate for supervised GSP disaggregation using AMI data, due to its consistent operational pattern and high-magnitude power draws. By tuning the scaling factor, graph classifier threshold, and sequencing, F-measures from 0 to 0.62 were demonstrated for the electric resistance water heater. Other controllable appliances with characteristics similar to the electric resistance water heater, such as single-speed HVAC units or clothes dryers with few operational modes, are expected to have similar performance and may be evaluated further in future work.

3.2 Unsupervised GSP-based NILM

In this section, the disaggregation of devices using an unsupervised GSP-based algorithm is investigated. Devices disaggregated include the water heater, HVAC system, and level 2 electric vehicle (EV) charger. The EV charger is included in this study due to

the presence of two EV chargers in the Birmingham, Alabama APC dataset. This device was not present in the ORNL Yarnell Station data used in the supervised study.

3.2.1 Methodology

In this section, the unsupervised graph signal processing algorithm is presented, along with the figures of merit used to evaluate algorithm performance.

3.2.1.1 Unsupervised GSP-based algorithm

We investigate the application of the unsupervised graph signal processing (GSP) algorithm developed by Zhao et al. (2016) to new datasets, with the application of a dynamic time warping approach for signal labelling. A brief explanation of the GSP algorithm is provided here. In this graph signal processing-based approach, the aggregate power signal is represented as a graph with nodes $V = \{v_1, v_2, \ldots, v_n\}$ and a weighted adjacency matrix *A*, which describes the similarity of the graph elements. Edges in the graph and their corresponding weights are defined by the adjacency matrix *A* (Equation (3.10)):

$$A_{i,j} = \exp\left(-\frac{(P(i)-P(j))^2}{\rho^2}\right) \tag{3.10}$$

This is a Gaussian kernel weighting function. The values *P* are the power draw values for each node, and ρ is a scaling factor. The smoothness of the graph signal is defined using Equation (3.11).

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{i,j}\left(s_j - s_i\right)^2 \tag{3.11}$$

This smoothness is equivalent to $s^T Ls$, where L is the graph Laplacian, calculated using Equation (3.12).

$$L = D - A \tag{3.12}$$

where D is a diagonal matrix given by

$$D_{i,i} = \sum_j A_{i,j} \tag{3.13}$$

To find the smoothest signal, an unconstrained quadratic minimization problem is formulated as:

$$\arg\min_s \left\| s^T Ls \right\|_2^2 \tag{3.14}$$

This optimization problem has been demonstrated (Boyd, 2004) to have the closed-form solution

$$s^* = L^{\#}_{2:N,2:N}(-s_1)L^T_{1,2:N} \tag{3.15}$$

Intervals of operational events are determined by identifying significant changes in power draw. To avoid detecting power fluctuations due to standby operation or small fluctuations in baseload, an initial threshold T_0 is set and all changes in power following the conditions of Equation (3.16) form an initial set of candidates Π.

$$\Delta P_i \in (-\infty, -T_0) \cup (T_0, \infty) \tag{3.16}$$

After identifying the initial set of event candidates, a graph is built for the candidate set Π. All values of ΔP_i are associated with nodes v_i on a graph and the adjacency matrix

for the candidate set is calculated. The value of s_1 in the graph signal s is set to 1 if $\Delta P_1 > T_0$ and -1 otherwise. All other values in s_i are initially set to 0. Then all samples similar to s_1 are clustered together. Clustering is performed by determining whether each s_j^* is greater than a constant value q. If $s_j^* > q$, the value is added to the first cluster and removed from the candidate set.

The mean and standard deviation of the first cluster are denoted by μ and σ, respectively. The relative standard deviation R of the cluster is then calculated for the cluster (Equation (3.17)). A lower value of R indicates a high-quality cluster, implying that the points will be closely grouped together around the mean.

$$R_i = \left| \frac{\sigma_i}{\mu_i} \right| \tag{3.17}$$

The mean values of the two clusters with the highest R, one from the positive clusters and one from the negative, are used to determine new threshold values, T_P and T_N. A new set of candidate values is selected using Equation (3.18) and the process is repeated.

$$\Pi = \Delta P \in (-\infty, T_N) \bigcup (T_P, \infty) \tag{3.18}$$

Clusters are formed until all samples have a designated cluster. After clusters are formed, the relative standard deviation R is compared to a heuristically determined value K. If $R_i > K$, the cluster is considered low-quality. The low-quality clusters are grouped to form a new set of candidate values and the clustering process is performed again, this time with a scaling factor half that of the previously used scaling factor. The cluster refining step may be performed multiple times until all clusters satisfy $R_i > K$. The clusters correspond to

either ON or OFF events of various appliances; thus, there should be an even number of clusters with each group of two forming a complete set of ON/OFF events for a given appliance. If the number of clusters at this stage is odd, smaller clusters are added into the most similar larger size clusters.

Feature matching is then performed to group the positive and negative clusters together to form signals for each appliance. Pairs of clusters are formed by selecting clusters with the closest absolute mean values. GSP is then used to match each event in the positive cluster to a corresponding event in the negative cluster. The magnitude-wise largest cluster is matched first. The set of magnitude differences between events in the positive cluster and those in the negative cluster is represented by Φ_M, and the time difference between events in the positive and negative clusters is represented by Φ_T. Two graphs are formed from the clusters. $G_M = \{V_M, A_M\}$ is formed using a set of nodes V_M indexed by the elements of Φ_M and the adjacency matrix calculated using Equation (3.19)

.

$$A_{M\ i,j} = \exp\left(-\frac{\left| dist\left(\Phi_{Mi}, \Phi_{Mj}\right)\right|^2}{\rho^2} \right) \tag{3.19}$$

$G_T = \{V_T, A_T\}$ is formed using a set of nodes V_T indexed by the elements of Φ_T and the adjacency matrix calculated using Equation (3.20).

$$A_{T\ i,j} = \exp\left(-\frac{\left| dist\left(\Phi_{Ti}, \Phi_{Tj}\right)\right|^2}{\rho^2} \right) \tag{3.20}$$

The graph signal that minimizes the global smoothness of each graph is calculated, yielding s_M^* and s_T^*. An optimization problem is then formulated (Equation (3.21)) to match the features, with coefficients α and β, where $\alpha + \beta = 1$, chosen heuristically to tradeoff between magnitude difference and time difference. Here, $i = 1, \ldots, n$ where n is the number of candidate events.

$$\arg\max_i \left\{ \alpha s_{M_i}^* + \beta s_{T_i}^* \right\} \tag{3.21}$$

Any unpaired edges are passed to the next magnitude-wise smaller cluster and matching is performed again.

After all positive and negative cluster events are matched, the result is a set of signals that may then be labeled by comparing the signals to signals in a labeled database. To make this comparison, we use a dynamic time warping (DTW) approach. The DTW algorithm measures the similarity between two time series by comparing features of the time series without considering the rate that time passes between these features.

The development of the labeled signal database is an important task in both semi-supervised and unsupervised load disaggregation. Local unsupervised approaches may rely on user interactions to label disaggregated signals or require the user to isolate loads by turning off other appliances to acquire a labeled signal with which the disaggregated load may be matched. Alternatively, a pre-existing load database may provide a labeled set of appliances that may be used as a general labeled training set during disaggregation. For example, the signal of an electric resistance water heater measured in a laboratory setting may be used as the labeled signal for a disaggregation algorithm applied to a set of houses. This approach may be assumed in AMI-based applications, as no local user interaction is

required. In this study, we use one example operational period for each of the load types disaggregated as labeled signals for signal matching.

3.2.1.2 Performance metrics

F-measure (F_M) and total percent error (*TPE*) were used to evaluate the unsupervised algorithm following the same methodology presented for the supervised algorithm. *TPE* is of primary concern for DR applications, as determining the amount of energy available in a given period is of greater importance than replicating the signal interval-by-interval. However, it is important to consider both metrics during evaluation. To demonstrate the relationship between F_M and *TPE*, two signals were modeled as two-state Markov sources with transition probabilities $P(S_{i+1}^1 = 1|S_i^1 = 1) = 0$ and $P(S_{i+1}^1 = 1|S_i^1 = 0) = 1$, to reflect the always-on nature of the whole-house load, and $P(S_{i+1}^2 = 1|S_i^2 = 1) = 0.97$ and $P(S_{i+1}^2 = 1|S_i^2 = 0) = 0.95$ for the device signal, consistent with transition probabilities specified by Zhao et al. (2016). The transition probabilities affect the degree of overlap between the two signals, which has some impact on disaggregation accuracy. However, by specifying the transition probabilities of Signal 1 to reflect an always-on load, these signals are always overlapping, removing any impact of this effect from the analysis. Signal 2 was then disaggregated from Signal 1 using the unsupervised GSP algorithm. Constant values for μ_1 and μ_2, and σ_2 were set, while σ_1 was varied iteratively from 50 to 800. Figure 3.10 presents the relationship between F_M and *TPE* for $0.35 < F_M < 1.0$ resulting from the simulation. For $F_M > 0.8$, F_M and percent error have a linear correlation; however, for $F_M < 0.8$, there is no correlation between F_M and *TPE*. Thus, improvements in F_M that fall in the range below 0.8 do not necessarily indicate

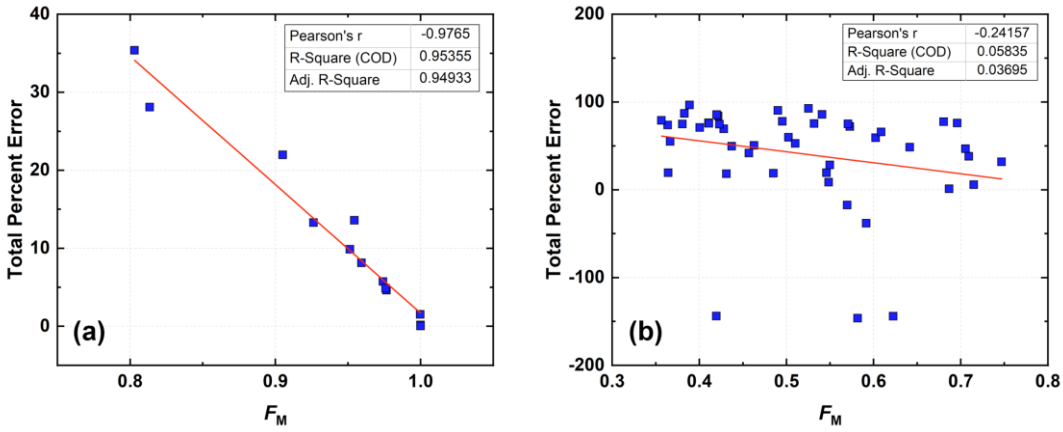

Figure 3.10: Correlation between F_M and percent error for (a) $F_M > 0.8$ and (b) $F_M < 0.8$

an improvement in disaggregation performance. Instead, it is likely that a disaggregated signal has been incorrectly matched with the target labeled signal in the signal database. While the misidentified signal may have greater agreement in terms of time of operation, the power draw attributed to the signal may be much different than the actual power draw.

3.2.1.3 Datasets

Publicly available datasets used in this study include the REDD, REFIT, and Pecan Street Dataport datasets (Kolter and Johnson, 2011b; Murray et al., 2017). In addition, a proprietary dataset collected at a site in Birmingham, Alabama is used. The REDD dataset is a freely available dataset composed of power data from eight houses in the Boston, Massachusetts area (Kolter and Johnson, 2011b). It is the most frequently cited dataset in NILM literature (Iqbal et al., 2021) and has been used often in the literature on graph signal processing-based load disaggregation (Liao et al., 2014; Stankovic et al., 2014; Zhao et al., 2015, 2016; He et al., 2018). This dataset is used to compare the algorithm with the performance results presented in Zhao et al. (2016) to verify agreement between the

performance of the GSP algorithms. The REFIT dataset is composed of data collected in the United Kingdom and appears frequently in GSP-based algorithm literature. In this study, we use the statistics of the REFIT dataset to demonstrate the characteristics of the datasets for which the GSP-based algorithm has been applied thus far.

The Pecan Street Dataport dataset is collected from sites in New York, Texas, Colorado, and California. Data from 25 houses in Texas and New York, and 23 houses in California are freely available to the public. The houses included in this dataset include solar PV generation. To compare the total energy consumption of these houses with the other houses included in this study, the solar generation was removed from the whole-house load by adding the amount of energy generated by solar PV to the amount of energy used from the electric grid.

The proprietary dataset was collected by Alabama Power Company (APC) in a neighborhood designed for smart grid research. The houses in this neighborhood are equipped with the same make and model of all major appliances. The appliances include a variety of load types, including two-state, finite-state, and continuously variable loads. Load types with more than two states are more difficult to disaggregate than those with two states due to greater variance in the signal. The disaggregation of appliance signals from this dataset from the mains signals of houses in the APC dataset, as well as the mains signals of the Dataport houses, was performed to investigate the limitations of the GSP-based approach. The houses included in this dataset have greater energy consumption and greater variance than those typically used in the literature and most of those included in the Pecan Street Dataport dataset. The mean and variance of the whole-house signal are important indicators of disaggregation potential. Whole-house signals with high means and

variances indicate a high degree of operational overlap between signals, and/or the presence of devices with many operational states. Such signals mask almost all individual signals in a house, making them impossible to differentiate. Figure 3.11(a) presents the energy consumption of the houses included in the APC dataset, along with the mean and variance of the whole-house signal as compared to the houses from the REDD and REFIT datasets appearing frequently in the literature. The whole-house signals of the APC houses have much greater means and variances than any of the houses included in these datasets. The houses in the APC dataset are more similar to the houses in the Pecan Street Dataport dataset (Figure 3.11(b)), with only two of the APC houses exceeding the energy consumption of those in the Pecan Street dataset. It is notable that the APC houses, while having similar or greater mean values, have lower variance than the houses in the Dataport dataset with similar energy use. The REDD dataset is used for comparing the algorithm with the performance results presented in Zhao et al. (2016) to ensure consistency, while the REFIT dataset is used only for additional comparison of signal characteristics between the APC dataset and those in the literature. The Pecan Street Dataport dataset is used to extend the results found using the APC dataset to a broader group of U.S. houses.

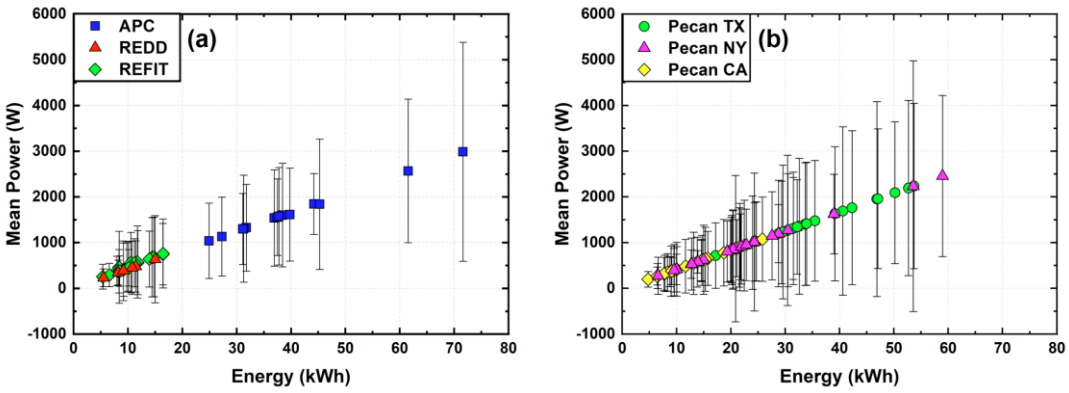

Figure 3.11: Energy use patterns of datasets

3.2.2 Results and discussion

In this section, the effect of splitting multi-state appliance signals into multiple binary-state signals within the labeled signal database is presented, along with a comparison of results across datasets, and a discussion of the implications of these findings for demand-side management.

