Data-driven customer energy behavior characterization for distributed energy management

Milad Afzalan

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Civil Engineering

Farrokh Jazizadeh
Jesus de la Garza
Bert Huang
Naren Ramakrishnan

June 15th, 2020
Blacksburg, Virginia

Keywords: Distributed energy management, Smart grid, Machine learning, Human-building interaction, Segmentation.

Data-driven customer energy behavior characterization for distributed energy management

Milad Afzalan

ABSTRACT

With the ever-growing concerns of environmental and climate concerns for energy consumption in our society, it is crucial to develop novel solutions that improve the efficient utilization of distributed energy resources for energy efficiency and demand response (DR). As such, there is a need to develop targeted energy programs, which not only meet the requirement of energy goals for a community but also take the energy use patterns of individual households into account. To this end, a sound understanding of the energy behavior of customers at the neighborhood level is needed, which requires operational analytics on the wealth of energy data from customers and devices.

In this dissertation, we focus on data-driven solutions for customer energy behavior characterization with applications to distributed energy management and flexibility provision. To do so, the following problems were studied: (1) how different customers can be segmented for DR events based on their energy-saving potential and balancing peak and off-peak demand, (2) what are the opportunities for extracting Time-of-Use of specific loads for automated DR applications from the whole-house energy data without in-situ training, and (3) how flexibility in customer demand adoption of renewable and distributed resources (e.g., solar panels, battery, and smart loads) can improve the demand-supply problem.

In the first study, a segmentation methodology form historical energy data of households is proposed to estimate the energy-saving potential for DR programs at a community level. The proposed approach characterizes certain attributes in time-series data such as frequency, consistency, and peak time usage. The empirical evaluation of real energy data of 400 households shows the successful ranking of different subsets of consumers according to their peak energy reduction potential for the DR event. Specifically, it was shown that the proposed approach could successfully identify the 20-30% of customers who could achieve 50-70% total possible demand reduction for DR. Furthermore, the rebound effect problem (creating undesired peak demand after a DR event) was studied, and it was shown that the proposed approach has the potential of identifying a subset of consumers (~5%-40% with specific loads like AC and electric vehicle) who contribute to balance the peak and off-peak demand. A projection on Austin, TX showed 16MWh reduction during a 2-h event can be achieved by a justified selection of 20% of residential customers.

In the second study, the feasibility of inferring time-of-use (ToU) operation of flexible loads for DR applications was investigated. Unlike several efforts that required considerable model parameter selection or training, we sought to infer ToU from machine learning models without in-

situ training. As the first part of this study, the ToU inference from low-resolution 15-minute data (smart meter data) was investigated. A framework was introduced which leveraged the smart meter data from a set of neighbor buildings (equipped with plug meters) with similar energy use behavior for training. Through identifying similar buildings in energy use behavior, the machine learning classification models (including neural network, SVM, and random forest) were employed for inference of appliance ToU in buildings by accounting for resident behavior reflected in their energy load shapes from smart meter data. Investigation on electric vehicle (EV) and dryer for 10 buildings over 20 days showed an average F-score of 83% and 71%. As the second part of this study, the ToU inference from high-resolution data (60Hz) was investigated. A self-configuring framework, based on the concept of spectral clustering, was introduced that automatically extracts the appliance signature from historical data in the environment to avoid the problem of model parameter selection. Using the framework, appliance signatures are matched with new events in the electricity signal to identify the ToU of major loads. The results on ~1500 events showed an F-score of >80% for major loads like AC, washing machine, and dishwasher.

In the third study, the problem of demand-supply balance, in the presence of varying levels of small-scale distributed resources (solar panel, battery, and smart load) was investigated. The concept of load complementarity between consumers and prosumers for load balancing among a community of ~250 households was investigated. The impact of different scenarios such as varying levels of solar penetration, battery integration level, in addition to users' flexibility for balancing the supply and demand were quantitatively measured. It was shown that (1) even with 100% adoption of solar panels, the renewable supply cannot cover the demand of the network during afternoon times (e.g., after 3 pm), (2) integrating battery for individual households could improve the self-sufficiency by more than 15% during solar generation time, and (3) without any battery, smart loads are also capable of improving the self-sufficiency as an alternative, by providing ~60% of what commercial battery systems would offer.

The contribution of this dissertation is through introducing data-driven solutions/investigations for characterizing the energy behavior of households, which could increase the *flexibility* of the aggregate daily energy load profiles for a community. When combined, the findings of this research can serve to the field of utility-scale energy analytics for the integration of DR and improved reshaping of network energy profiles (i.e., mitigating the peaks and valleys in daily demand profiles).

Data-driven customer energy behavior characterization for distributed energy management

Milad Afzalan

GENERAL AUDIENCE ABSTRACT

Buildings account for more than 70% of electricity consumption in the U.S., in which more than 40% is associated with the residential sector. During recent years, with the advancement in Information and Communication Technologies (ICT) and the proliferation of data from consumers and devices, data-driven methods have received increasing attention for improving the energy-efficiency initiatives.

With the increased adoption of renewable and distributed resources in buildings (e.g., solar panels and storage systems), an important aspect to improve the efficiency by matching the demand and supply is to add *flexibility* to the energy consumption patterns (e.g., trying to match the times of high energy demand from buildings and renewable generation). In this dissertation, we introduced data-driven solutions using the historical energy data of consumers with application to the flexibility provision. Specific problems include: (1) introducing a ranking score for buildings in a community to detect the candidates that can provide higher energy saving in the future events, (2) estimating the operation time of major energy-intensive appliances by analyzing the whole-house energy data using machine learning models, and (3) investigating the potential of achieving demand-supply balance in communities of buildings under the impact of different levels of solar panels, battery systems, and occupants energy consumption behavior.

In the first study, a ranking score was introduced that analyzes the historical energy data from major loads such as washing machines and dishwashers in individual buildings and group the buildings based on their potential for energy saving at different times of the day. The proposed approach was investigated for real data of 400 buildings. The results for EV, washing machine, dishwasher, dryer, and AC show that the approach could successfully rank buildings by their demand reduction potential at critical times of the day.

In the second study, machine learning (ML) frameworks were introduced to identify the times of the day that major energy-intensive appliances are operated. To do so, the input of the model was considered as the main circuit electricity information of the whole building either in lower-resolution data (smart meter data) or higher-resolution data (60Hz). Unlike previous studies that required considerable efforts for training the model (e.g, defining specific parameters for mathematical formulation of the appliance model), the aim was to develop data-driven approaches to learn the model either from the same building itself or from the neighbors that have appliance-level metering devices. For the lower-resolution data, the objective was that, if a few samples of buildings have already access to plug meters (i.e., appliance level data), one could estimate the operation time of major appliances through ML models by matching the energy behavior of the buildings, reflected in their smart meter information, with the ones in the neighborhood that have similar behaviors. For the higher-resolution data, an algorithm was introduced that extract the

appliance signature (i.e., change in the pattern of electricity signal when an appliance is operated) to create a processed library and match the new events (i.e., times that an appliance is operated) by investigating the similarity with the ones in the processed library. The investigation on major appliances like AC, EV, dryer, and washing machine shows the >80% accuracy on standard performance metrics.

In the third study, the impact of adding small-scale distributed resources to individual buildings (solar panels, battery, and users' practice in changing their energy consumption behavior) for matching the demand-supply for the communities was investigated. A community of ~250 buildings was considered to account for realistic uncertain energy behavior across households. It was shown that even when all buildings have a solar panel, during the afternoon times (after 4 pm) in which still ~30% of solar generation is possible, the community could not supply their demand. Furthermore, it was observed that including users' practice in changing their energy consumption behavior and battery could improve the utilization of solar energy around >10%-15%. The results can serve as a guideline for utilities and decision-makers to understand the impact of such different scenarios on improving the utilization of solar adoption.

These series of studies in this dissertation contribute to the body of literature by introducing data-driven solutions/investigations for characterizing the energy behavior of households, which could increase the *flexibility* in energy consumption patterns.

ACKNOWLEDGMENT

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Farrokh Jazizadeh, for all his support, guidance, and encouragement. His advice and mentoring during my PhD studies at Virginia Tech have helped to prepare for the next chapter of my career and future research efforts. I would like to thank my committee members, Dr. Jesus de la Garza, Dr. Bert Huang, and Dr. Naren Ramakrishnan for all their constructive feedbacks, recommendations, and support.