3.2.2.1 Performance limits and signal splitting

Zhao et al. (2016) analyzed the performance limits of the unsupervised GSP algorithm by considering the disaggregation of two signals modeled as two-state Markov sources. The signals were given the transition probabilities of $P(S_{i+1}^1 = 1|S_i^1 = 1) = 0.95$, $P(S_{i+1}^1 = 1|S_i^1 = 0) = 0.03$, $P(S_{i+1}^2 = 1|S_i^2 = 1) = 0.94$, and $P(S_{i+1}^2 = 1|S_i^2 = 0) = 0.05$ to simulate transition probabilities commonly seen in devices within the datasets they analyzed. The "on" state of each signal follows a Gaussian distribution with a given mean (μ) and standard deviation (σ). By setting a mean value of signal 1, μ_1, and varying the variances of the two signals (σ_1^2, σ_2^2), it was determined that the minimum difference between the mean of signal 1 and the mean of signal 2 required to achieve $F_M > 0.8$ is given by

$$\Delta\mu = a\sigma_1 + b\sigma_2 + c \tag{3.22}$$

Thus, it is demonstrated that F_M is a function of the mean and variance of the signal targeted for disaggregation and the whole-house signal. Devices such as finite-state machines with more than two operational states and continuously variable devices have greater variance in operational power than two-state devices, which will negatively impact disaggregation. To reduce signal variance, the labeled signal may be split into multiple two-state devices.

This approach may also be applied to noisy signals by filtering the signal to reduce variance. An example of splitting a finite-state machine signal into multiple two-state signals is presented in Figure 3.12. In this example, a multi-state water heater signal (WH) is split into three two-state components: heat pump (WH 1), electric resistance stage 1 (WH 2), and electric resistance stage 2 (WH 3). The increased mean value and reduced variance of the WH 3 portion of the signal improve the chances of accurate disaggregation of this signal from the whole-house load. In this example, the amount of energy use included in the WH 3 portion of the signal consumes approximately 50% of the total device energy consumption.

Despite the reduced variance of a two-state signal, some two-state devices have noisy signals due to downsampling, such as an electric vehicle (EV) charger, which may be active for several full 15-minute intervals but then switch states to "off" midway through a 15-minute interval. The same approach taken in splitting a multi-state device into a two-state device may be applied to the problem of reducing variance for a noisy two-state

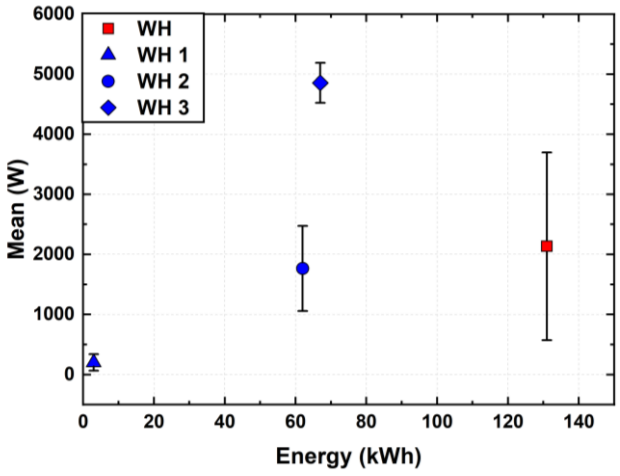

Figure 3.12: Filtering signal to transform a multi-state water heater signal into three signals with reduced variance

device. For the EV charger, a threshold of 7500 W is set, and all "on" periods lower than this threshold are removed from the signal. Figure 3.13(a) presents an example of the $\Delta\mu$ limit for an EV charger before filtering the signal. The $\Delta\mu$ value for the EV charger does not exceed the $\Delta\mu$ boundary as calculated using Equation (17). After filtering the signal, the $\Delta\mu$ value is now well outside the $\Delta\mu$ boundary, implying that the signal is easily detectable (Figure 3.13 (b)). The filtered portion of this EV signal captures approximately 94% of the total energy consumption contained in the original signal.

The signals of other devices, such as the continuously variable signals of some HVAC systems, may also be split into multiple two-state signals. However, the portion of these signals that is detectable may not represent much of the total energy consumption of the device. For example, filtering the HVAC signal in the APC dataset into a signal with a mean greater than 3000 W includes only 1% of the energy use from the original signal. While the signal may now be disaggregated, detecting a signal with such low energy consumption has reduced utility in assessing demand response capabilities.

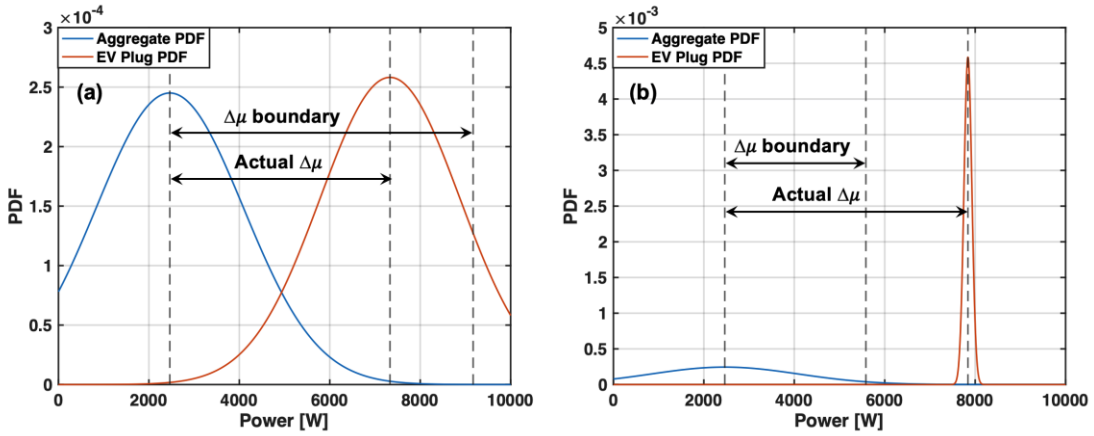

Figure 3.13: $\Delta\mu$ boundary for an unfiltered EV charger signal and the associated aggregate signal

3.2.2.2 Initial threshold values and scaling factor

In addition to the mean and variance of the signals, the scaling factor and initial threshold for candidate selection has an effect on disaggregation. In the literature, the initial threshold for candidate selection is set at a low value of 10 W. The purpose of the threshold is to exclude any parasitic loads that may be present in the whole-house signal. It is set to a low value to exclude these portions of the signal while allowing events corresponding to state changes of lower-power devices such as lighting to be included in the candidate set. This is practical for datasets such as REDD and REFIT, where devices with low signal magnitudes are being disaggregated. However, when appliances with greater power draw are targeted for disaggregation, these thresholds may be set higher to exclude lower magnitude devices from being included in the initial candidate set. This approach limits the events in the candidate set to those more similar to the targeted device, resulting in higher quality clusters after cluster refinement. For example, by changing the threshold $\pm T_0$ from 20 to 200, F_M of the EV charger increased from ~0.7 to ~0.99. In addition, changing the initial scaling factor was shown to improve disaggregation performance. Due to the dynamic nature of the scaling factor in this algorithm, Zhao et al. (2016) did not see an improvement in disaggregation accuracy as a result of modifying the scaling factor. However, the loads used in their demonstration had very low variances, ranging from 2 to 30 W. This resulted in a power range of 1000 to ~1300 W being investigated. It is possible that at much higher signal variances and magnitudes, the initial setting of scaling factor has an effect on disaggregation. For example, increasing scaling factor from 20 to 200 improved disaggregation of a 4800 W water heater signal from F_M = ~0.7 to ~0.8. While

changing the scaling factor has a much less pronounced effect than altering the initial threshold values, it does offer some improvement to disaggregation accuracy.

3.2.2.3 Results across datasets

The GSP algorithm was first applied to houses in the APC dataset and then extended to houses in the Dataport dataset. The devices targeted for disaggregation were the HVAC system, electric hybrid heat pump water heater, and electric vehicle (EV) charger. The disaggregation results for the initial, unfiltered signals used for signal identification are presented in Figure 3.14(a). Prior to splitting the labeled signals, no appliances exceeded the minimum required F-measure of 0.8. The signals were then split into several two-state signals and disaggregation was performed again. Figure 3.14(b) presents the accuracy of disaggregation after filtering the labeled signals. After splitting the signals, the second stage of the electric resistance water heater (WH 3), at ~5000 W, and the EV charger (EV 2), at ~8000 W, were capable of exceeding the minimum required accuracy. While the second stage of the electric resistance water heater accounts for 50%

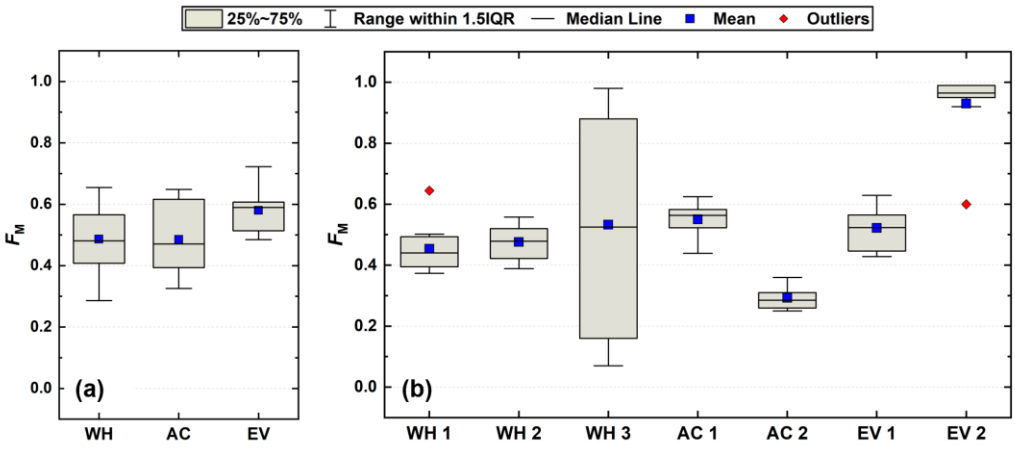

Figure 3.14: Device disaggregation accuracy for (a) unfiltered and (b) filtered labeled device signals

or less of the energy consumption of the hybrid heat pump water heater, it could account for a greater portion of the energy consumption of a typical electric resistance water heater that does not contain a heat pump element. The filtered EV charger signal accounted for ~96% of the total energy consumption represented by the original signal. Four of the fourteen houses included in the APC dataset used the electric resistive elements of the hybrid heat pump water heater during the trial period, and only two of the houses had active EV chargers. To extend the study to other houses in the neighborhood, two signals were modeled as two-state Markov sources with transition probabilities determined using the transition probabilities of the signals available in the dataset. For the whole-house load, $P(S_{i+1}^1 = 1|S_i^1 = 1) = 0$ and $P(S_{i+1}^1 = 1|S_i^1 = 0) = 1$, to reflect the always-on nature of the whole-house load. $P(S_{i+1}^2 = 1|S_i^2 = 1) = 0.99$ and $P(S_{i+1}^2 = 1|S_i^2 = 0) = 0.85$ for the EV charger and $P(S_{i+1}^2 = 1|S_i^2 = 1) = 0.99$, $P(S_{i+1}^2 = 1|S_i^2 = 0) = 0.50$ for the water heater. The mean and variance in the "on" state were determined from the mean and variance of the respective two-state signals. The mean and variance of the whole-house load was specified for each individual house in the dataset, and then the disaggregation of the device load from the whole-house load was evaluated. The water heater signal had a maximum F_M of ~0.8 for houses with lower variance and mean values, similar to those in the California portion of the Dataport dataset. Figure 3.15(a) presents the results of EV charger disaggregation in terms of F-measure. The EV charger demonstrates much greater accuracy than the water heater across all houses in California, New York, and the APC (Alabama) dataset. The Texas houses, while presenting lower total energy consumption than the Alabama houses, exhibited greater variance, resulting in greater difficulty when disaggregating the EV charger. Figure 3.15(b) presents the total percent error for EV

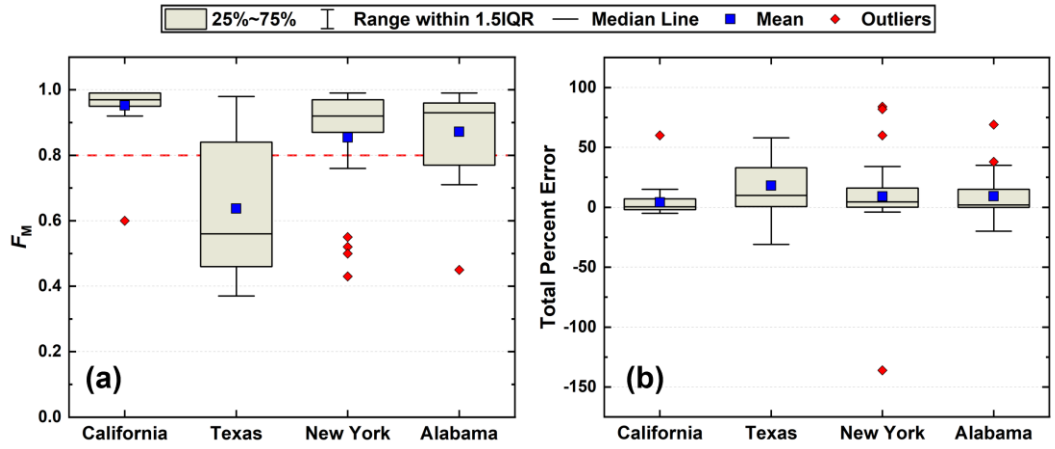

Figure 3.15: Electric vehicle disaggregation accuracy across APC and Pecan Street datasets

charger disaggregation across four states, with most houses falling within the ±25% range. This is a promising result for EV charger disaggregation, which may aid in characterizing the demand response capability of EV batteries for demand response activities using only AMI smart meter data.

Figure 3.16 presents the disaggregation accuracy of the third stage of the hybrid heat pump water heater signal (WH 3), in which both electric resistance elements are

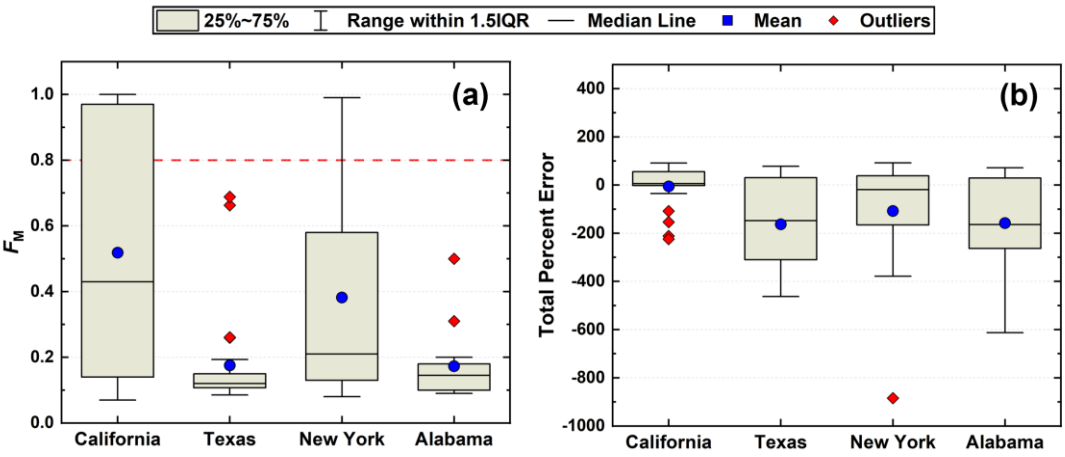

Figure 3.16: Water heater disaggregation accuracy across APC and Pecan Street datasets

operating. Separating this portion of the signal from the aggregate water heater signal increased the signal mean and reduced variance; however, these changes only made a significant difference on disaggregation accuracy in the California homes, which had low power draw.