Throughout the last 4 years in Blacksburg, I had the chance of getting to know some wonderful friends and colleagues. I would like to thank Armin, Farnaz, Sogand, Wooyoung, and other friends and colleagues in VCEMP and INFORM Lab. With their presence, this whole journey has been much more enjoyable to me.

Finally and most importantly, I would like to thank my parents, for all their unconditional love and support through all different stages. Thanks to my older siblings, Elham and Ali, for always being there beside me and for all the joyful moments. I am very blessed to have you as the family. You are my motivation, and this is dedicated to you!

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction and motivation

With the widespread adoption of distributed energy resources and advanced metering devices, future energy and distribution systems will operate in a considerably different environment. However, how to adapt to new elements for efficient operation is challenging and not fully understood. To this end, digitalization and data-driven solutions are becoming a powerful tool for the transition to sustainable energy. The unprecedented bulk of generated data from users and metering devices provides opportunities for learning and prediction of customers' behavior and to improve the resource coordination.

Residential buildings account for more than 37% of electricity consumption share in the United States [1], with a total consumption of 1.38 trillion kWh in 2017 [2]. Furthermore, the ever-growing concerns on the environmental and carbon emission impact [2, 3], the increasing trend in the urban population [4], and the high cost of supplying the peak demand at critical times [5] necessitate proposing efficient solutions for the management of the power and energy system. To this end, the integration of clean renewable energy resources and energy efficient transportation systems are considered as viable solutions [6], and there is currently a nationwide trend for adopting efficient resources by policies. For example, the state of California had installed solar panels with the capacity of over 10 GW [7], accounting for more than 16% of net electricity generation [8], and the number of electric vehicles is projected to reach around 2 million by 2040 [9]. Despite providing beneficial outcome, with the proliferation of highly volatile distributed renewable resources such as solar generation and the increased adoption of energy-efficient transportation system, the operation of the power system will be facing several challenges. Specifically, (1) with the high increase in the electricity load demand, an increase in the generation is expected. However, due to the high-cost and environmental factor described above, the number of controllable power plants is actually decreasing [10] and other energy resources will be replaced, and (2) with the high penetration of intermittent and fluctuating renewable energy resources, sending the surplus energy back to the grid can cause power instability. Subsequently, this can cause damage to household's appliances [11].

As a result of such evolutions, management of energy demand and supply is becoming more challenging, and there is an increased need to add *flexibility* to the smart grid, in terms of reshaping the energy demand profiles at needed times. In other words, the increased adoption of renewable

energy resources requires the accommodation of flexible and controllable loads. Accordingly, reshaping the aggregate energy demand profiles (from a community of homes) requires making alteration to typical usage patterns at the scale of individual homes based on users' (occupants') behavior [12]; however, it has been shown that demand profiles among different homes (or even one along subsequent days) are highly stochastic and do not follow the same trend [13]. Therefore, the complex and varied interactional behavior among users bring about challenges for characterizing their consumption patterns [14].

During recent years, the advancements in Information and Communication Technologies (ICT) have resulted in generating a vast amount of consumption data from residential homes. Specifically, with the roll-out of smart meters at the national scale and advancement in the advanced metering infrastructures (AMI) [15], opportunities for data analytics on the consumption data of residential homes have emerged in the energy sector. As a result, recent studies have focused on learning and predicting the characteristics of homes' consumption styles, in order to reveal actionable patterns for demand-side management (DSM) (e.g., [13, 16-25]. However, how to utilize such data to adapt to the future distribution systems with a high level of solar, electric vehicle (EVs), smart loads, and the problem of peak demand supply is challenging and not fully understood [14, 26].

1.1. Problem statement and research gaps

The research problem addressed in this dissertation is concerned with how interactional consumption behavior data of residential homes can be leveraged to make informed decisions for improving the flexibility of the energy system. The aim is to understand the dynamic complex human-appliance/building interaction across many households to reveal distinguishing factors of usage patterns and present actionable insight based on users' consumption behaviors.

Prior studies have looked into making occupants more aware about their energy usage, in the form of energy eco-feedback [27-33] for direct engagement of individual homes for DSM. However, considering that such approaches mainly look at energy data at the resolution of monthly or daily basis, they are not applicable to the concept of *flexibility*, in which the fluctuation of energy consumption at hourly basis is needed for distributed energy management. Flexibility is defined as the modification of household daily energy profiles through changing the power draw, the operation duration, and/or the activation time of individual devices [34]. Therefore, the focus of

this dissertation is to reveal actionable feedback based on data analytics for energy planners to enable strategies to improve flexibility of energy usage and to adopt automation scenarios. Specifically, the motivation stems from the fact that imposing a change in the usage behavior through direct feedback is challenging [35] and not as effective as emerging automation platforms [36].

Through our literature review, research gaps were identified as follows.

- Limited understanding of flexibility potential of residential loads for DSM applications according to human-building interaction and their varied level of contribution.
- Limited efforts in leveraging individual load data for improved energy management and automated control.

Based on the essence of the problem statement, I have defined the requirements for addressing the problems as follows:

1) The solution should be capable of identifying a limited subset of proper users for DSM programs while also accounting for a target amount of energy savings. In other words, it needs to avoid the rebound effect (i.e., creating unforeseen peak demand at off-peak time of a DR event, due to simultaneous contribution of many customers in DR).

2) Due to the importance of smart loads contribution in distributed energy management paradigm, the solution should increase the information gain by inferring flexible appliance time-of-use events from investigating daily profile load shapes.

3) Given the importance of demand and supply balancing in the presence of renewables, the approach needs to account for the inherent difference in household's load shapes and investigate how inclusion of distributed energy resources (e.g., solar, battery, users' flexibility) can improve the decentralized energy management and reduce the reliance on the grid.

1.2. Research questions

Based on the aforementioned objectives, the core questions addressed in this work are summarized as follows:

1. How human interactional behavior can be leveraged for efficient operation of flexible loads at the community level?

2. How could common aggregate load shapes be leveraged in inferring the pattern of flexible appliance time-of-use?
3. How does users' complementarity in energy consumption styles and the integration of different distributed energy resource improve the load flexibility?

1.3. Contributions

In this dissertation, we have investigated data-driven solutions for characterizing the energy behavior of residential customers for addressing the demand flexibility. Namely, our contribution include: (1) presenting a segmentation approach with application to DR, which divides the community of consumers based on energy saving. The approach characterizes the features of human-building interaction (frequency, consistency, and peak time usage) for ranking the households based on their potential for peak demand shaving, (2) inference of flexible appliance time-of-use for DR through characterizing the similarities in whole-house energy data (readily available data) using machine learning. Unlike similar studies, the approach does not require the task of in-situ training or model parameter-tuning, and (3) the investigation of a community's self-dependency (using its own renewable generation for meeting its demand) under the realistic uncertainties in demand and generation, in addition to the impact of different levels of distributed energy resources.

These contributions have been presented/published in the following journal papers:

First study (Chapter 2):

- **[37] Afzalan, Milad**, and Farrokh Jazizadeh. "Residential loads flexibility potential for demand response using energy consumption patterns and user segments." *Applied Energy* 254 (2019): 113693. DOI: https://doi.org/10.1016/j.apenergy.2019.113693
- **[38] Afzalan, Milad**, and Farrokh Jazizadeh. "Data-driven identification of consumers with deferrable loads for demand response programs." *IEEE Embedded Systems Letters* (2019). DOI: 10.1109/LES.2019.2937834

Second study (Chapters 3-5):

- **[39] Afzalan, Milad**, and Farrokh Jazizadeh. "A Machine Learning Framework to Infer Time-of-Use of Flexible Loads: Resident Behavior Learning for Demand Response" *IEEE Access* (2020), DOI: https://doi.org/10.1109/ACCESS.2020.3002155
- **[40] Afzalan, Milad**, and Farrokh Jazizadeh. "An automated spectral clustering for multi-scale data." *Neurocomputing* 347 (2019): 94-108. DOI: https://doi.org/10.1016/j.neucom.2019.03.008
- **[41] Afzalan, Milad**, Farrokh Jazizadeh, and Jue Wang. "Self-configuring event detection

in electricity monitoring for human-building interaction." *Energy and Buildings* 187 (2019): 95-109. DOI: https://doi.org/10.1016/j.enbuild.2019.01.036

Third study (Chapter 6):

- **Afzalan, Milad**, and Farrokh Jazizadeh. "Quantified investigation of peer-to-peer energy trading between prosumers and consumer: An empirical analysis " *To be submitted to Applied Energy*.