3.2.2.4 Demand response implications

Results indicate that non-intrusive load monitoring of electric vehicle chargers using GSP-based disaggregation may be possible with a high degree of accuracy in some U.S. regions and disaggregating electric resistance water heater signals may also be possible, although with lower accuracy. Using EV energy use profiles developed by NREL for the study of uncoordinated residential EV charging, a profile of mean hourly energy consumption was generated (Figure 3.17(a)) (Muratori, 2017). Using this EV schedule, it is found that an average of ~2.67 kWh of energy is used by an EV during peak demand hours each day. The total number of EVs registered in each of the four states, as well as the number of publicly available EV ports is provided by the DOE Alternative Fuels Data Center (DOE, 2021). In addition to information on publicly available EV ports, the California Energy Commission (CEC) also provides information on private shared EV ports within California (CEC, 2022). Shared private EV chargers are those in locations such as apartment buildings or workplaces. No information on private EV chargers, such as those located in the garage of a single-family home, is provided. Therefore, the ratio of Level 2 chargers to total chargers, as determined using the available information, is applied to the total number of EVs registered in each state to determine the maximum number of vehicles charging on a Level 2 charger during peak demand response hours. The total amount of energy consumed by EVs charging on Level 2 chargers in each state is then

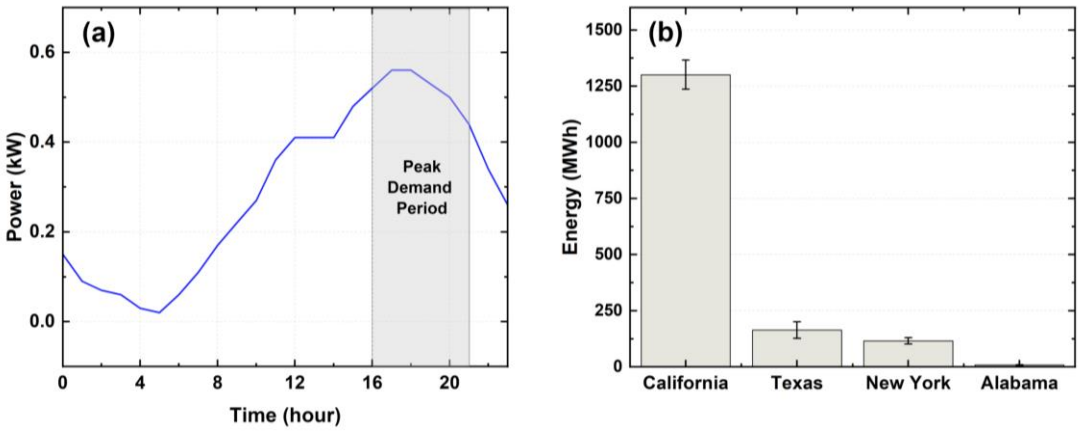

Figure 3.17: EV chargers (a) energy consumption during peak demand hours and (b) estimate of energy available for DR

determined and is presented in Figure 3.17(b). The state of California has the greatest potential for DR activity using EVs, with ~1300 MWh energy shifting potential during the peak demand period. A similar analysis was performed for the water heater. Scheduling information was found in the Building America House Simulation Protocol (Figure 3.18(a)) (Hendron and Engebrecht, 2010). The number of electric water heaters in each state and mean water heater energy consumption per household was determined using data

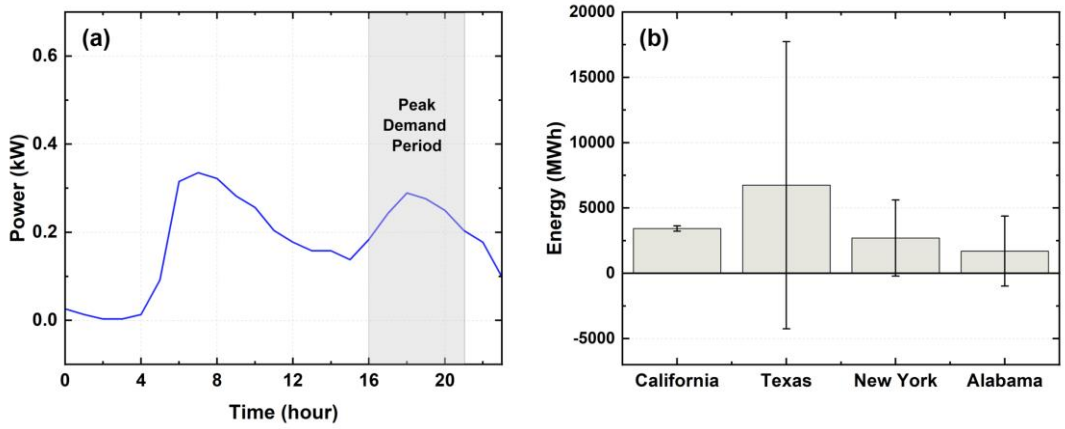

Figure 3.18: Water heater (a) energy consumption during peak demand hours and (b) estimate of energy available for DR

from the EIA Residential Energy Consumption Survey (EIA, 2015). Applying the prediction accuracies for water heater load disaggregation, the range of estimates in DR energy availability for each state using disaggregated signals is presented in Figure 3.18(b). Despite having lower water heater energy availability, California is once again the best candidate for DR estimation using GSP-based load disaggregation, due to the greater accuracy associated with load disaggregation in these houses.

3.2.3 Unsupervised GSP algorithm conclusions

An unsupervised GSP-based load disaggregation algorithm was applied to new datasets containing houses with more complex energy use profiles that are more representative of those in the United States than datasets presented previously in the GSP-based disaggregation literature. It was demonstrated that electric vehicle charger signals may be disaggregated with $F_M > 0.80$ and $TPE < 25\%$ from houses in the California, New York, and APC (Alabama) datasets. Electric resistance water heaters also had some success in disaggregation, with California being the best candidate for GSP-based disaggregation of water heaters. Further analysis may be performed on a larger AMI dataset to further investigate whole-house load characteristics by state and the associated impact on load disaggregation.

CHAPTER 4. ENERGY FORECASTING

Energy forecasts are an essential part of electric grid management, allowing grid managers to anticipate demand in advance and adjust the electricity supply or shape the demand curve accordingly. Due to an increase in technologies that enable demand-side management, such as smart appliances and home energy management systems, automated residential energy management is becoming a viable option for demand-side management (DSM) (Seo et al., 2016). In addition to advances in technologies that enable DSM, advanced metering infrastructure (AMI) meters are facilitating more interaction between customers and electric utilities. AMI meters are rapidly replacing manual electric meters across the United States, with approximately 56% of homes currently equipped with AMI meters (EIA, 2019). AMI meters allow power consumption data to be collected by the electric utility remotely over smaller intervals, typically every fifteen minutes or hour, rather than manually reading the total energy consumption once per billing cycle. In addition, AMI meters provide two-way communication between the meter and the electric utility. In this study, signals at a 1-hour resolution are used to simulate a signal that may be received by an electric utility from an AMI meter. These signals are used to forecast energy consumption of several appliances that may be considered for demand-side management.

Water heater, dishwasher, clothes dryer, and HVAC system energy consumption is forecasted in this study. These devices are chosen because they allow for load control that is responsive and does not interfere with the comfort of the end user when executed properly. The HVAC system and water heater have inherent thermal energy storage that facilitates load shifting without causing discomfort for the occupants, while operation of

the dishwasher and clothes dryer may simply be rescheduled to accommodate the needs of the electric grid. The HVAC system is considered to be one of the most effective sources of demand-side flexibility (Asadinejad et al., 2018; Malik et al., 2019). However, the energy consumption of these devices is often challenging to predict. Modern HVAC systems are often equipped with variable speed fans and compressors, which results in a continuously variable load shape. In addition, the power draw of these systems is sensitive to external factors such as ambient temperature, which may vary significantly from day to day. Modern HVAC systems also often include built-in energy management systems, whether it is a simple programmable thermostat or a more advanced device that learns occupancy patterns and preferences and adjusts HVAC operation automatically. In the case of automated adaptation, energy use patterns have a more pronounced effect than usual. System energy consumption may fluctuate more frequently and to a greater degree to adapt to changes in occupancy schedules and the internal heat gains generated by occupants and their associated activities, such as cooking. Many modern houses also contain multiple space-conditioning zones, with different devices serving each zone. For example, a house may have a separate HVAC system for the first and second floor. In such houses, heat transfer between zones introduces an additional layer of complexity to the energy forecast, as the operation of one HVAC system is affected by the operation of the other (Cui et al., 2019). The resulting power signals may be much more chaotic than the signals of binary or finite state machines found in the house and, therefore, more difficult to forecast.

Modern water heaters also pose new challenges in energy forecasting. Many water heaters today contain thermostats that adapt control of the device to various modes of operation that are set by the user. Hybrid electric air-source heat pump water heaters, which

contain both a heat pump and electric resistive elements, are becoming increasingly popular, and their dual mechanisms of heating make forecasting water heater energy consumption increasingly challenging. The complexity of these devices, combined with the stochasticity introduced by occupant-influenced energy use patterns, results in a high degree of forecast performance variability.

Clothes dryers and dishwashers also offer an opportunity for load control, as a delay in their operation is typically not much of an inconvenience for the user. However, these devices are typically not used every day, and may be used during different time intervals on the days they are used. The energy consumption of these devices is difficult to predict due to the sporadic nature of the associated operation schedules.

The use of AMI data for load forecasting is desirable due to the ubiquity of AMI smart meters and the abundance of associated data. In this study, it is assumed that the appliance power signals are perfectly disaggregated from the AMI mains real power signal. Several forecasting methods using this simulated AMI data are assessed and discussed based on signal characteristics associated with occupant behavior.

In this study, the forecastability of appliances with various energy consumption patterns is investigated using real power signals from a set of houses located in Birmingham, Alabama. Appliance power forecastability is the degree to which a model may accurately predict the appliance energy consumption over a given look-ahead period. Although much work has been done to develop accurate models for energy consumption (Beaudin and Zareipour, 2015; Jin et al., 2017; Mariano-Hernández et al., 2021; Wang et al., 2021; Jin et al., 2022), at present, the literature lacks studies detailing aspects of the appliance power signals that may influence model accuracy. Forecastability studies have,

however, been conducted in the field of wind energy (Feng et al., 2019). Wind energy forecasts often vary greatly in accuracy from one period to another; therefore, an understanding of the factors influencing forecastability is vital to accurately predicting generation capability over a given look-ahead period. The forecastability of a time series may be correlated with features of the training and forecasting data sets. The best modeling methods may be selected, but characteristics of the time series may prevent the model from accurately forecasting the energy consumption during that time. It is important for grid managers to have an idea of the uncertainty for a given forecast to be able to schedule grid resources accordingly. In this study, the impact of occupant energy usage patterns on appliance forecastability is investigated. Forecasting studies often present comparisons of algorithms for a given appliance, and/or a comparison of the performance of an algorithm when applied to several appliance types. In such studies, the variation in algorithm performance may be attributed to the operational characteristics of the individual devices. In this study, we compare algorithm performance across a set of identical devices, each installed in a different house. It is demonstrated that there is significant variation in algorithm performance that is dependent not only on the appliance type, but also on the energy use patterns in the residence. Knowledge of the relationship between energy use patterns and forecast performance will allow grid managers to select houses that may satisfy a predefined accuracy requirement for forecasting, based on their energy use patterns.

4.1 Data sets

The data set used in this study was collected from a neighborhood developed for smart grid research in Birmingham, Alabama. All houses in the neighborhood have the

same construction characteristics and are equipped with major appliances of the same manufacturer and model. Typically, houses within a given data set include a variety of appliance types and models, as is the case in most publicly available data sets (Kolter and Johnson, 2011a; Anderson et al., 2012; Makonin, 2016; Steven et al., 2017). Additionally, the houses included in such data sets may not be located in the same geographic region and may have a variety of construction characteristics, including insulation R-values, exterior cladding types, and geometries. In this study, the single location, presence of identical appliance models across all houses, and homogeneity of building construction characteristics allow for the investigation of the effect of occupant behavioral patterns and local weather on the forecastability of appliance energy consumption. Occupant behavior is defined as the usage patterns of appliances within the house and may vary significantly from one house to another. In the case of the HVAC system, energy consumption is influenced by the occupant through executing temperature set point changes and on/off commands, specifying modes of operation, influencing internal heat gains, and introducing heat exchange with the external environment by opening windows and doors. Water heater energy consumption is influenced by the occupants by executing temperature set point changes, hot water consumption, and, occasionally, on/off commands. Operation of the clothes dryer and dishwasher have no significant external influences other than the behavior of occupants. The energy consumption of these devices is dictated by the schedules imposed upon them by the occupants. Despite the similarity of devices and building construction, occupant-influenced energy use patterns vary significantly from house to house, and we demonstrate how these variations may influence the forecastability of hourly real power consumption.

Figure 4.1 presents an example of HVAC, water heater, clothes dryer, and dishwasher real power signals sampled at 1-hour intervals from one house in the data set. The HVAC systems installed throughout the neighborhood are Carrier Infinity systems. In these systems, both the outdoor unit compressor and indoor air handling unit fan have continuously variable load profiles. The HVAC system operates more frequently than any of the other devices but has a higher degree of variability in power draw magnitude. The heat pump water heater is a Rheem hybrid heat pump water heater. In the hybrid heat pump units, water may be heated using an air-source heat pump, electric resistive elements, or both, operating in parallel. The combination of heating technologies results in a load profile that appears as a finite-state machine with intermittent operation patterns. The dishwasher and clothes dryer are both finite-state machines, with the load varying according to the

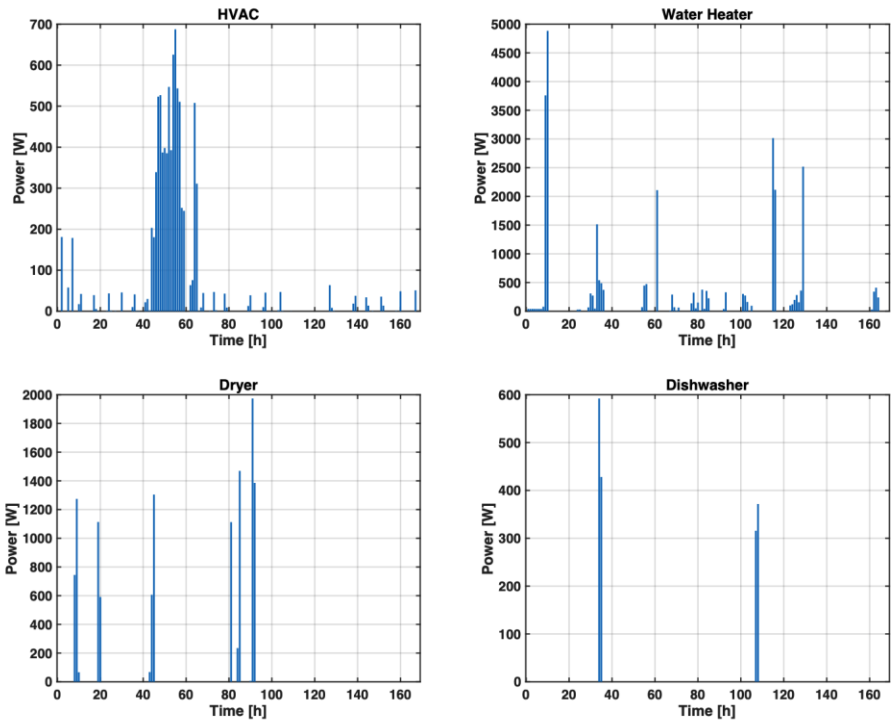

Figure 4.1: Hourly load profiles for each appliance over a one-week period

mode of operation selected. These devices operate much more intermittently than the HVAC system or the water heater. The operating cycles of each device are often less than the one-hour sampling rate that is used in this study to represent low-resolution AMI data. In some cases, an operational event takes place during the transition between hours, resulting in the energy consumed during this operational cycle being assigned partially to each of the two hour-long intervals. This effect often results in devices that are finite-state machines appearing as though they have more operational states than are actually present, and thus, a more complex signal. However, accuracy loss related to power draw magnitude may be countered by increased accuracy of the probability of an operational event over the longer time interval.

The months of November through March are considered in this study. These months correspond to the winter residential time-of-use rate structure of the local utility. Energy consumption among houses monitored over this period varies considerably, with most houses in the data set ranging from 3,000 to 7,000 kWh of energy consumption over the five months, despite the similarity between the houses. Five houses were chosen for evaluation, which represent a range of energy consumption patterns. Figure 4.2 presents the energy consumption of each of the houses, as well as the percentage of energy consumed by each of the appliances across the set of houses. Total seasonal energy consumption in this set of houses ranges from approximately 3,000 kWh to approximately 9,300 kWh. Of the appliances considered, the HVAC system consumes the most energy in all of the houses except House 1, which has greater water heater energy consumption. It would typically be unusual for the water heater to consume more energy than the HVAC system, but this may be explained by considering the particular water and energy

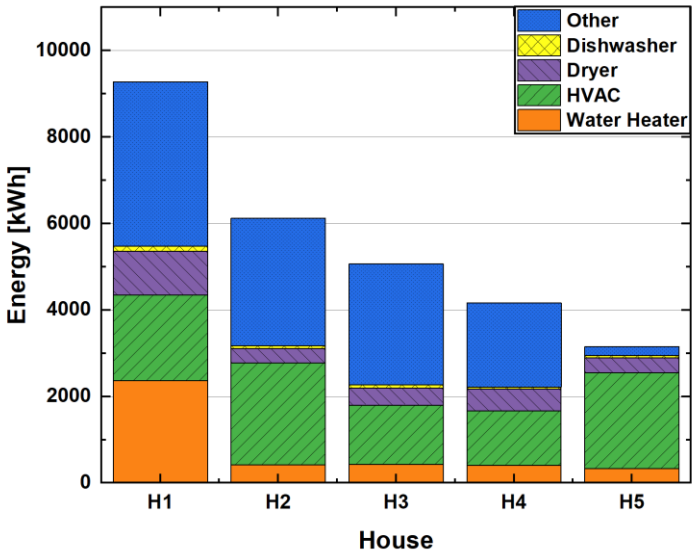

Figure 4.2: Total energy consumption by house

consumption patterns of House 1. The water heater in House 1 has a greater rate of water usage than all other houses in the neighborhood, averaging approximately 30 liters per hour in the winter season. A higher rate of hot water usage implies that the heat pump alone is unable to meet the heating demands and must be assisted by the electric resistive elements. The resistive elements consume ~5000 W of power, as opposed to the ~450 W of the heat pump. Indeed, the electric resistive elements operated during ~50% of the House 1 water heater operational cycles. Additionally, the frequent use of appliances that contribute to internal heat gains, like the clothes dryer, decrease the heating load in the house, reducing HVAC energy consumption. In the remaining houses, the water heater consumes the next greatest amount of energy, followed by the clothes dryer and dishwasher. It is notable that the dishwashers were used sparingly in all houses in the neighborhood, consuming approximately 1% of the total energy, on average.

This wide range in energy consumption occurs in spite of the similar house construction and uniformity of appliance type. Local weather is also ruled out as a contributor to variance in energy consumption across the houses, as all houses are located in the same neighborhood. Variation in energy consumption is due only to variation in occupant behavior, and we investigate how this variation affects appliance-level energy forecasting.

4.2 Methodology

To provide insight into the effect of increasing intermittency in appliance time series data on model performance, models of several types were used to forecast energy consumption using low-resolution real power data and compared. Autoregressive integrated moving average (ARIMA), support vector regression (SVR), Long short-term memory (LSTM) networks, and models using the mean hourly energy consumption were evaluated. The predictive performance of the models is compared with the predictions of a baseline model based on the mean energy consumption profiles of each appliance, to determine the effect of appliance energy consumption on forecast accuracy. The objective of this study is to understand the impact of time series characteristics on model performance, rather than to develop the best model; therefore, basic versions of each forecasting model were used for demonstration. Because the data used in these forecasts are assumed to be collected using AMI smart meters, only the power signal is included as a predictive feature. This approach has the added benefit of allowing better comparison of model performance between univariate models, such as ARIMA, and models capable of multivariate input, such as LSTM.

Before modeling, the real power data were checked for stationarity using the Augmented Dickey-Fuller test (Fuller, 1996). None of the time series were found to be stationary using this method. Each data set was also checked for completeness, with all data sets achieving a completeness >98%. The data were originally sampled at 1-minute intervals and were downsampled to 1-hour intervals before forecasting by averaging the energy consumption over each hour, to simulate data collected by an AMI meter. Multiplying the average hourly power by a time constant results in the total energy consumed in the hour, which is the value provided by an AMI smart meter.

The Kruskal-Wallis H Test was performed to evaluate the significance of temporal characteristics on the energy consumption of each appliance. The Kruskal-Wallis H Test is a one-way analysis of variance test that determines where two or more sample distributions originate from the same distribution (Kruskal and Wallis, 1952). The test determines whether at least one of the samples within the set is statistically different from the others. Both hour of day and day of the week were evaluated using this method. Results of this analysis are discussed in Section 4, where the relationship to modeling performance is discussed.

The ARIMA model was developed using the Statsmodels Python package (Seabold and Perktold, 2010). ARIMA models are a basic form of time series modeling that have been used in the literature for energy forecasting in commercial buildings (Hadri et al., 2019) and aggregate, grid-level energy forecasting. The number of autoregressive lags for the ARIMA model is set to two, while the order of differencing is one, and the number of moving average lags is zero.

Support vector machine (SVM) algorithms are typically used for classification. A plane is fit between clusters of data to identify classification categories. SVR performs regression using the hyperplane, rather than using it to cluster data, as in typical SVM classification. Thus, the success of the SVR algorithm depends on the cluster distinction. If data are loosely clustered, the SVR hyperplane will be overfit to the training data set and be unable to generalize to future forecasting periods. The SVR model was developed using the Scikit-learn Python library function for SVR. Hour of day and mean hourly real power draw were used as inputs to the model, and a sigmoid kernel function was specified.

LSTM models allow for the persistence of temporal trends in the model over long time intervals. The LSTM model was developed using the Keras Python toolkit and is a stacked three-layer network that uses the sigmoid activation function, with a batch size and epoch number of 200. Batch size and epoch number were chosen based on a grid search optimization scheme in which the optimal values were those that produced the lowest error, as measured by MAAPE, for a given house.

A mean value model was used as a baseline for model comparison. The mean value model is simply the mean of the hourly energy consumption of a given device. Use of the mean value model implies that the assumption that the energy consumption values across the training period for each hour are normally distributed. In reality, these values are often not normally distributed. For this reason, a second mean value model was developed in which Gaussian mixture models (GMM) were used to separate energy consumption into several normal distributions. The mean values of each distribution were found and then weighted according to the probability that a value would fall into a given distribution. The unaltered mean value model and the GMM mean value model were used as a baseline for

comparison among the SVR, ARIMA, and LSTM models, to determine if these methods offer any advantage over a basic estimation of energy consumption based on the mean hourly value.

Forecasting was performed for each day in a twenty-eight-day period, with the preceding ninety days of each look-ahead period used as training data. A walk-forward modeling approach was used, with the model being trained for each of the 28 days using data from the preceding 90 days, rather than increasing the number of training days for each training period. The training period was chosen based on an iterative evaluation of training intervals ranging from 5 to 110 days. After 90 days, no improvement was demonstrated, on average, in the models. For devices such as the HVAC, this is most likely due to seasonal variations in energy consumption patterns. Including a greater number of days risks the inclusion of data from a different season, with different energy consumption patterns. The lack of consistency in energy consumption patterns, influenced by weather trends, results in a less accurate forecast.

The model performance is evaluated using mean arctangent absolute percent error (MAAPE) (Kim and Kim, 2016), mean absolute error (MAE), and root mean square error (RMSE). When choosing error metrics, it is important to consider the end use of the forecast. In forecasting the flexibility of energy resources for demand response, the total energy availability is of interest, rather than agreement of the forecast on a point-by-point basis. For this reason, MAAPE is the primary error metric, while MSE and RMSE are used to provide additional insight into the effects of the time series characteristics on model performance. Mean absolute percent error (MAPE) (Equation (4.1)) is commonly used in the literature but may, however, encounter problems with division by zero when the

measured power of a device is zero. This issue is encountered often when considering appliances that operate infrequently.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{4.1}$$

To avoid division-by-zero errors, Kim and Kim (2016) proposed the MAAPE as an alternative to MAPE for intermittent demand forecasting. MAAPE has subsequently been used in energy forecasting literature (Liu et al., 2022; Yang et al., 2022), and is used in this study due to its establishment as an alternative metric to MAPE. MAAPE takes the arctangent of the MAPE, which results in a value that is restricted to the range zero to $\pi/2$. MAAPE is the primary error metric used in this study, as it is a scale-independent metric that captures information about the performance of the algorithms in terms of estimating total energy consumption, while not being affected by the high degree of intermittency of the operation of some appliances. In this study, MAAPE is reported in degrees, resulting in a range of 0 to 90, with 90 corresponding to an instance of division by zero, or an infinite value. Thus, 90 is the maximum possible inaccuracy achievable, and 0 represents a perfect forecast.

$$MAAPE = \frac{1}{n}\sum_{i}^{n}\arctan\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right) \tag{4.2}$$

MAE and RMSE are scale-dependent metrics that give an idea of the algorithm performance on a point-by-point basis. MAE gives an idea of the degree of deviation from the true data and increases linearly as the deviation increases. RMSE, however, amplifies the impact of large deviations from true data. Comparing the MAE and RMSE can provide

insights about the degree to which individual, large deviations are influencing the algorithm performance.

$$MAE = \frac{1}{n}\sum_{i}^{n}|y_i - \hat{y}_i| \tag{4.3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4.4}$$

These metrics are often used in the literature, but also have limitations when evaluating sparse time series. Most notably, sparse time series result in low MAE and RMSE values due to the high number of zero-energy intervals in the time series. Both errors will be larger in time series with more frequent power draws, even if the error of prediction during actual operational periods is identical.

4.3 Results and discussion

Results of the Kruskal-Wallis H-Test demonstrate a strong correlation between hour of day and energy consumption for both the HVAC system and the HPWH. There are also significant correlations between the day of the week and energy consumption values for both the HPWH and HVAC system for some of the houses. The clothes dryer and dishwasher, however, do not exhibit a correlation between hour of day and energy consumption for most houses. For the HVAC system, a strong correlation between hour of day and energy consumption is a reasonable outcome, because HVAC operation is closely tied to the outdoor ambient temperature, which has a strong correlation with time of day. HVAC energy consumption also has a strong seasonal correlation with energy consumption. In winter months, energy consumption for the HVAC system peaks in the

night hours, when temperatures are coldest, and peaks at mid-afternoon during the summer months, when outdoor ambient temperatures are hottest. Occupant behavior has a greater influence on water heater energy consumption than ambient temperature, but the water heater tends to follow a similar pattern day-to-day, as hot water energy consumption is tied to activities like showering and cooking. These activities tend to follow similar schedules each day and are performed relatively consistently. The dishwasher and clothes dryer, however, do not have a significant correlation between day of week in most cases, as their usage schedules are more irregular. To illustrate the variability of appliance energy consumption patterns from house to house, Figure 4.3 and Figure 4.4 present the energy

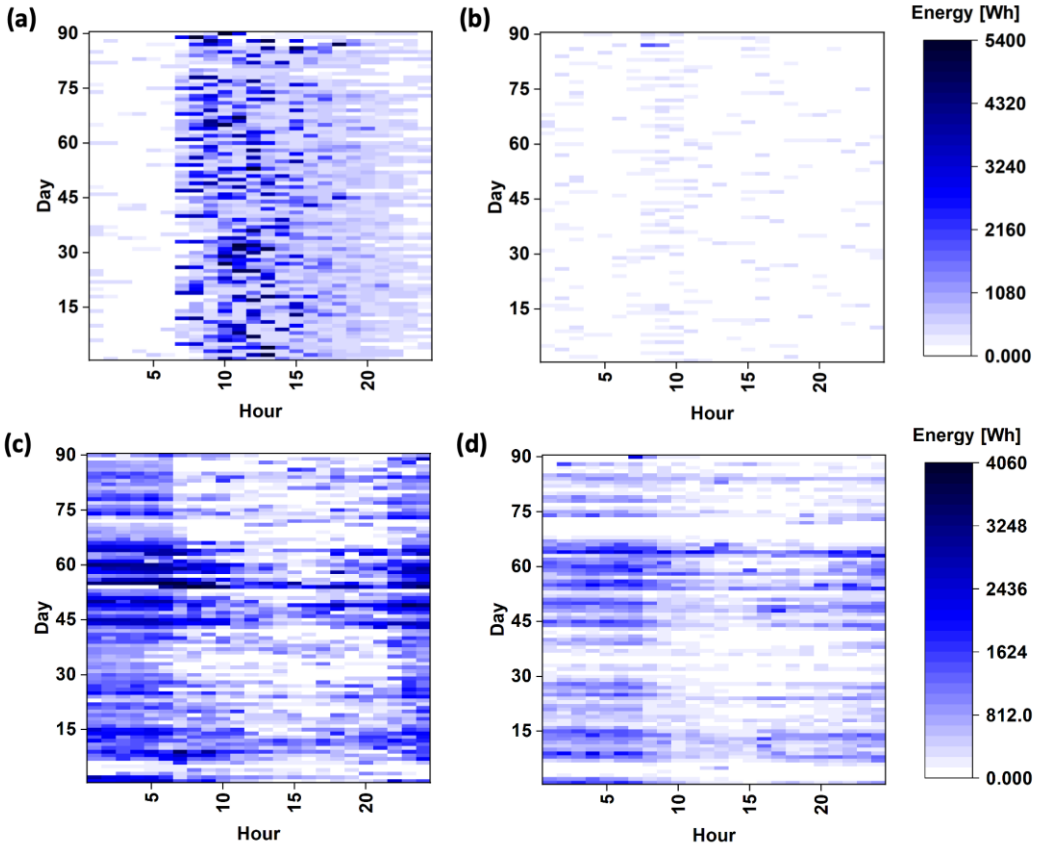

Figure 4.3: Energy consumption by hour over 90-day training period for (a) HPWH greatest energy consumption, (b) HPWH least energy consumption, (c) HVAC greatest energy consumption, (d) HVAC least energy consumption