1.4. Dissertation structure

The rest of the dissertation is structured as follows: In chapter 2, the segmentation approach is presented, and its findings and implications are presented. Chapter 3 presents the inference of time-of-use events of flexible loads using smart meter data. Chapters 4 and 5 presents the self-configuring framework for inference of time-of-use events with high-resolution data. Chapter 6 present the quantified results on the load balancing concept amongst prosumers and consumers. Chapter 7 concludes the dissertation by summarizing the discussions and future directions.

Chapter 2: Residential loads flexibility potential for Demand Response using energy consumption patterns and user segments

Afzalan, Milad, and Farrokh Jazizadeh. "Residential loads flexibility potential for demand response using energy consumption patterns and user segments." *Applied Energy* 254 (2019): 113693. DOI: https://doi.org/10.1016/j.apenergy.2019.113693
Afzalan, Milad, and Farrokh Jazizadeh. "Data-driven identification of consumers with deferrable loads for demand response programs." *IEEE Embedded Systems Letters* (2019). DOI: 10.1109/LES.2019.2937834

Abstract

Demand response (DR) is considered an effective approach in mitigating the ever-growing concerns for supplying the electricity peak demand. Recent attempts have shown that the contribution from the aggregate impact of flexible individual residential loads can add flexibility to the power grid as ancillary services. However, current DR schemes do not systematically distinguish the varying potential of user contribution due to the highly-varied usage behaviors. Thus, this paper proposes a data-driven approach for quantifying the potential of individual flexible load users for participation in DR. We introduced a metric to capture the predictability of usage in a future DR event using the historical consumption data for different load types. The metric helps to sort the users of flexible loads in a community according to their potential for load shifting scenarios. We then evaluated the applicability of the metric in the DR context to assess the extent of energy reduction for different segments of users. In our analysis, we included electric vehicle, wet appliances (dryer, washing machine, dishwasher), and air conditioning. The analysis of real-world data shows that the approach is effective in identifying suitable user segments with higher predictive potential for demand reduction. We also presented a cross-appliance comparison for assessing the flexibility potential of different user segments. As a case study based on Pecan Street Project, the findings suggest that potentially ~140MW demand reduction might be achieved in Austin, TX through only 20% participation of the selected flexible loads for the residential sector during a 2-hour event.

2.1. Introduction

With moving towards distributed and decentralized energy management [42, 43], it is important to add flexibility to the power consumption pattern at the individual household-level to enable

efficient operation of the power system, grid-interactive efficient buildings, and self-adaptive smart grids. Accordingly, demand response (DR) is considered as a cost-effective technique for demand reduction and ancillary services in comparison to conventional and cost-intensive techniques for expanding generation capacity or network augmentation. From the automation perspective, DR schemes vary from methods based on direct user engagement to more automated ways, in which the load operation can be automatically scheduled with Home Energy Management (HEM) systems. The advances in communication technologies and embedded communication modules in appliances [44-46] have paved the way for automation of flexible load operation, which in turn could reduce the users' burden for manual load shifting [47] and response fatigue [48]. Specifically, automation scenarios could enable the implementation of effective and acceptable dynamic pricing [49-51] in the electricity market such as real-time pricing (RTP), which is typically difficult to implement for manual scheduling of energy use due to high variation in pricing and higher uncertainty in user response [49, 52]. From the HEM automation perspective, several energy-intensive devices, such as wet appliances (tumble dryers, washing machines, and dishwashers), electric vehicles (EV), air conditioning (AC) systems, and water heaters could be adopted for flexible operation by receiving the signal from an operator. For example, wet appliances or EV can shift their operation or charging time to a later time during peak demand, or AC systems can adjust the temperature setpoint within the users' preference ranges. In this context, flexibility has been formally defined as the potential for modification of appliance power profile by adjusting power draw, the operation duration, and/or the activation time [34]. Several experimental studies have shown the applicability of load control through automation by leveraging the flexibility of individual loads at needed times with little or no impact on users' convenience [44, 53, 54].

In recent years, several research efforts have shown that the aggregate contribution of deferring individual loads in residential units could facilitate the peak demand reduction and provide ancillary services (e.g., [44, 54-57]). These studies have focused on assessing and quantifying the flexibility, offered by individual loads, as well as the associated response from users for automated DR on networks of households. In assessing load flexibility potentials, a number of factors should be taken into account. (1) Diversity in appliance types and interactional behaviors [13] makes the aggregate load profiles widely different and brings about a high level of uncertainty in estimating flexibility across different households [48]. Even for users with the same appliance type, the user-

load interaction patterns could be considerably different, which results in differences in load flexibility potential. (2) In the adoption of automated DR technologies, targeted and stratified engagement of users is of economic importance in incentivization under the constraints of limited resources [58]. Specifically, introducing DR dynamic pricing scheme comes with constraints such as additional costs for enabling technologies and customer enrollment and marketing investment [59, 60]. Such barriers necessitate the identification and enrolment of only high-potential users, whose price responsiveness to dynamic pricing can be more beneficial [61]. (3) In idealistic scenarios through an unjustified selection of individual loads, the DR objective might not be satisfied, and the rebound effect (generating a new peak) can occur [62]. Synchronized activation of a large number of deferrable appliances to target a peak time [63] could actually jeopardize the system reliability and create another unforeseen peak [47]. Therefore, justified selection of participants under the uncertainty constraints of load type, load demand, and user behavior is the objective of our study on self-sustained smart grid operations.

Accounting for the above factors, it is imperative to consider the users' historical consumption and their variation of usage to effectively leverage the opportunities for load flexibility potentials. Specifically, individual loads' usage pattern has shown to be an important factor for energy demand characterization while its influence on DR objectives has been less explored [64, 65]. Although smart meter aggregate-level data analysis has been used for user segmentation in DR applications (e.g., [13, 66, 67]), such investigations on an individual load basis, with applications for automated DR programs, has been less explored. Accordingly, in this paper, we have proposed a data-driven approach for user segmentation by leveraging individual load consumption patterns and statistical indicators of user-load interactions. Relying on the proposed segmentation strategy, through a case study by using the data from the Pecan Street Project [68], we have further investigated the DR capacity of the targeted selection of households according to user-appliance interactions and their associated usage pattern. Furthermore, using the proposed segmentation approach, we presented and evaluated a comprehensive cross-appliance comparison of demand reduction potential at a community level using real-world data. The method is applied to a sample of more than 300 households, primarily located in Austin, TX.

The rest of the paper has been structured as follows. Section 2.2 covers the related literature. Section 2.3 defines the segmentation approach for ranking users and its constituent parameters. Section 2.4 presents the applicability of the method through empirical assessment for estimating

lower DR capacity and peak load reduction. Section 2.4.4 discusses the implications of the study, followed by concluding remarks and future research directions.

2.2. Research Background

The research background in this study has been presented from two perspectives of (1) residential load flexibility assessment for different types of loads regardless of user consumption behavior, and (2) leveraging users' consumption behavioral patterns in DR targeted operations.

2.2.1. Flexibility potential assessment of residential loads

Residential loads could contribute to the flexibility of grid operation through two classes of deferrable and thermostatically-controlled loads (TCL) [69]. Accordingly, one of the primary directions of research has been focused on assessing the potential of smart appliances operation for load shifting regardless of the patterns of user-load interaction/consumption. In what follows, a number of the major studies are described to provide an insight on the research directions into this domain. In one of the leading efforts, through a field study on 77 households with solar power in the Netherlands, Kobus et al. [54] have explored the demand shift from smart washing machines in a field experiment. They used dynamic pricing to encourage user compliance for automated load shifting within a 24-hour window at an optimum time. The user acceptance to shift the demand was found to be relatively low (14%) given the extended window for shifted operations. In a later study, D'hulst et al. [44] investigated the flexibility potential of all wet appliances (i.e., dryers, washing machines, and dishwashers), as well as EVs and water heaters, in more than 180 households in Belgium. Compared to the previous study, the engagement in smart and automated operation was increased (varying between 30~50%) mainly due to the increased incentive and authority of users in defining the allowable operational delay window. They have stated that varied levels of demand change could be observed from different load types at different times of the day. In this regard, EV and water heaters showed higher potential compared to wet appliances [44]. Klassen et al. [70] performed a field experiment for flexibility potential assessment of 188 households in the Netherlands for smart washing machines for both manual and automated DR. They showed the potential of automated DR and reported success in shifting the load to off-peak pricing times resulting in 31% of load reduction during the evening.

In addition to investigating the flexibility potential of deferrable appliances, studies have also sought to quantify the flexibility of thermostatically controlled load (TCL) (e.g., [57, 71, 72]).

Assessing the load flexibility for different ambient temperatures and setpoints of air conditioning systems [57, 72], as well as refrigerator and water heaters [57] for different building types comprise the main direction of research for flexibility quantification of TCLs. The use of elasticity component of the residential loads is another alternative for creating flexibility capacity [73]. Elasticity refers to reducing the power demand (e.g., reducing the heating load of a dryer) at the cost of increasing the duration time assuming that smart appliances could offer such flexibilities. These efforts have adopted simulation as a methodology for assessing flexibly potential for different building types and geographical locations without explicitly accounting for diversity in user consumption patterns, which stem from differences in interaction habits of users with appliances.

In recent years, in order to leverage the aforementioned load flexibility potentials, Home Energy Management (HEM) systems have been introduced and explored (e.g., [53, 54, 74-83]). These efforts mainly have focused on optimal load scheduling for a single user and adopting effective DR dynamic pricing. In this work, looking at a network of buildings, we have investigated the quantification of load flexibility potential from the user segmentation perspective as the first step of targeting and prioritizing homes for distributed energy management initiatives.

2.2.2. User identification for DR: consumption data analysis

Several previous efforts have focused on the identification of suitable users for DR (e.g., [13, 67, 84, 85]) by using historical behavior of households, reflected in aggregate power consumption data. To this end, studies that leveraged historical data are mainly categorized based on clustering the daily load profiles or developing DR selection functions. In the first category of studies [13, 86-88], clustering has been used on daily aggregate consumption profiles of households to identify various repeating load shapes and their similarities and differences. Different techniques such as two-stage k-means [13, 86], hierarchical clustering [19], self-organizing maps (SOM) [16], and fast search and find of density peaks (FSFDP) [87] have been used for the user segmentation purposes. The selection of suitable users for participation has been carried out based on the shape and power magnitude of clusters, the variability of load shapes, and the distribution of different load shapes in each household [13, 86, 87]. Within this context, the rationale for user selection for direct DR control is to identify households with high-consumption and low-variability (i.e., less variation across subsequent days) load shapes [87]. Therefore, the user selection through aggregate load profiling has been mainly focused on visual analytics of load profile clusters. However, the

potential opportunities for demand reduction based on user selection have not been explored in a new set of data as a test set for further evaluations. On the other hand, selection functions as a quantitative approach for identification of user consumption patterns have been also investigated. For example, using aggregate power, Mammen et al. [67] proposed a function that couples the consistency of consumption, peak consumption, and customer response (learned from the historical events) to categorize user behavior. They studied the potentials of the proposed function by investigating the trade-off between participation fairness and the selection of users with higher consumption using a historical dataset from 60 apartments to simulate DR event. In another example, Holyhead et al. [89] proposed a residential DR approach based on mixed integer programming to target users based on relevance (the likelihood of using deferrable appliances at peak times) and willingness to respond positively to DR requests.

As a common trend, the research efforts on user selection according to consumption patterns have focused on aggregate power data, which could mask the information from individual loads and thus reduce the potential for context-aware automated DR applications. Analysis of individual loads' dynamics could bring about higher information gain on user behavior for context-aware operations. Although several efforts have looked into individual load contributions for DR (as discussed in section 2.2.1), they have not accounted for variance in user interactions with different load types and their impact on the efficacy of the DR process. In other words, all the users and their interaction with the flexible appliances were treated similarly. Nonetheless, there has been a number of research efforts that explored the data-driven impact of individual loads. In a recent effort, Malik et al. [71] investigated the contribution of AC units in summer on peak demand reduction using data from selected houses in Australia. Clustering was used to characterize different consumption patterns for AC units in different households, and a load control strategy was employed to assess the possible demand reduction for each cluster type. It was concluded that around 9% of total peak demand can be reduced trough moderate change in temperature setpoints. In another study [90], a comparison on flexibility potential of different appliances such as EVs, ACs, pool pumps, and lights was performed. However, this study did not account for differences among households in providing flexibility potential.

As the review of the literature shows, in DR operations, behavioral patterns of users in interacting with individual appliances have been less investigated. A Large body of segmentation methods in literature has looked at the aggregate (i.e., whole-house) segmentation, except few recent studies

[71, 91] that focused on AC load segmentation. In other words, to the best of our knowledge, there is no comparative study on flexibility potential of different appliance types according to the user interaction patterns with individual loads. Therefore, we have proposed to leverage human-appliance interaction patterns as another level of information in identifying suitable users for DR engagement. We specifically look at the load flexibility potentials from the perspective of engaging users with different behavioral patterns. To this end, we have proposed a multi-dimensional metric to characterize user behavioral patterns for targeted engagement of users. The objective was to investigate the impact of user interaction with each flexible appliance separately by envisioning emerging Demand-Side Management (DSM) technologies [92] for the smart operation of individual loads for modern grid operation. We have further evaluated the DR capacity of engaging users with different interaction patterns while accounting for the probability of compliance to accept DR requests.

2.3. Methodology

As noted, in this study, we are proposing to learn from the user-appliance interaction patterns to characterize user behavioral traits for effective user segmentation and more informed user engagement. Therefore, the methodology in this study describes a proposed multi-dimensional metric for user behavior identification, as well as simulation studies for evaluating the DR capacity of engaging different groups of households in the load shifting process by accounting for user behavior and compliance.

2.3.1. Data-driven user behavior characterization

In this method, by assuming that the historical data at the appliance level are available, we have proposed to leverage the historical consumption data to quantitatively identify the behavioral patterns of user-appliance interactions that could benefit the goal of load shifting across a community. Through statistical analysis of the historical interactions, we have sought to create a potential score for each household for different load types or appliances. To this end, considering N users in a community, we assign a score S_{ij}^{n} in which i is the user and j is the appliance type. S_{ij}^{n} is calculated by leveraging the daily power consumption for different households and appliances (P_{ijk}) for day k. $P_{ijk}(t)$ is defined as the daily power consumption profile for user i ($i \in [1,..,N]$), appliance type j ($j \in [1,..,J]$), day index k ($k \in [1,..,K]$), and t is time index of the day

($t \epsilon [1:T]$, e.g., $T = 24$ for the hourly electricity consumption data). In order to characterize the user-appliance interaction patterns as S_{ij}^{n}, as illustrated in Figure *2-1*, we have proposed to consider three attributes (i.e., dimensions) [93], namely (i) frequency of use, (ii) consistency of use, and (iii) magnitude of demand during the peak time. All these attributes are measured for a specific load type (j):

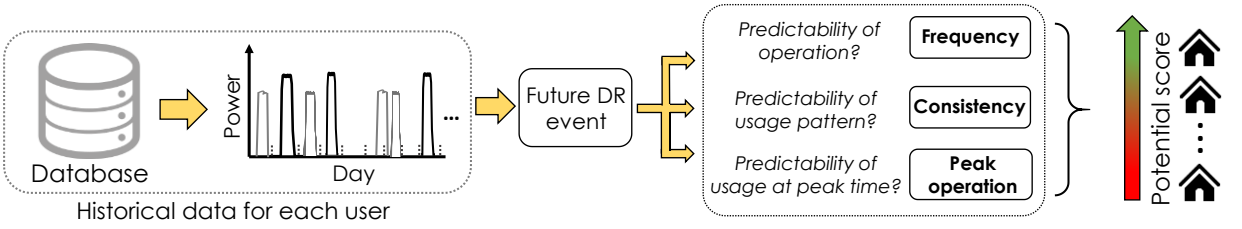


Figure 2-1. Data-driven scoring system for characterization of user-appliance interactions.

Frequency of operation: It is intuitive that some of the targeted appliance types might not be operated by some users on a regular basis. Therefore, it is important to understand the frequency of operation and the tendency of users to use an appliance or a device on a regular basis. Therefore, we have quantified the frequency of operation (FS_{ij}^{n}), in the range of [0,1] as follow:

$$FS_{ij} = \frac{\left|\{k \mid \max(P_{ijk}(t)) > \tau_j \, , k \in (1:K)\}\right|}{K} \quad (1)$$

in which τ_j is a threshold value related to the minimum power draw that an appliance class has. Therefore, FS_{ij} measures the ratio of the number of days that an appliance has been activated compared to the total number of historical days (K) in the analysis. τ_j is used to eliminate the inherent impact of noise in the data or standby power to avoid false detection of operations.