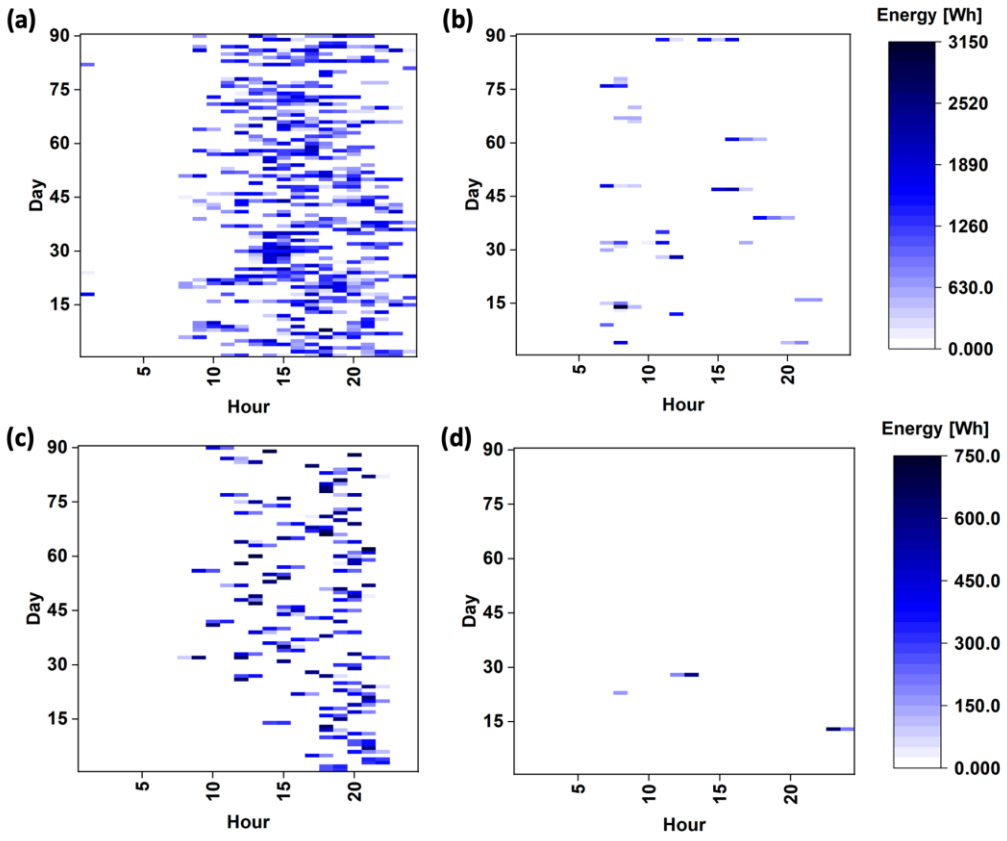

Figure 4.4: Energy consumption by hour over 90-day training period for (a) clothes dryer greatest energy consumption, (b) clothes dryer least energy consumption, (c) dishwasher greatest energy consumption, (d) dishwasher least energy consumption

consumption of each appliance in the house with greatest energy consumption and the house with the least energy consumption.

Each of the model types was applied to the HVAC, HPWH, dishwasher, and clothes dryer data. To determine the optimal training time horizon, MAAPE was calculated for each of the five houses for time intervals from 5 days to 115 days for each of the appliance types. The forecast response to variation in training time interval was significantly different for each of the forecasting methods considered. For the ARIMA models, forecasting performance across model types was fairly consistent for all training periods greater than

10 days. The SVR models, however, had greater variability, particularly for the hybrid heat pump water heater. Across the model types, the forecast performance peaked at approximately 90 days.

Forecasting results for the clothes dryer are presented in Figure 4.5, along with the energy consumption associated with the dryer for each house. The intermittency of the clothes dryer renders it the most difficult to forecast of all appliances considered. It is apparent that the MAAPE increases as the dryer energy consumption decreases. The baseline mean value models provide better results, as measured by MAAPE, than the other modeling methods, with a minimum MAAPE of ~29 in the house with the greatest energy consumption. The operation of the clothes dryer results in so few operational events that applying the GMM approach to the mean value models provides no benefit over treating each distribution as normally distributed. ARIMA and SVR models have the poorest performance in every house. The ARIMA model often functions as a persistence model in intermittent demand situations, and in the case of the clothes dryer model, forecasts an

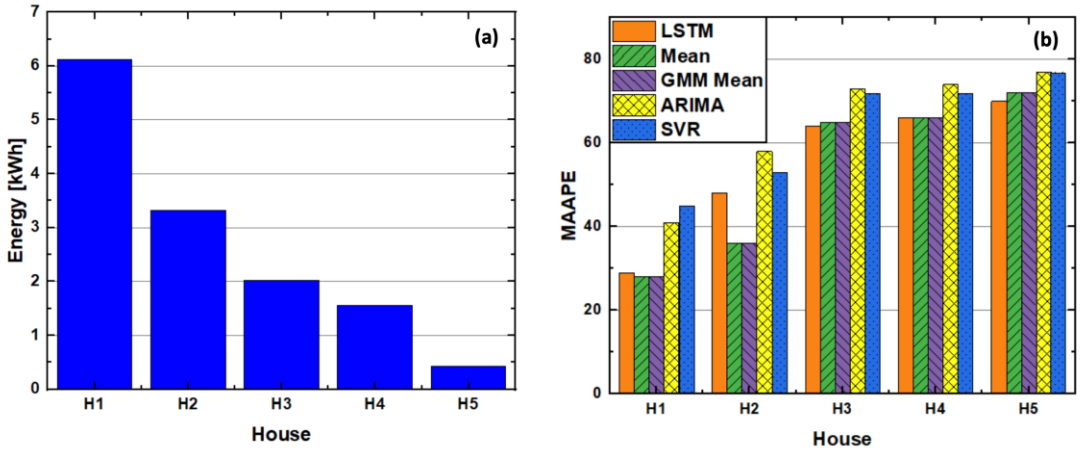

Figure 4.5: Comparison of (a) mean hourly energy consumption of clothes dryer with (b) MAAPE of clothes dryer energy forecast

energy consumption value of zero for most forecast periods. The SVR model fits a hyperplane that also approximates a value close to zero, on most days, for the clothes dryer, and forecasts 0.1 Wh being used every hour. The LSTM is better able to approximate the hourly energy consumption but does not perform better than the mean value model in most houses. The LSTM model performs better than the baseline models (mean and GMM mean) in the two houses with the greatest energy consumption, in which the hour of day is more strongly correlated with clothes dryer energy consumption. However, all the clothes dryer models perform so poorly that this advantage is insignificant.

Forecasting performance using the scale-dependent parameters MAE and RMSE are presented in Figure 4.6. RMSE was consistently higher than MAE by ~100 W, demonstrating that much of the error is due to error in forecasting outliers, which, in the case of a highly intermittent time series, would be all operational events.

The dishwasher models performed similarly to the clothes dryer, with a clear correlation between MAAPE and mean daily energy consumption. MAAPE of the dishwasher forecast is presented along with the mean daily energy consumption in Figure

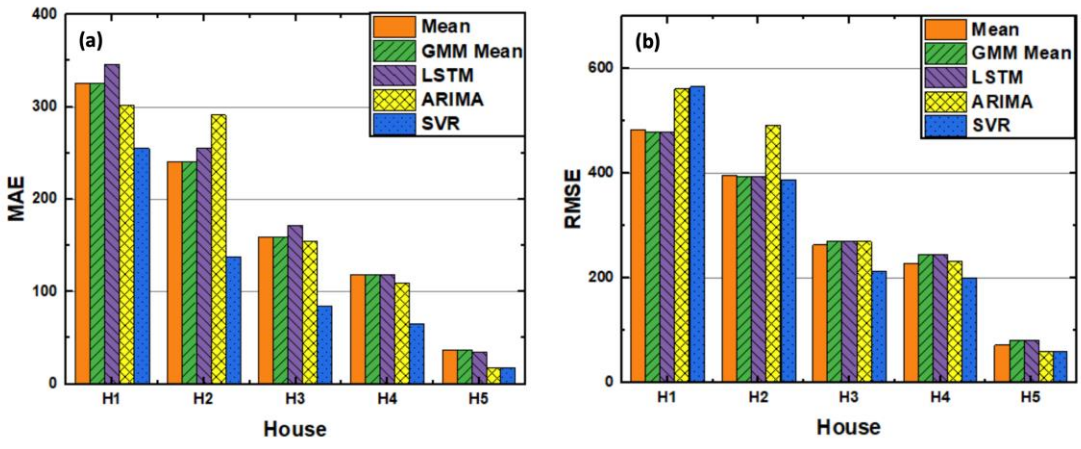

Figure 4.6: Comparison of (a) MAE and (b) RMSE of clothes dryer energy

4.7. MAAPE values for the dishwasher are >60 for four out of five houses considered, reflecting poor performance in the presence of a high degree of intermittency. Although the dishwasher operation exhibits some correlation with hour of day, the use of this device is not consistent on a daily basis and exhibits no correlation with a specific day of the week in most houses. RMSE and MAE followed a pattern similar to that of the clothes dryer, with RMSE being significantly higher than MAE across all houses (Figure 4.8). These results demonstrate that for devices with a similar degree of intermittency and little to no correlation with hour of day, a forecast based on more sophisticated modeling techniques offers no advantage over simply forecasting the mean hourly energy consumption value of the device, as measured during a training period.

MAAPE for the HVAC system is presented with the energy consumption levels of the HVAC system across all houses in Figure 4.9. On average, the LSTM model exhibited the highest accuracy of all models evaluated, as measured by MAAPE, with a mean MAAPE across houses of ~34. This may be attributed to the HVAC system dependence

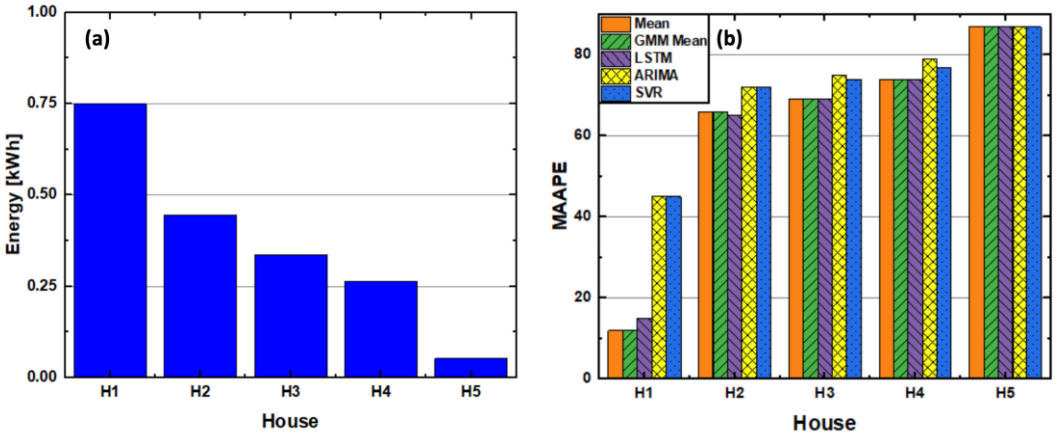

Figure 4.7: Comparison of (a) mean hourly energy consumption of dishwasher with (b) MAAPE of dishwasher energy forecast

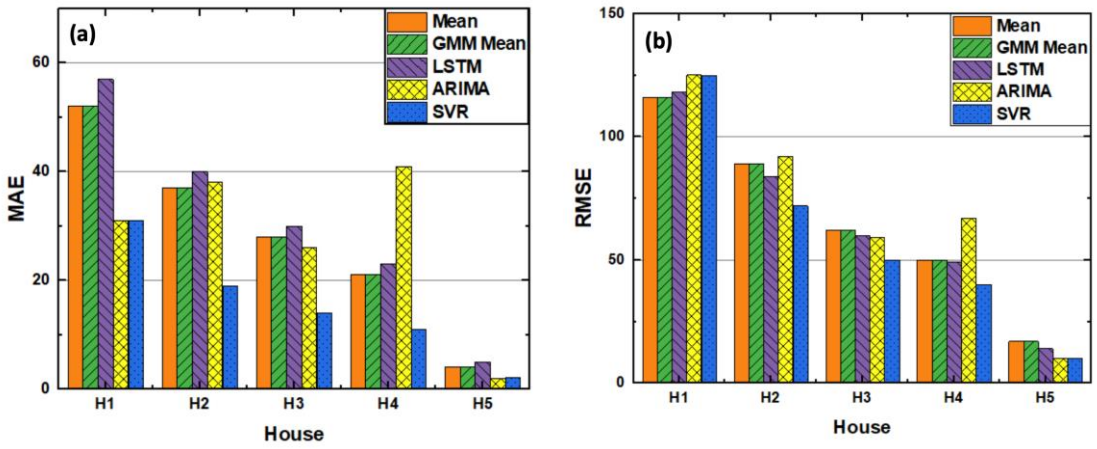

Figure 4.8: Comparison of (a) MAE and (b) RMSE of dishwasher energy forecast

on ambient temperature. Despite not including temperature information in the model directly, hour of day and ambient temperature are correlated in each season. Thus, ambient temperature is indirectly accounted for in the model. The impact of heat exchange between the conditioned space within the house and the external environment, as well as between space conditioning zones, is defined by the principles of physics rather than the behavior of an occupant, which reduces the stochasticity of the appliance signal. Thus, the LSTM is

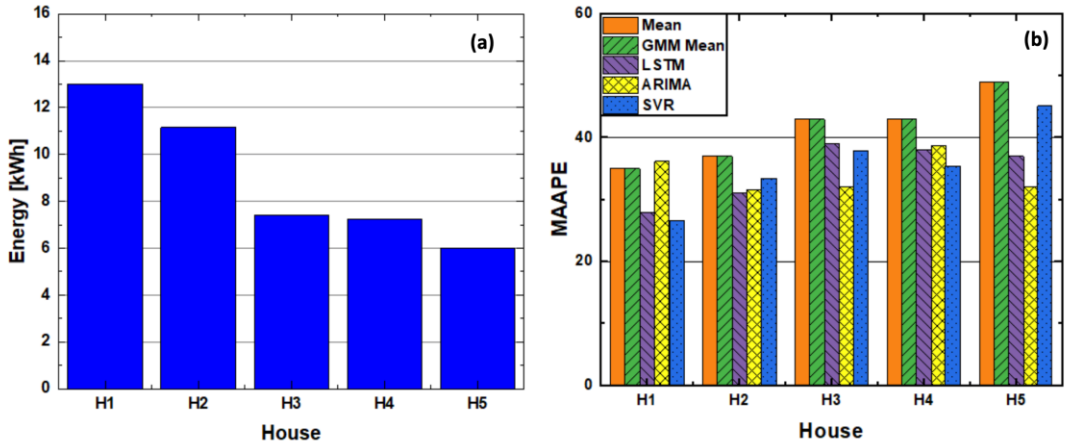

Figure 4.9: Comparison of (a) mean hourly energy consumption of HVAC with (b) MAAPE of HVAC energy forecast

able to infer patterns of operation dependent on time that are not present in other appliance signals.