Consistency of operation: An important factor in targeting users in a DR scheme is to understand the consistency of usage [94]. This attribute aims at measuring the extent to which a user's behavior is deterministic or stochastic across subsequent days. Accordingly, from the utility perspective, it is more effective to invest in users with higher consistency, as this factor reflect the likelihood of following the expected usage pattern during the DR event. As a demonstration, we have shown the charging patterns of EVs for two users across 10 days in Figure *2-2*. As shown, user 1 has shown to be more consistent by repeating the same pattern for different days, while user 2 is more sporadic and less predictable in usage patterns. We have defined the consistency of operation (CS_{ij}^{n}), measured in the range of [0,1], as follows:

$$CS_{ij}^{n} = 1 - RMS_{ij}^{n} \tag{2}$$

in which RMS_{ij}^{n} is the root mean square error (RMS_{ij}), normalized across all users in the community using min-max normalization. The non-normalized RMS_{ij} is defined as follows:

$$RMS_{ij} = \begin{cases} \sum_{k \in K_{op}} \sqrt{\sum_{t=1}^{T} \left[P_{ijk}^{n}(t) - \bar{P}_{ij}^{n}(t)\right]^2} & j \in \begin{Bmatrix} EV, dryer, washing\ machine, dishwasher \\ or\ other\ deferrable\ loads \end{Bmatrix} \\ \sum_{k \in K_{op}} \sqrt{\sum_{t=t_1-\frac{t_2-t_1}{2}}^{t_2+\frac{t_2-t_1}{2}} \left[P_{ijk}^{n}(t) - \bar{P}_{ij}^{n}(t)\right]^2} & j = AC\ or\ TCLs \end{cases} \tag{3}$$

in which K_{op} is the set of days that an appliance was operational (defined in the numerator in Eq. (1)), $\bar{P}_{ij}^{n}(t)$ is the average of normalized daily profiles over the span of K historical days, $[t_1 : t_2]$ is a potential DR timeframe, and $P_{ijk}^{n}(t)$ is the normalized values of power consumption profile on each given day as calculated as follows:

$$P_{ijk}^{n}(t) = \frac{P_{ijk}(t)}{\max(P_{ijk}(t))}, k \in K_{op} \tag{4}$$

The RMS_{ij} (Eq. (3)) measures the deviation of the observed value compared to the average across all days that an appliance was operated. For AC or in general TCL loads, we limit the consistency measurement to the vicinity of DR timeframe, due to the fact that they might have multiple daily cycles. For deferrable loads, since the number of activations is limited per day, the consistency could be measured across the entire day. The normalization in Eq. (4) is to avoid biases in comparing the errors given the same appliance class shows varying levels of power draw across multiple users.

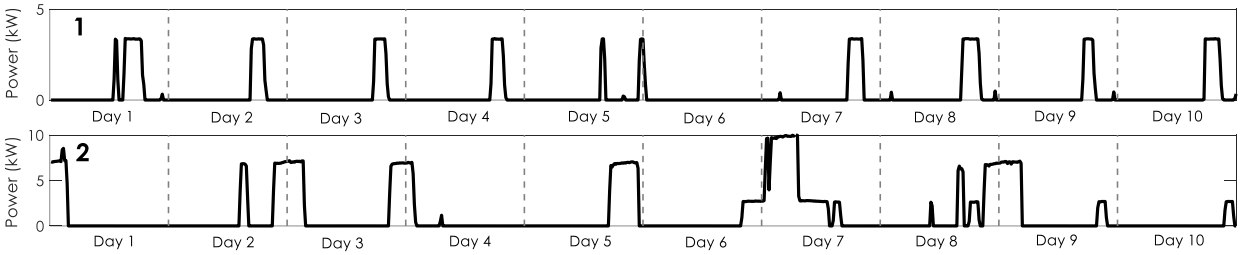


Figure 2-2. Consistency of usage pattern for two users across 10 days

Peak time operation: using a pre-defined demand management timeframe, for example, a DR event timeframe, the historical usage pattern during the event timeframe could be characterized. Accordingly, users with higher consumption during DR timeframes are more suitable for load shifting or shedding. Assuming a DR timeframe of $[t_1: t_2]$, we define PS_{ij} as:

$$PS_{ij} = \sum_{k \in K_{op}} \int_{t_1}^{t_2} P_{ijk}(t)\, dt \tag{5}$$

PS_{ij} indicates the energy consumption during the DR timeframe across historical days. Using min-max normalization for all users, the $PS_{ij}^n \epsilon\, [0,1]$ is calculated to account for power draw variations for the same load type.

Potential score: Using the aforementioned attributes (dimensions) of frequency (FS_{ij}^n), consistency (CS_{ij}^n), and peak time use (PS_{ij}^n), we define S_{ij} as

$$S_{ij} = FS_{ij}^n * CS_{ij}^n * PS_{ij}^n \tag{6}$$

The attributes are multiplied to penalize the S_{ij} if either of them is low (e.g., users that show frequent and consistent use with low energy consumption during DR timeframe). The superscript n for all attributes indicated that all values are mapped in the range of [0:1] over the entire community.

For better interpretability, the min-max normalization of S_{ij} across the entire community is performed to obtain the normalized potential score $S_{ij}^n \epsilon [0,1]$:

$$S_{ij}^n = \frac{S_{ij} - \min(S_{ij})_{\forall i}}{\max(S_{ij})_{\forall i} - \min(S_{ij})_{\forall i}}, i \in [1, \dots, N] \tag{7}$$

S_{ij}^n is applied to rank the users for different load types (j). As the first stage of user engagement, this metric has been intended to be used for user engagement prioritization in a DR scheme for different load types. In other words, the score is used to consider the *relevance* factor, as the suitability for providing flexibility. However, user compliance should be also considered. If the consumption patterns are driven by operational urgency, it is less likely that users comply. On the other hand, if consumption patterns are habitual, it is more likely that users comply. Nonetheless, the level of flexibility in compliance is another factor that should be considered. The compliance could be quantified through direct communication or through statistical analysis in historical user response to DR events.

2.3.2. User-centered load shifting framework

In order to quantify the aggregate impact of load shifting/shedding across a community for different groups (i.e., segments) of users, we have adopted a load shifting framework that leverages the user characterization model in section 2.3.1. Figure *2-3* illustrates the general framework, which leverages the ground truth data to simulate the impact of load shifting/shedding to off-peak time following a set of standard protocols, described in the following sections.

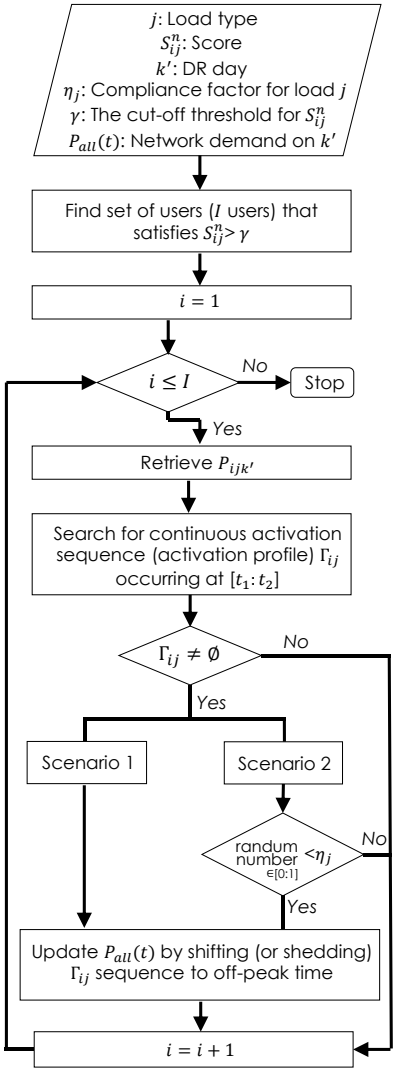


Figure 2-3. Load shifting/shedding simulation framework based on user-appliance interaction patterns

2.3.2.a. Load shifting/shedding simulation scenarios

In order to quantify the DR capacity of the proposed score in load shifting, we have simulated the use of the above framework through a case study of a real-world community. In this simulation,

for each DR event day, we retrieved the daily consumption profile for user i and appliance j. During the DR timeframe $[t_1: t_2]$, the operational status of an appliance class is examined by searching for available continuous activation load sequences (i.e., adjacent data points) above the limiting τ_j. If no operation is detected, the power profile remains the same. Otherwise, the activation profile, $\Gamma_{ij} = [t_{beg}: t_{end}]$ is identified, in which $t_1 \leq t_{beg} < t_2$ (i.e., activation start is within the DR timeframe), and t_{end} could be either within DR timeframe or stretched to a later time. For load shifting from Γ_{ij}, we have considered two scenarios:

Scenario 1: Minimum temporal deferral: In this scenario, we have assumed that the deferred load will be immediately operated once the DR timeframe (i.e., the peak time) is over. Studies have shown that there is a higher probability that users shift the loads to an immediate timeframe after the DR event [95]. In other words, users prefer the minimum delay in the activation start. In this case, we assume the new activation cycle will happen at the $\Gamma_{ij}^* = [t_2: t_2 + t_{end} - t_{beg}]$ timeframe.

Scenario 2: Temporal deferral in a user flexibility window: In this scenario, each user defines an allowable flexibility window for each load type upon agreement for load deferral [44]. The flexibility window is within the comfort limit of users. In this way, the loads could be shifted to off-peak time when the electricity price is lower. Therefore, the new activation cycle will happen at the $\Gamma_{ij}^* = [t_{beg'}: t_{end'}]$ in which $t_{beg'} > t_2$ and $t_{end'} - t_{beg} < max_d$. Here, max_d is the maximum allowable delay time defined by the user.

2.3.2.b. User compliance simulation scenarios

To present the applicability of the method, we select multiple days that were *not* previously considered as the historical days for ranking the users. To this end, we randomly select five days after the historical days to quantify the energy reduction at the peak time and report the average result. We select multiple days to impose adequate variations for empirical demonstration. In quantifying the energy reduction during peak time, we have considered two alternatives for emulating user response:

Scenario 1: Maximum potential:

In this scenario, we allow for the load shifting of a load only if its consumption time coincides with the peak timeframe to evaluate the maximum load shifting potential. In other words, it was assumed that users always comply with a DR request signal. Therefore, the results represent the

upper bound for energy reduction potential. Since deferring the load to an immediate timeframe after the DR window is more consistent with user tendency [95], we used the minimum deferral scenario (described in section 2.3.2.a) for shifting the operation cycle for deferrable loads right after the DR timeframe ends (t_2).

Scenario 2: User-compliance factor:

Successful DR events consider users' preferences. User compliance and the flexibility window (allowable time for load deferral) are contextual attributes that drive the users' response to a DR event. User compliance indicates whether a user accepts a DR signal or configurations for automated operations of smart appliances for participation. The flexibility window defines the allowable timeframe for load deferral without compromising user comfort. In the literature, there exist a few experimental and field studies on characterizing users' response for automated load deferral. In our simulations, we have adopted the user response output from a large-scale experimental pilot project over a community of houses monitored for 3 years [44]. The compliance factor, reflecting the average of acceptance to DR signals for each load, are as presented in Table *2-1*. For the flexibility window, the probability distributions shown in Figure *2-4* were adopted from the same study, which reported the average flexibility windows for EV, dryer, washing machine, and dishwasher as 5.6, 8.1, 7.3, and 8.5 hours, respectively [44]. AC was not included in Figure *2-4* as the control mechanism is different and is based on temperature setpoints.

Table 2-1. Compliance factor (parameters partially adopted from [44])

Device	Compliance factor threshold
EV*	0.60
Dryer	0.31
Washing machine	0.29
Dishwasher	0.56
AC**	0.50

* was not specified in ref [44] and we assumed its compliance factor threshold . In the reference, it was stated that a majority of smart configurations occurred in the evening.

** AC was not included in the pilot study, and we assumed its compliance factor threshold.

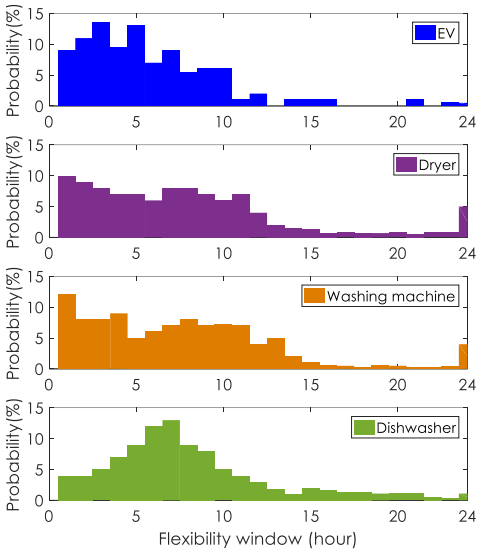


Figure 2-4. The probability distribution of flexibility windows for different load types (adopted from [44]).

For each owner of the deferrable loads, we used the *compliance threshold* specified in Table *2-1* as the average fixed response of community for engaging in a DR event. A random number from a uniform distribution (between 0 to 1) was generated for each user as the *compliance response*. If the compliance response was lower than the compliance factor threshold, DR participation (i.e., agreement) was positive. Upon agreement, the value of the flexibility window (duration of load deferral) was selected from the probability distribution function (shown in Figure *2-4*). We ran the results 10 times on each DR day in order to account for the fact that different users have varying compliance factors, and the values shown in Table *2-1* are used as the average of the community. In this scenario, to account for the possibility of load shifting to a next day (if the flexibility window allows), we extracted 48-hour power profiles. Figure *2-5* shows different combinations of scenarios from section 2.3.2.a and section 2.3.2.b.

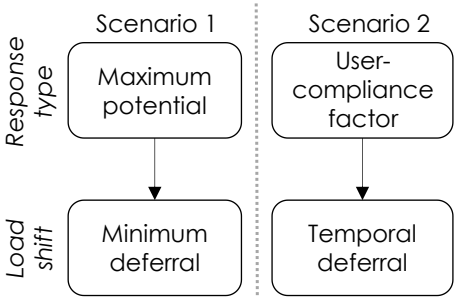


Figure 2-5. Combination of scenarios in simulation analyses.

2.3.3. Case study community

We hypothesized that user selection based on human-appliance interactions could be effective in realizing a large ratio of possible demand reduction in the entire community by engaging a small portion of users (for a specific load type). To assess this hypothesis, we used the real-world consumption data from the Dataport Pecan Street project [68], which is an ongoing project on monitoring the appliance-level and aggregate-level data for more than 1000 households, primarily located in Austin, TX, from 2011. Monitored houses have participated at different time periods. Therefore, some of them have opted-in and opted-out from the beginning of the project, and the number of houses has been subject to change over the years. In this study, we have retrieved the data for all households that were monitored during July and August 2015, reflecting high AC use as well. We used 15-minute resolution data, indicating that each daily profile includes $T = 96$ datapoints. 15-minute resolution data was used since (1) it was enough to capture the operation/charging of considered flexible appliances in this study and (2) could be acquired through smart meters [96]. Daily power profiles of appliances were used to examine the complex human-building interaction at the level of individual loads.

Consumption daily profiles for EV, dryer, washing machine, dishwasher, and AC were extracted for different users. The total number of households that participated in the data collection process during the period of our study was 307. The data set information is presented in Table 2-2. The second column shows the number of households or owners, and the third column shows the number of households with viable datasets. In some of the houses, either the devices were not used over a long span of time or the datasets had considerable missing data points. The last column shows the number of daily profiles for different load types.

Table 2-2. Characteristics of the selected community data set.

Device	# of owners	# of viable datasets	# of daily profiles
EV	85	78	4909
Dryer	184	175	10797
Washing machine	190	93	11220
Dishwasher	224	176	13208
AC	282	276	16560

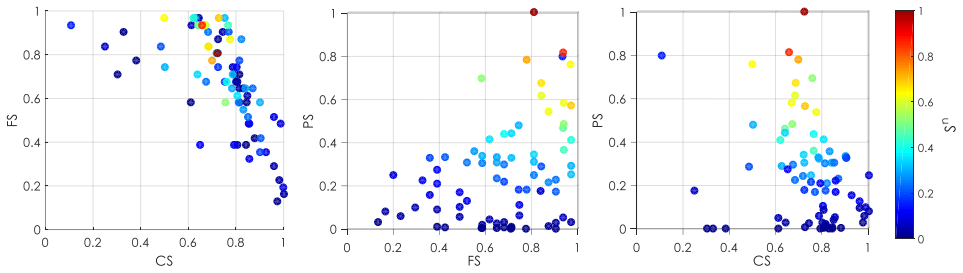
2.4. Data analysis and results

The assessment of the hypothesis, the demonstration of the potential score, the results of load shifting under different scenarios, and a cross-comparison of the flexibility for different load types at different time-of-use across a 24-hour period have been presented in this section.