MAE and RMSE of the HVAC system are presented in Figure 4.10 and follow the same trend as the MAE and RMSE of the clothes dryer and dishwasher, with the RMSE values being greater than the MAE values, implying difficulty in predicting spikes in energy consumption. The HVAC modeling results demonstrate that for a device with characteristics similar to the HVAC system, LSTM models provide the best method of forecasting, but the accuracy of these models depends on the energy consumption of the device as measured during the training interval. Depending on the acceptable threshold for model accuracy, a forecast of a similar signal using an LSTM model may not be capable of forecasting energy consumption for houses with low levels of HVAC energy consumption. By selecting such a threshold, and evaluating HVAC systems in this manner, grid managers may select a set of devices that would be most useful for demand management purposes.

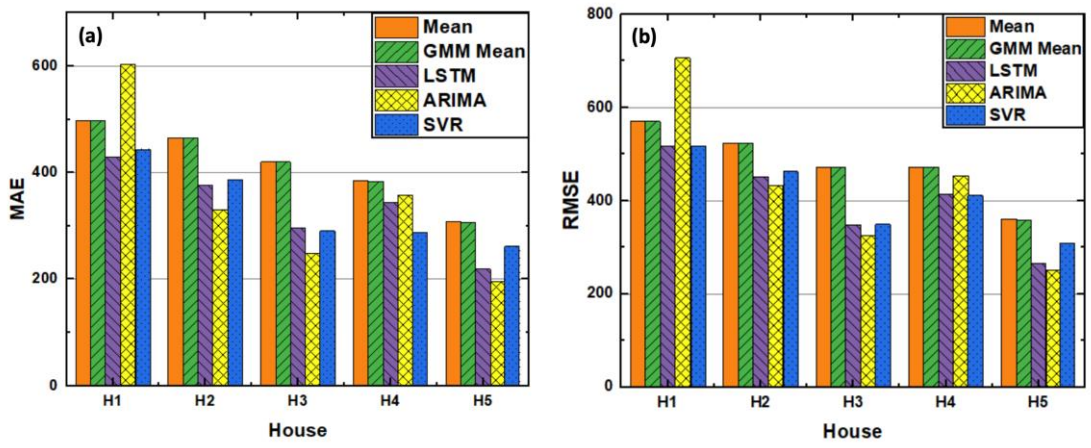

Figure 4.10: Comparison of (a) MAE and (b) RMSE of HVAC energy forecast

The heat pump water heater exhibited the greatest variability in forecasting performance (Figure 4.11). MAAPE for the heat pump water heater appears to be independent of water heater mean energy consumption over the training interval, with the lowest MAAPE of ~11 occurring in a house with one of the lowest energy consumption values. The hybrid heat pump water heater has two modes of operation. If the electric resistance elements operate sporadically, the resulting signal has much more variability than the signal generated from a water heater that is operating only in heat pump mode. The electric resistance elements often operate during periods of high hot water demand, when the heat pump is incapable of heating water to the set point temperature at a sufficient rate. This results in higher energy demand with a more sporadic signal, which may be more difficult to forecast. In House 1, the electric resistance elements operated during approximately 50% of the operational events, resulting in a signal that was more difficult to predict than the water heater signal of the house with the lowest level of water heater energy consumption. House 4 has the highest model accuracy, due to a slightly higher level

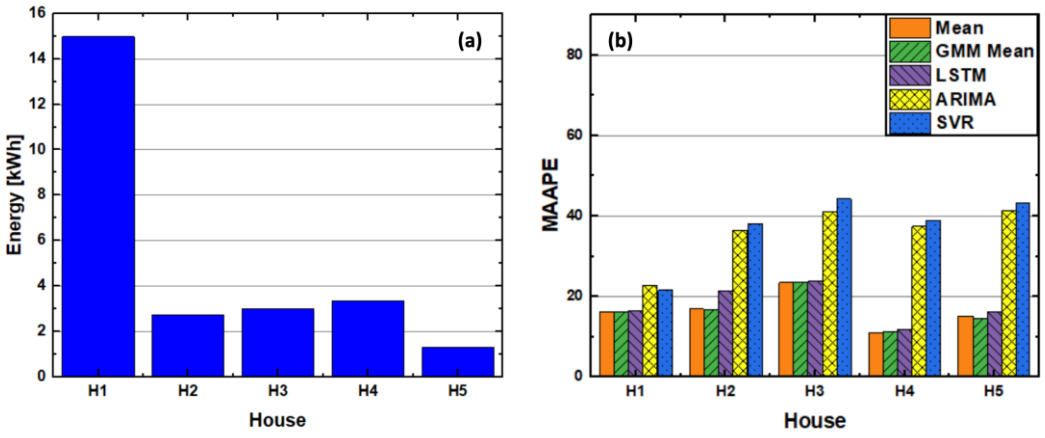

Figure 4.11: Comparison of (a) mean hourly energy consumption and (b) MAAPE of water heater energy forecast

of energy consumption combined with a water heater that operated only in heat pump mode. The consistent mode of operation resulted in a more predictable signal. Performing a comparable analysis on devices used for DSM purposes may guide grid managers in selecting appropriate devices for DSM. For example, if a maximum forecast MAAPE of 20 was allowable for these water heaters, devices may be required to consume energy in excess of 10 kWh or operate in a single mode (heat pump or electric resistance) to be eligible for a DR program.

The same variation of MAE and RMSE with energy consumption is observed in Figure 4.12. It may also be noted that, when considering water heater energy forecasting in one house, the methods that produce the lowest MAAPE do not always produce the lowest MAE or RMSE. This trend may be seen in other appliances as well but is particularly evident in the water heater. For example, the SVR and ARIMA models have much greater MAAPE values than the LSTM and baseline models; however, the MAE and RMSE values for these models are lower than those of the LSTM and baseline models, indicating that the models are better able to predict the energy consumption of these devices

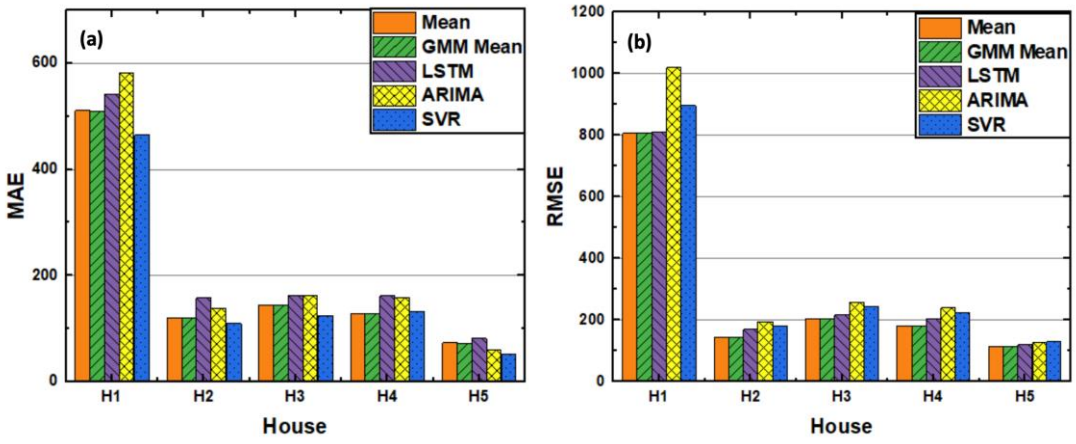

Figure 4.12: Comparison of (a) MAE and (b) RMSE of water heater energy forecast

on a point-by-point basis, but not in predicting the total energy consumption over the prediction interval.

Across all appliances, higher forecast accuracy was achieved at greater levels of energy consumption. It was demonstrated that MAE and RMSE are proportional to energy consumption and intermittency, and that scale-independent metrics that evaluate total energy consumption such as MAAPE are preferable to those evaluating the forecast on a point-by-point basis (MAE and RMSE) when evaluating energy forecast accuracy of intermittent signals. For appliances with the greatest intermittency, such as clothes dryers and dishwashers, the baseline mean hourly value models were shown to be better predictors of energy consumption than more sophisticated modeling methods. For devices influenced by a predictable external trend, such as the case with HVAC systems and ambient temperature, LSTMs have greater accuracy, as they are able to retain pertinent information about these trends over long intervals. The results demonstrate the importance of considering the time series characteristics of an appliance when selecting a modeling method, and the variability of forecast performance caused by variation in usage pattern, even when considering identical devices.

4.3.1.1 Performance limitations

In this section, the models are applied to a set of synthetic time series that represent variations in usage patterns. Hourly probabilities for "on" events were generated using a normal distribution, where x is the hour of day and α is standard deviation, varied from 0.1 to 8.

$$\delta_\alpha(x) = \frac{1}{|\alpha|\sqrt{\pi}} e^{-\left(x/\alpha\right)^2} \tag{4.5}$$

The resulting probability distributions were scaled between 0 and 0.95, resulting in the distributions presented in Figure 4.13. Scaling was performed to specify the same maximum probability for all distributions. To develop the synthetic signals, the probability distributions were applied to a set of random numbers to indicate whether the simulated device was on or off during a given interval. Seven days of the resulting signals are presented in Figure 4.14. Figure 4.14(a) represents a device which has a short duration of operation occurring at approximately the same time each day. Figure 4.14(b) is similar to a device such as the clothes dryer (Figure 4(a)), with hours of operation clustered around the same several-hour period but occurring for more than one operational interval per day. A device such as the water heater of Figure 4.3(a) may be represented by Figure 4.14(c), and a device such as the HVAC system (Figure 4.3(d)) is represented by Figure 4.14(d).

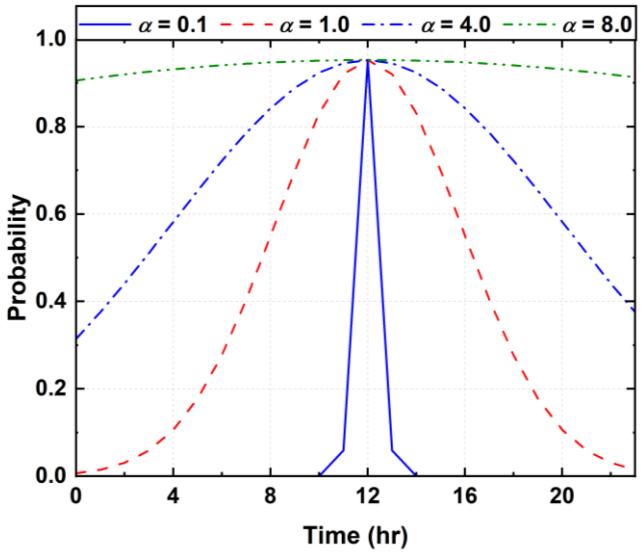

Figure 4.13: Probability distributions of synthetic signals

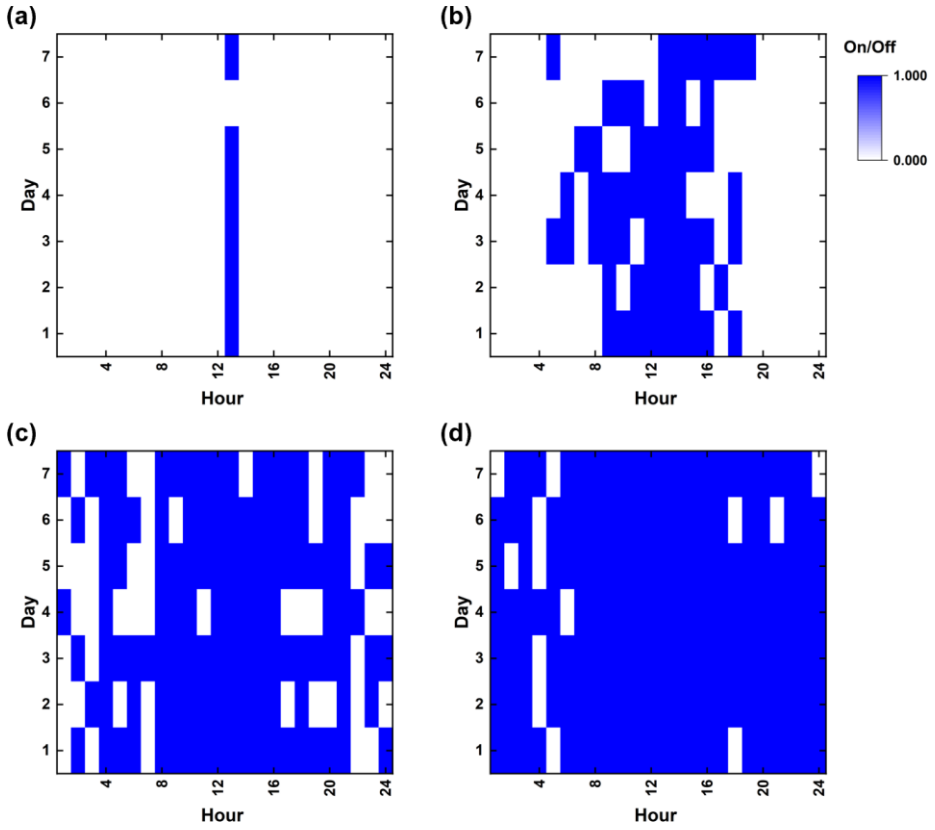

Figure 4.14: Energy consumption by hour over 7-day test period for distributions with (a) α=0.1, (b) α=1.0, (c) α=4.0, (d) α=8.0

Modeling the signals using this method removes any variability due to day-to-day shifts in operational time or variance in power draw, and isolates intermittency and cycle duration or frequency as the variables influencing forecast accuracy.

Each of the four forecasting methods was applied to the problem of determining the operational state over seven 24-hour periods. Seven days were forecast to allow observation of changes in performance due to the stochasticity of the simulated time series. Figure 4.15 and Figure 4.16 present the results of the application of each forecasting method to the various time series.