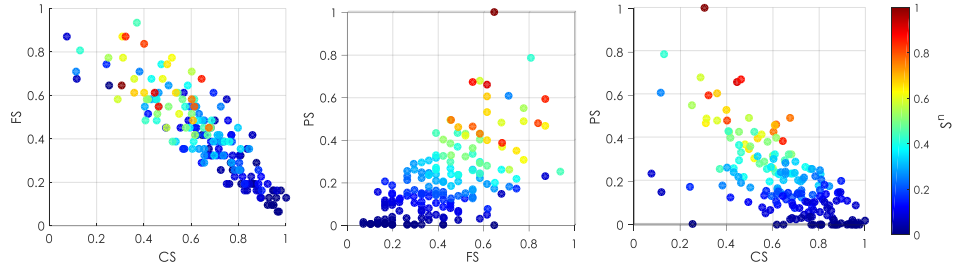
2.4.1. The effectiveness of the potential score

In this study, except otherwise specified, a DR timeframe of $[t_1: t_2]$ from 17:00 to 19:00 was used. This timeframe is compatible with our peak time observations in the aggregate consumption patterns of the case study community, as well as the common timeframe in practice (e.g., [97]). The 2-hour duration has been also commonly investigated in the literature [66, 89, 98]. Nonetheless, the selected timeframe is an input to the model, and the score S^n could be calculated accordingly. Through a sensitivity analysis (as follows), we selected one month of historical days (K) in the analysis. For τ_j, values of 1kW, 0.8kW, 0.3kW, 0.5kW, and 0.5kW were used for EV, dryer, washing machine, dishwasher, and AC, respectively. These values were selected based on typical power draw values of appliances [99, 100] in addition to visual inspection of the dataset.

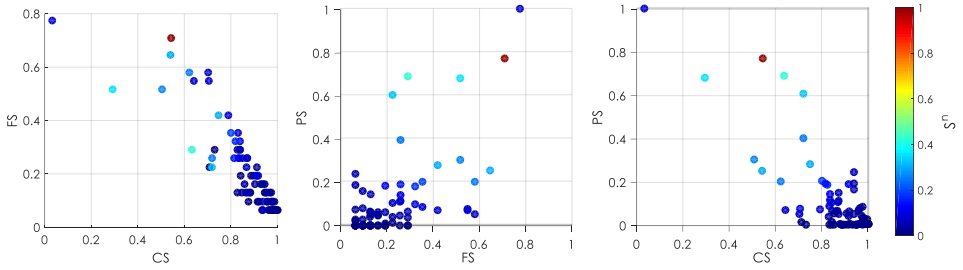
The distribution of the potential scores (S^n) for EV, dryers, washing machine, dishwasher, and AC are presented in Figure 2-6. Each subplot represents one load type and each data point represents one of the households in the community. Each of the axes represents one of the dimensions (i.e., FS^n, CS^n, PS^n) and variations of S^n values have been illustrated by a color heat map. In subplot (a), it could be seen that the potential scores for EV charging are distributed across all three attributes in the community, indicating highly varied usage styles and operational frequencies in the community. In subplots (b-d), which represent wet appliances, there is a higher concentration in distribution of houses around lower values of FS^n, indicating these appliances are mainly not operated on a regular basis. This is also consistent with the expected daily routines of users. However, higher CS^n values could be observed, reflecting on consistency of the time-of-use of the appliances by users. In the case of AC in subplot (e), a majority of data points show higher values of FS^n, reflecting regular use of AC on a daily basis. Given that the data set is associated with summer days, the remaining few data points with low FS^n could be associated with unoccupied houses.



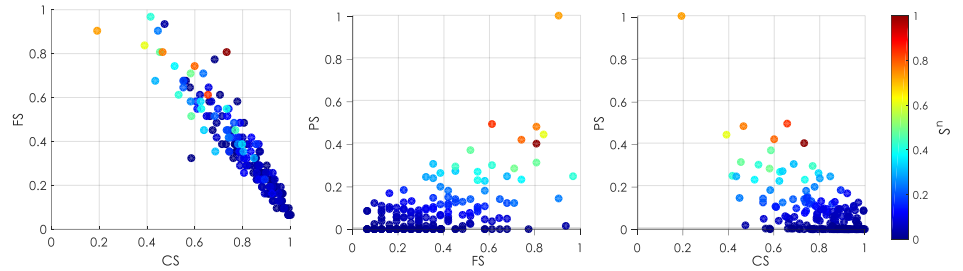
a) EV



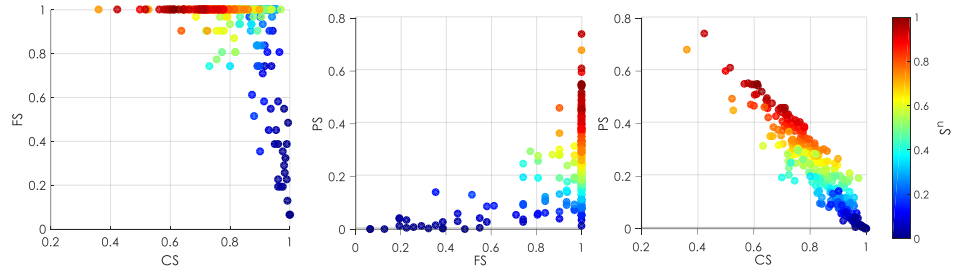
b) Dryer



c) Washing machine



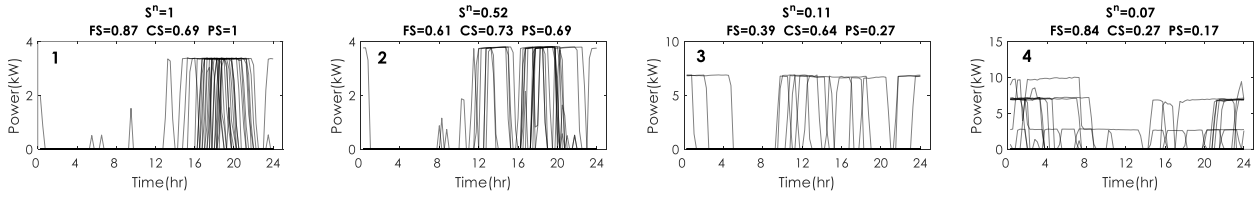
d) Dishwasher



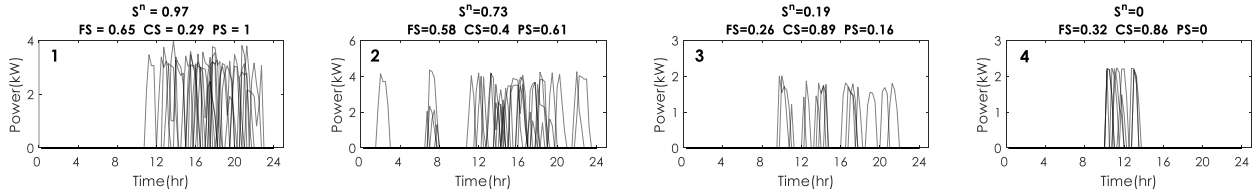
e) AC

Figure 2-6. Distribution of S^n for different a) EV owners, b) dryer owners, c) washing machine owners, d) dishwasher owners, and e) AC owners.

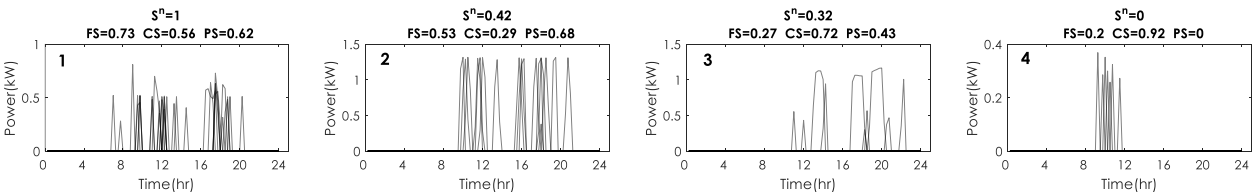
The power variations for different potential scores have been presented in Figure 2-7 to provide an insight into their interpretation for several representative users. Each subplot illustrates the consumption profiles across all historical days ($K = 30$) with their associated potential scores and relevant dimensional values. To provide an example explanation of these visualizations, we have elaborated the observations for the EV data. User 1 charges the EV almost every day (26 out of 30 days, $FS^n = 1$) with a relatively consistent pattern ($CS^n = 0.69$), with the highest consumption in the community during the desired DR timeframe ($PS^n = 0.1$) resulting in a potential score of $S^n = 1$. User 2 has a relatively high potential score ($S^n = 0.52$). Compared to User 1, this user charges EV less frequently ($FS^n = 0.61$), with lower consumption values ($PS^n = 0.27$) while being more consistent across the activation days ($CS^n = 0.73$). User 3 and user 4 both have lower a potential score ($S^n = 0.11$ and $S^n = 0.07$, respectively). User 4 charges EV on a regular basis (higher FS^n) with low temporal consistency (lower CS^n) with a charging time that rarely coincides with the DR timeframe (lower PS^n). Therefore, utilities will be less interested in involving these users in a DR event. As can be seen from usage patterns for different load types in Figure 2-7, user 1 and 2 with higher S^n are more suitable for getting involved in a DR event as they have high frequent peak usage and more predictable behavior while user 3 and 4 manifest lower values for either of the dimensions.