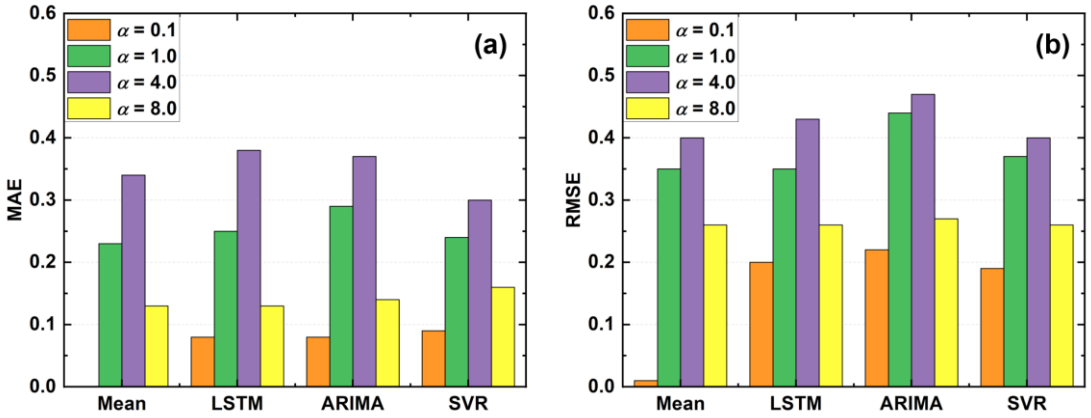

Figure 4.16: Comparison of (a) MAE and (b) RMSE of forecasts using synthetic data

Forecasting time series following the distribution in which $\alpha = 0.1$ was demonstrated to exhibit the greatest accuracy in terms of MAE and RMSE. This result may be attributed to the concentration of operational events during one hour of the day, along with a high likelihood that they will operate during that period. In terms of MAAPE, however, the SVR method of forecasting struggled to predict this signal. This is due to difficulty in fitting a plane to a time series with such intermittent sharp peaks in energy consumption. On a point-by-point basis, as well as in estimation of total energy consumption, the baseline mean

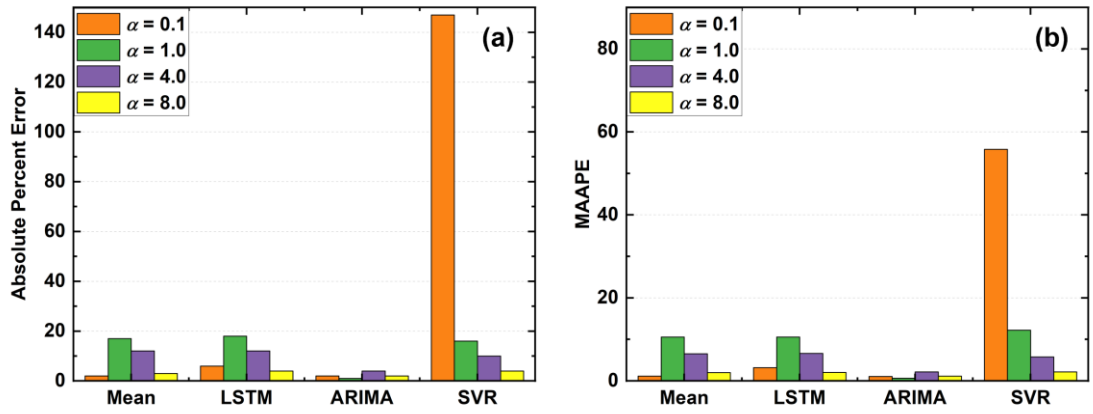

Figure 4.15: Synthetic data forecast (a) absolute percent error and (b) MAAPE

value model was the best predictor of this signal. The next-greatest accuracy was exhibited for forecasts on the time series $\alpha = 8.0$. This distribution has a high probability of operation, with little variation over the course of the day. This signal was also best predicted by the baseline mean value model. The lowest accuracy in terms of MAE and RMSE was demonstrated for the time series following the probability distribution in which $\alpha = 4.0$, followed closely by the time series in which $\alpha = 1.0$. These distributions have greater daily stochasticity in operation as the hour deviates from mid-day. Considering MAE, RMSE, and MAAPE, the SVR forecast was the best predictor of these signals. Comparing accuracies across the various probability distributions, it may be concluded that a greater probability of operation for any given hour has the greatest impact on forecastability for all devices. A strong correlation between hour of day and device operation, as with $\alpha = 0.1$, improves forecastability for most methods, but only if a high probability of operation during that period is maintained. The ARIMA method was the most accurate in modeling the synthetic time series in terms of total percent error; however, this may be attributed to the consistency of day-to-day operational probabilities that are not typically seen in real-world situations, as well as the shorter simulation look-ahead period, as compared to the other modeling methods. In terms of hour-by-hour predictive capability, the SVR and baseline mean value model exhibited the greatest accuracies.

4.4 Conclusion

The stochasticity of residential energy consumption results in a time series that is challenging to predict. In this study, the effect of energy use patterns on appliance-level energy forecasting is investigated using several methods of forecasting. The clothes dryer and dishwasher are the appliances with the greatest intermittency and least correlation

between hour of day and energy consumption; these appliances demonstrate the poorest forecast performance with MAAPE values >60 across all model types in most houses. The HVAC system is best modeled by an LSTM network, with a mean MAAPE of ~34 across all houses but has a high degree of performance variability that is dependent upon the amount of energy consumed. The heat pump water heater demonstrated the best forecast performance of all devices modeled, with MAAPE values <20 for the mean value models, which may be attributed to frequent operation with little seasonal variation. However, the water heater also exhibited the greatest amount of variability in forecast accuracy, with less correlation to the amount of energy consumed during the training period. This variation may be attributed to the use of multiple heating mechanisms, the operation of which are affected by the highly stochastic, occupant-controlled hot water demand.

It was demonstrated that even when controlling for house type, appliance type, and geographic location, occupant behavior has a significant impact on energy consumption patterns from house to house, which translates to a difference in energy forecasting capability. Houses in which the appliances of interest have low energy demand or highly intermittent demand may not be good candidates for demand response load control. With this information, assessment of the energy consumption patterns of houses in a given area using disaggregated AMI data could allow for an estimation of the enrollment potential in demand response programs in future work. This study highlights the importance of considering the variation in forecastability not only among device types, but among individual devices of a given type, based on the energy use patterns. It was demonstrated that optimal algorithm choice is also dependent upon energy use patterns. The framework presented in this study may be used to expand this approach to other device types and used

by grid managers to evaluate the suitability for a device to be used for demand response purposes.

CHAPTER 5. CONCLUSION

5.1 Overview

A study on graph signal processing (GSP) -based load disaggregation and energy forecasting methods for very low-resolution electricity data is presented. For load disaggregation, two approaches for disaggregating electrical loads from aggregate signals measured using AMI meters were investigated. The algorithms consist of a supervised and unsupervised version of GSP-based disaggregation. GSP is a promising new approach for disaggregating very low-resolution loads. For load forecasting, three common methods of time series forecasting are presented and compared to a baseline mean value model. Impacts of energy use patterns on the suitability of forecasting methods is presented.

The supervised approach to GSP is presented first. The supervised approach is not ideal for load disaggregation using AMI data, as it requires labeled appliance signals collected locally to be used in the disaggregation process. However, it was demonstrated in the literature to provide minor performance advantages over the unsupervised GSP algorithm. The algorithm structure of the supervised method is less complex than that of the unsupervised method and allowed for investigation of the effect of parameters used in both the supervised and unsupervised algorithms on disaggregation performance. It was demonstrated that by varying the scaling factor and threshold values, some devices could achieve more favorable disaggregation results. However, the appliances analyzed achieved maximum disaggregation accuracy of $F_M = 0.62$, which is lower than the desired minimum accuracy of $F_M = 0.80$. In addition to varying scaling factor and signal classifier threshold values, the disaggregation sequence is also demonstrated to be important in supervised

disaggregation. This implies that supervised disaggregation requires a labeled set of signals for not only the target device, but other devices that may be disaggregated prior to the target device as well.

The unsupervised approach to GSP is presented next. An unsupervised approach to load disaggregation is ideal for an AMI-based disaggregation application, as labeled data sampled locally is not required. The unsupervised algorithm uses GSP to iteratively cluster similar signal elements representing appliance state changes. The resulting signals represent many devices present in the aggregate load and are automatically disaggregated in sequence from greatest power draw (corresponding to greatest change in power state) to least. Therefore, the unsupervised approach does not require the consideration of disaggregation sequencing. This is an advantage over the supervised approach, as the only signal required to perform this analysis is a labeled signal of the target appliance. This signal need not be collected locally, as dynamic time warping (DTW) is used to match the labeled signal with the disaggregated signal. DTW removes consideration of intermittency from signal matching, considering only the similarity of intervals representing state changes in each signal. It was demonstrated that filtering the labeled signals increased accuracy of the unsupervised GSP approach by reducing the signal variance and increasing the signal mean. The relationship between the variance and mean of the appliance signal and aggregate signal was discussed. The unsupervised GSP algorithm was applied to a dataset from a smart grid research site in Birmingham, Alabama, as well as data from the Pecan Street Dataport. The houses included in these datasets have greater whole-house signal means and variances, reducing disaggregation capabilities. It was demonstrated that, despite the increased complexity of these signals, the signal of the EV charger could be

disaggregated in most houses, achieving $F_M > 0.8$. Success was also achieved in disaggregating the electric resistance elements of the hybrid heat pump water heater, although limited to houses with lower energy consumption, such as those in California. As mentioned previously, F_M did not exceed 0.8 for any of the devices using the supervised method, with the maximum achieved F_M of approximately 0.62. This result occurred despite the reduced overall power consumption of the ORNL house compared to those of the APC and Pecan Street Dataport datasets, indicating that the unsupervised algorithm using filtered signals exhibited better performance.

ARIMA, SVR, and LSTM networks were evaluated for forecasting appliance power signals. Forecasting outcomes were benchmarked against a model developed using the mean hourly power draw of the device. Results of this analysis demonstrated the effect of energy use patterns on forecasting accuracy. Houses with greater levels of energy consumption also exhibited stronger correlation between hour of day and operational state, as well as increased probabilities of device operation during all hours. It was demonstrated, using synthetic data, that the probability of operation has a more pronounced effect on disaggregation accuracy than the correlation between hour of day and operational state. The mean value model used as a benchmark performed similarly to the LSTM network for the real-world data, with greater accuracy than the SVR and ARIMA methods. However, on the synthetic data, ARIMA and SVR were demonstrated to have superior performance. This may be attributed to the repetitive patterns and high probability of operation seen in the synthetic dataset.

5.2 Contributions

This work makes the following contributions:

- State-of-the-art NILM algorithms were evaluated on datasets of greater complexity than those presented in the literature. These datasets are more representative of single-family households within the United States and contain several appliances that would be suitable for DR activities. Datasets included in the literature typically contain devices with low power draw and low variance in the power signal. By evaluating more complex datasets, a better estimate of the expected appliance disaggregation capabilities in U.S. houses was made.
- The GSP-based NILM algorithms were applied in a new context, with the objective of disaggregating appliances that may be used in DR control schemes. Disaggregation in the literature does not typically include these appliance types, or these sampling rates. Results demonstrated that the level 2 EV charger and electric resistive water heater were candidates for disaggregation in many houses. Any device with a similar power signal would be capable of achieving similar results.
- A study investigating load forecasting capabilities of individual residential appliances was presented. Forecasting of individual residential appliance loads is essential for successful direct load control and is not often presented in the literature. The study compared ARIMA, SVR, and LSTM modeling approaches to a baseline mean hourly energy model. Results demonstrated that a simple mean hourly model was capable of predicting energy consumption as well as more advanced methods, in most cases. The accuracy of the forecasting algorithm was

demonstrated to be dependent on the usage patterns of the device, as well as the complexity of the device signal.

5.3 Recommendations for future work

The present study investigated the application of GSP-based NILM approaches to data with greater complexity than datasets presented in the literature, with the objective of demonstrating load disaggregation capabilities in U.S. houses.

- It was demonstrated that by comparing the mean and variance of an aggregate electrical (whole-house) signal to the target signal desired for disaggregation, an estimation of the potential disaggregation accuracy can be made. In the future, this work could be extended to a larger dataset of AMI meter data, performing this analysis for any desired target signal.
- Electric vehicle (EV) chargers were demonstrated to have the greatest potential for disaggregation using GSP. In the datasets considered, very few EV chargers were present, resulting in these signals being used as representative signals for Level 2 EV chargers across all houses in the dataset. Future work may use a dataset that includes more EV charger data to further investigate the effect of differences in patterns of operation and other signal characteristics on disaggregation performance.
- In this study, the suitability of various load forecasting methods to forecasting appliances with various usage patterns was investigated. A dataset that includes more EV chargers could also be used to investigate the application of various load forecasting methods to this signal type.

- Load forecasting methods were applied to a set of devices in real-world houses, and the outcomes of this application was further analyzed using synthetic time series data. Future work may use more synthetic data to further analyze the effect of usage patterns on forecasting outcomes.
- Additional signal features may be available from AMI meter data, such as reactive power. In the future, these features may be incorporated into disaggregation algorithms, in addition to the use of real power, to improve disaggregation outcomes.

While this study investigates the application of NILM and forecasting methods to improve currently available DR estimation tools, future research may also include topics related to direct load control of appliances for DR. These topics would bridge the gap between estimation of available DR capacity and executing DR commands to achieve load shaping objectives. Such topics may include collecting occupancy data to improve disaggregation outcomes or load forecasts; gathering energy use data directly from device APIs, such as those available directly from the electric vehicle interface or HVAC thermostat; and assessing the computational requirements for residential DR schedule optimization, among others. In addition to analyzing DR scheduling at the individual house level, future work may consider the interaction among houses within a group. Topics may include assessing aggregate forecast error among houses, assessing the impact of seasonality on the availability of energy for DR activities, and evaluating the impact of selecting a subset of houses for higher-frequency sampling.

REFERENCES

Abdel-Aal, R. E. and A. Z. Al-Garni (1997). "Forecasting Monthly Electric Energy Consumption in Eastern Saudi Arabia Using Univariate Time-Series Analysis," *Energy*. Vol. 22(11), pp. 1059-1069 DOI: https://doi.org/10.1016/S0360-5442(97)00032-7.

Abledu, G. (2013). "Modeling and Forecasting Energy Consumption in Ghana," *Journal of Energy Technologies and Policy*. Vol. 3, pp. 1-9.

Abosedra, S., A. Dah and S. Ghosh (2011). "Demand for Electricity in Lebanon," *International Business & Economics Research Journal (IBER)*. Vol. 8 DOI: 10.19030/iber.v8i1.3083.

Aiad, M. and P. H. Lee (2016). "Unsupervised Approach for Load Disaggregation with Devices Interactions," *Energy and Buildings*. Vol. 116, pp. 96-103 DOI: https://doi.org/10.1016/j.enbuild.2015.12.043.

Anderson, K., A. Ocneanu, D. R. Carlson, A. Rowe and M. Bergés (2012). "Blued : A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research," *Proceedings of*.

Asadinejad, A., A. Rahimpour, K. Tomsovic, H. Qi and C.-f. Chen (2018). "Evaluation of Residential Customer Elasticity for Incentive Based Demand Response Programs," *Electric Power Systems Research*. Vol. 158, pp. 26-36 DOI: 10.1016/j.epsr.2017.12.017.

Basu, K., V. Debusschere, S. Bacha, U. Maulik and S. Bondyopadhyay (2015). "Nonintrusive Load Monitoring: A Temporal Multilabel Classification Approach," *IEEE Transactions on Industrial Informatics*. Vol. 11(1), pp. 262-270 DOI: 10.1109/TII.2014.2361288.

Batra, N., J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh and M. Srivastava (2014). *Nilmtk: An Open Source Toolkit for Non-Intrusive Load Monitoring*. Fifth International Conference on Future Energy Systems. ACM e-Energy, Cambridge, UK.

Beaudin, M. and H. Zareipour (2015). "Home Energy Management Systems: A Review of Modelling and Complexity," *Renewable and Sustainable Energy Reviews*. Vol. 45, pp. 318-335 DOI: https://doi.org/10.1016/j.rser.2015.01.046.

Boyd, S. P. (2004). *Convex Optimization*. Cambridge University Press, New York.

Callaway, D. S. and I. A. Hiskens (2011). "Achieving Controllability of Electric Loads," *Proceedings of the IEEE*. Vol. 99(1), pp. 184-199 DOI: 10.1109/JPROC.2010.2081652.

CEC (2022). "Electric Vehicle Chargers in California," *California Energy Commission*. DOI: https://www.energy.ca.gov/zevstats.

Cui, B., C. Fan, J. Munk, N. Mao, F. Xiao, J. Dong and T. Kuruganti (2019). "A Hybrid Building Thermal Modeling Approach for Predicting Temperatures in Typical, Detached, Two-Story Houses," *Applied Energy*. Vol. 236, pp. 101-116 DOI: 10.1016/j.apenergy.2018.11.077.

Deb, C., F. Zhang, J. Yang, S. E. Lee and K. W. Shah (2017). "A Review on Time Series Forecasting Techniques for Building Energy Consumption," *Renewable and Sustainable Energy Reviews*. Vol. 74, pp. 902-924 DOI: https://doi.org/10.1016/j.rser.2017.02.085.

DOE (2021). Alternative Fuels Data Center, Access Date: DOI.

DOE, U. S. (2016). *Advanced Metering Infrastructure and Customer Systems: Results from the Smart Grid Investment Grant Program*. U. S. D. o. Energy.

Dong, B., C. Cao and S. E. Lee (2005). "Applying Support Vector Machines to Predict Building Energy Consumption in Tropical Region," *Energy and Buildings*. Vol. 37(5), pp. 545-553 DOI: https://doi.org/10.1016/j.enbuild.2004.09.009.

EIA (2015). "Residential Energy Consumption Survey," U.S. Energy Information Administration, Washington D.C.

EIA (2019). "Annual Electric Power Industry Report," United States Energy Information Association, Washington, D.C.

EIA (2020). "Annual Energy Outlook 2020," U.S. Energy Information Administration, Washington D.C.

EIA (2022). *Hourly Electric Grid Monitor, Form Eia-930*. U.S. Energy Information Administration.

Elafoudi, G., L. Stankovic and V. Stankovic (2014). "Power Disaggregation of Domestic Smart Meter Readings Using Dynamic Time Warping," *Proceedings of 2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*.

EPA (2022). "Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2022," U.S. Environmental Protection Agency.

EPRI (2019). "Communication Protocol Mapping Guide 1.0, Openadr 2.0 to Ansi/Cta-2045-A: Requirements for Exchanging Information between Openadr 2.0 Clients and Ansi/Cta-2045 Technologies," Electric Power Research Institute, Palo Alto, CA.

Feng, C., M. Sun, M. Cui, E. K. Chartan, B.-M. Hodge and J. Zhang (2019). "Characterizing Forecastability of Wind Sites in the United States," *Renewable Energy*. Vol. 133, pp. 1352-1365 DOI: https://doi.org/10.1016/j.renene.2018.08.085.

FERC (2018). "Assessment of Demand Response and Advanced Metering," Federal Energy Regulatory Commission, Washington, D.C.

Fugate, D. (2019). "Ornl-Nilms Project Yarnell Station Research House," Oak Ridge, TN.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*. Wiley, New York.

Green, C. and S. Garimella (2021). "Residential Microgrid Optimization Using Grey-Box and Black-Box Modeling Methods," *Energy and Buildings*. Vol. 235, p. 110705 DOI: https://doi.org/10.1016/j.enbuild.2020.110705.

Hadri, S., Y. Naitmalek, M. Najib, M. Bakhouya, Y. Fakhri and M. Elaroussi (2019). "A Comparative Study of Predictive Approaches for Load Forecasting in Smart Buildings," *Procedia Computer Science*. Vol. 160, pp. 173-180 DOI: https://doi.org/10.1016/j.procs.2019.09.458.

Hansen, A., J. Staggs and S. Shenoi (2017). "Security Analysis of an Advanced Metering Infrastructure," *International Journal of Critical Infrastructure Protection*. Vol. 18, pp. 3-19 DOI: 10.1016/j.ijcip.2017.03.004.

Hao, H., D. Wu, J. Lian and T. Yang (2018). "Optimal Coordination of Building Loads and Energy Storage for Power Grid and End User Services," *IEEE Transactions on Smart Grid*. Vol. 9(5), pp. 4335-4345 DOI: 10.1109/TSG.2017.2655083.

Hart, G. W. (1992). "Nonintrusive Appliance Load Monitoring," *Proceedings of the IEEE*. Vol. 80(12), pp. 1870-1891 DOI: 10.1109/5.192069.

He, K., L. Stankovic, J. Liao and V. Stankovic (2018). "Non-Intrusive Load Disaggregation Using Graph Signal Processing," *IEEE Transactions on Smart Grid*. Vol. 9(3), pp. 1739-1747 DOI: 10.1109/TSG.2016.2598872.

Hendron, R. and C. Engebrecht (2010). "Building America House Simulation Protocols," United States Department of Energy, Oak Ridge, TN.

Hong, W.-C. (2009). "Electric Load Forecasting by Support Vector Model," *Applied Mathematical Modelling*. Vol. 33(5), pp. 2444-2454 DOI: https://doi.org/10.1016/j.apm.2008.07.010.

Iqbal, H. K., F. H. Malik, A. Muhammad, M. A. Qureshi, M. N. Abbasi and A. R. Chishti (2021). "A Critical Review of State-of-the-Art Non-Intrusive Load Monitoring Datasets," *Electric Power Systems Research*. Vol. 192, p. 106921 DOI: https://doi.org/10.1016/j.epsr.2020.106921.

Jain, R. K., K. M. Smith, P. J. Culligan and J. E. Taylor (2014). "Forecasting Energy Consumption of Multi-Family Residential Buildings Using Support Vector Regression: Investigating the Impact of Temporal and Spatial Monitoring Granularity on Performance Accuracy," *Applied Energy*. Vol. 123, pp. 168-178 DOI: https://doi.org/10.1016/j.apenergy.2014.02.057.

Jin, N., F. Yang, Y. Mo, Y. Zeng, X. Zhou, K. Yan and X. Ma (2022). "Highly Accurate Energy Consumption Forecasting Model Based on Parallel Lstm Neural Networks," *Advanced Engineering Informatics*. Vol. 51, p. 101442 DOI: https://doi.org/10.1016/j.aei.2021.101442.

Jin, X., K. Baker, D. Christensen and S. Isley (2017). "Foresee: A User-Centric Home Energy Management System for Energy Efficiency and Demand Response," *Applied Energy*. Vol. 205, pp. 1583-1595 DOI: 10.1016/j.apenergy.2017.08.166.

Johnson, M. and A. Willsky (2012). "Bayesian Nonparametric Hidden Semi-Markov Models," *Journal of Machine Learning Research*. Vol. 14.

Kaytez, F., M. C. Taplamacioglu, E. Cam and F. Hardalac (2015). "Forecasting Electricity Consumption: A Comparison of Regression Analysis, Neural Networks and Least Squares Support Vector Machines," *International Journal of Electrical Power & Energy Systems*. Vol. 67, pp. 431-438 DOI: https://doi.org/10.1016/j.ijepes.2014.12.036.

Kim, N., M. Kim and J. K. Choi (2018). "Lstm Based Short-Term Electricity Consumption Forecast with Daily Load Profile Sequences," *Proceedings of 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*.

Kim, S. and H. Kim (2016). "A New Metric of Absolute Percentage Error for Intermittent Demand Forecasts," *International Journal of Forecasting*. Vol. 32(3), pp. 669-679 DOI: https://doi.org/10.1016/j.ijforecast.2015.12.003.

Kolter, J. Z. and T. Jaakkola (2012). *Approximate Inference in Additive Factorial Hmms with Application to Energy Disaggregation*. Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics. D. L. Neil and G. Mark. PMLR, Proceedings of Machine Learning Research, Vol. 22 pp. 1472--1482.

Kolter, J. Z. and M. J. Johnson (2011a). "Redd: A Public Data Set for Energy Disaggregation Research," *Proceedings of Proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*.

Kolter, J. Z. and M. J. Johnson (2011b). *Redd: A Public Data Set for Energy Disaggregation Research*. SustKDD Workshop on Data Mining Applications in Sustainability.

Kruskal, W. H. and W. A. Wallis (1952). "Use of Ranks in One-Criterion Variance Analysis," *Journal of the American Statistical Association*. Vol. 47, pp. 583-621 DOI: 10.2307/2280779.

Liao, J., G. Elafoudi, L. Stankovic and V. Stankovic (2014). *Non-Intrusive Appliance Load Monitoring Using Low-Resolution Smart Meter Data*. IEEE International Conference on Smart Grid Communications.

Liu, R., T. Chen, G. Sun, S. M. Muyeen, S. Lin and Y. Mi (2022). "Short-Term Probabilistic Building Load Forecasting Based on Feature Integrated Artificial Intelligent Approach," *Electric Power Systems Research*. Vol. 206, p. 107802 DOI: https://doi.org/10.1016/j.epsr.2022.107802.

Makonin, S. (2016). *Ampds2: The Almanac of Minutely Power Dataset (Version 2)*. Harvard Dataverse.

Malik, A., N. Haghdadi, I. MacGill and J. Ravishankar (2019). "Appliance Level Data Analysis of Summer Demand Reduction Potential from Residential Air Conditioner Control," *Applied Energy*. Vol. 235, pp. 776-785 DOI: https://doi.org/10.1016/j.apenergy.2018.11.010.

Mariano-Hernández, D., L. Hernández-Callejo, A. Zorita-Lamadrid, O. Duque-Pérez and F. Santos García (2021). "A Review of Strategies for Building Energy Management System: Model Predictive Control, Demand Side Management, Optimization, and Fault Detect & Diagnosis," *Journal of Building Engineering*. Vol. 33, p. 101692 DOI: https://doi.org/10.1016/j.jobe.2020.101692.

Muratori, M. (2017). *Impact of Uncoordinated Plug-in Electric Vehicle Charging on Residential Power Demand - Supplementary Data*. NREL.

Murray, D., L. Stankovic and V. Stankovic (2017). "An Electrical Load Measurements Dataset of United Kingdom Households from a Two-Year Longitudinal Study," *Scientific Data*. Vol. 4(1), p. 160122 DOI: 10.1038/sdata.2016.122.

Oğcu, G., O. F. Demirel and S. Zaim (2012). "Forecasting Electricity Consumption with Neural Networks and Support Vector Regression," *Procedia - Social and Behavioral Sciences*. Vol. 58, pp. 1576-1585 DOI: https://doi.org/10.1016/j.sbspro.2012.09.1144.

Piette, M. A., G. Ghatikar, S. Kiliccote, E. Koch, D. Hennage, P. Palensky and C. McParland (2009). "Open Automated Demand Response Communications Specification (Version 1.0)."

Rahimpour, A., H. Qi, D. Fugate and T. Kuruganti (2015). "Non-Intrusive Load Monitoring of Hvac Components Using Signal Unmixing," *Proceedings of 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*.

Sandryhaila, A. and J. M. F. Moura (2013). "Classification Via Regularization on Graphs," *Proceedings of 2013 IEEE Global Conference on Signal and Information Processing*.

Seabold, S. and J. Perktold (2010). "Statsmodels: Econometric and Statistical Modeling with Python," *Proceedings of 9th Python in Science Conference*.

Seo, J., J. Jin, J. Y. Kim and J.-J. Lee (2016). "Automated Residential Demand Response Based on Advanced Metering Infrastructure Network," *International Journal of Distributed Sensor Networks*. Vol. 12(2), p. 4234806 DOI: 10.1155/2016/4234806.

Somani, A., D. Wu, K. Kalsi, A. Reiman, J. Alam, R. Fan, K. Mongird and P. Balducci (2018). "Preliminary Locational Net-Benefit Analysis (Lbna) Using Building Loads Represented as Virtual Batteries," Pacific Northwest National Laboratory, Alexandria, VA.

Stankovic, L., V. Stankovic, J. Liao and C. Wilson (2016). "Measuring the Energy Intensity of Domestic Activities from Smart Meter Data," *Applied Energy*. Vol. 183, pp. 1565-1580 DOI: https://doi.org/10.1016/j.apenergy.2016.09.087.

Stankovic, V., J. Liao and L. Stankovic (2014). "A Graph-Based Signal Processing Approach for Low-Rate Energy Disaggregation," *Proceedings of 2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES)*.

Steven, F., K. Tom, D. Vanda, H. Tarek, F. Farid, C. Michael and W. Lynda (2017). *Refit Smart Home Dataset*.

The MathWorks, I. (2021). *Matlab Release 2021a*. The MathWorks, Inc., Natick, Massachusetts, United States.

Twanabasu, S. R. and B. A. Bremdal (2013). *Load Forecasting in a Smart Grid Oriented Building*. 22nd International Conference on Electricity Distribution (CIRED 2013), Stockholm, Sweden.

Vagropoulos, S. I., G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou and A. G. Bakirtzis (2016). "Comparison of Sarimax, Sarima, Modified Sarima and Ann-Based Models for Short-Term Pv Generation Forecasting," *Proceedings of 2016 IEEE International Energy Conference (ENERGYCON)*.

Wang, S., X. Deng, H. Chen, Q. Shi and D. Xu (2021). "A Bottom-up Short-Term Residential Load Forecasting Approach Based on Appliance Characteristic Analysis and Multi-Task Learning," *Electric Power Systems Research*. Vol. 196, p. 107233 DOI: https://doi.org/10.1016/j.epsr.2021.107233.

Williams, C. K. I. (2021). "The Effect of Class Imbalance on Precision-Recall Curves," *Neural Computation*. Vol. 33(4), pp. 853-857 DOI: 10.1162/neco_a_01362.

Xuemei, L., L. Jin-hu, D. Lixing, X. Gang and L. Jibin (2009). "Building Cooling Load Forecasting Model Based on Ls-Svm," *Proceedings of 2009 Asia-Pacific Conference on Information Processing*.

Yan, K., W. Li, Z. Ji, M. Qi and Y. Du (2019). "A Hybrid Lstm Neural Network for Energy Consumption Forecasting of Individual Households," *IEEE Access*. Vol. 7, pp. 157633-157642 DOI: 10.1109/ACCESS.2019.2949065.

Yang, W., J. Shi, S. Li, Z. Song, Z. Zhang and Z. Chen (2022). "A Combined Deep Learning Load Forecasting Model of Single Household Resident User Considering Multi-Time Scale Electricity Consumption Behavior," *Applied Energy*. Vol. 307, p. 118197 DOI: https://doi.org/10.1016/j.apenergy.2021.118197.

Zhao, B., K. He, L. Stankovic and V. Stankovic (2018a). "Improving Event-Based Non-Intrusive Load Monitoring Using Graph Signal Processing," *IEEE Access*. Vol. 6, pp. 53944-53959 DOI: 10.1109/ACCESS.2018.2871343.

Zhao, B., L. Stankovic and V. Stankovic (2015). "Blind Non-Intrusive Appliance Load Monitoring Using Graph-Based Signal Processing," *Proceedings of 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*.

Zhao, B., L. Stankovic and V. Stankovic (2016). "On a Training-Less Solution for Non-Intrusive Appliance Load Monitoring Using Graph Signal Processing," *IEEE Access*. Vol. 4, pp. 1784-1799 DOI: 10.1109/ACCESS.2016.2557460.

Zhao, B., L. Stankovic and V. Stankovic (2018b). "Electricity Usage Profile Disaggregation of Hourly Smart Meter Data," *Proceedings of 4th International Workshop on Non-Intrusive Load Monitoring*.

Zhao, B., M. Ye, L. Stankovic and V. Stankovic (2020). "Non-Intrusive Load Disaggregation Solutions for Very Low-Rate Smart Meter Data," *Applied Energy*. Vol. 268, p. 114949 DOI: https://doi.org/10.1016/j.apenergy.2020.114949.

Zhao, H.-x. and F. Magoulès (2012). "A Review on the Prediction of Building Energy Consumption," *Renewable and Sustainable Energy Reviews*. Vol. 16(6), pp. 3586-3592 DOI: https://doi.org/10.1016/j.rser.2012.02.049.