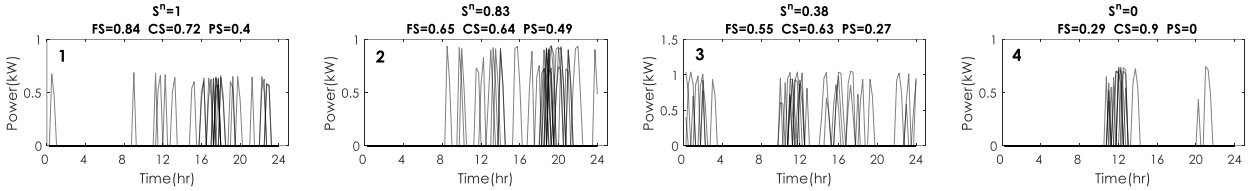
a) EV



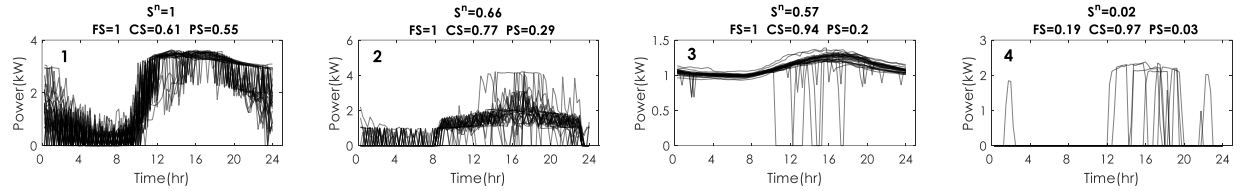
b) Dryer



c) Washing Machine



d) Dishwasher

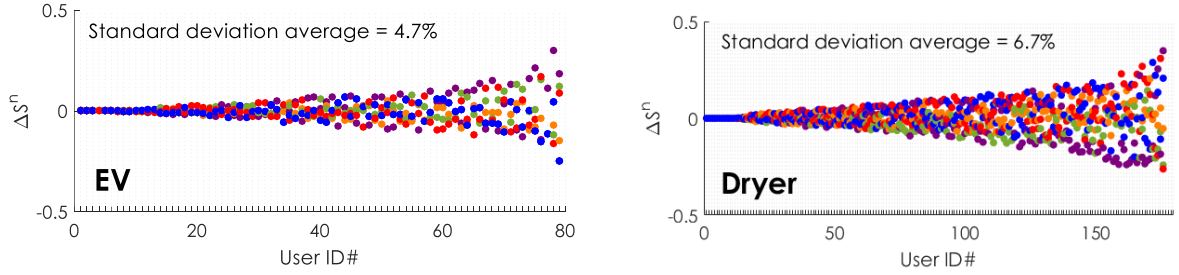


e) AC

Figure 2-7. The daily consumption profiles of four representative users with varied potential scores on 30 subsequent days and their associated dimensional values for: a) EV, b) dryer, c) washing machine, d) dishwasher, and e) AC.

2.4.1.a. Sensitivity analysis:

In order to show the sensitivity of S^n on the number of historical days K, we performed a sensitivity analysis for different K values of 20, 30, 40, 50, 60 days. Figure 2-8 illustrates the results for different load types. For each user ID# (i.e., house ID) on the horizontal axis, 5 different data points corresponding to different K values have been plotted on a vertical line. Users were sorted according to the lowest deviation from the average values of S^n over all K days. In some cases that houses have opted out from the program during a longer timespan (K>30), data points are missing. In some cases, variations of K could result in changes in frequency, consistency, and peak usage, given the potential changes in behavioral traits in user behavior. Nonetheless, the results in Figure 2-8 show a low amount of sensitivity in terms of change in S^n in total. Specifically, the average amount of standard deviation for S^n across all households varied between 2% to 7% for different considered load types, as shown in Figure 2-8.



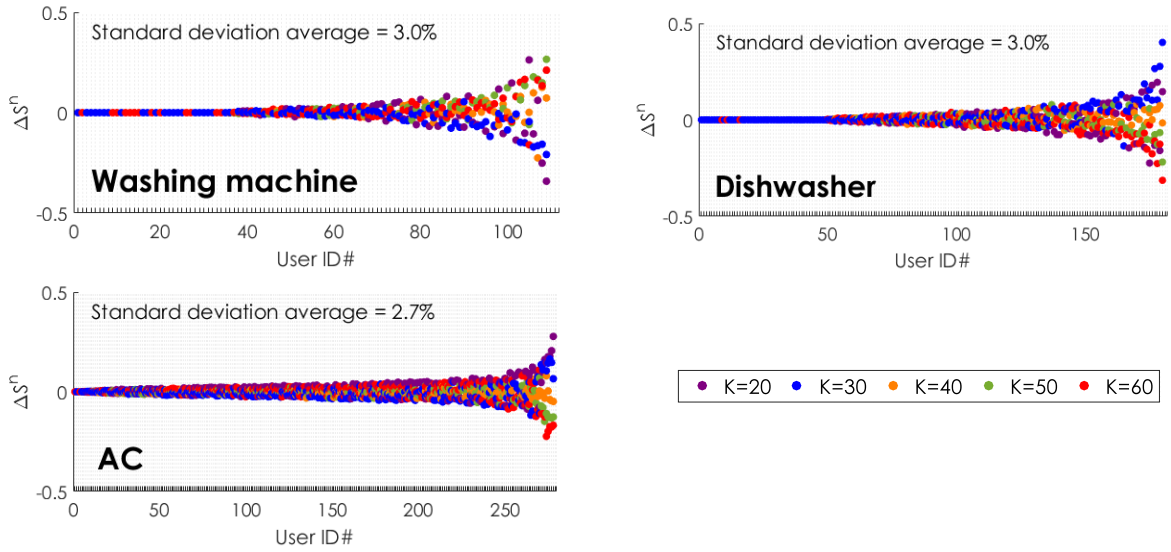


Figure 2-8. Sensitivity analysis result for quantifying S^n across 20, 30, 40, 50, and 60 days for different load types.

Based on the trend in Figure 2-8, we hypothesized that, through varying K, S^n do not show significant changes for each user. We performed the Kruskal-Wallis test to examine whether distributions of ΔS^n for different households are significantly different for different K values ($K = \{20,30,40,50,60\}$). The Kruskal-Wallis was selected for the comparison of multiple groups of days (different K values) as a non-parametric method since ΔS^n values did not follow the normal distribution. For each user, we measured the deviation of S^n for each specific K from its average (over the entire K vector). The Kruskal-Wallis test was performed on the considered 5 groups (different K values), in which each group contained ΔS^n values for users in the community. The null hypothesis was that the population medians for all groups of K values are all equal for the case-study community. The *p-values* in analysis for all flexible appliances were above 0.05 (as shown in Table *2-3*), indicating the fact that the null hypothesis is accepted and ΔS^n belongs to the same distribution in all cases.

Table 2-3. Kruskal-Wallis test on the set of K (different number of days).

	EV	Dryer	Washing machine	Dishwasher	AC
P-value*	0.99	0.10	0.30	0.23	0.70

* P-value > 0.05 indicates the equal population medians of ΔS^n for the set of K

2.4.2. Empirical assessment of DR capacity

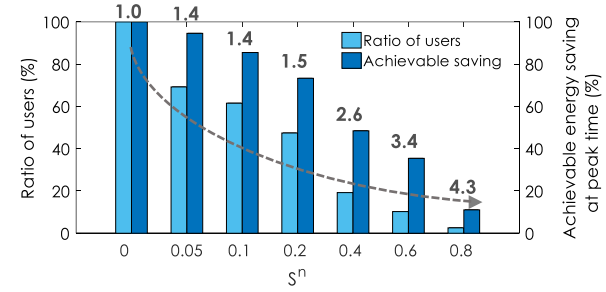
In this section, we have evaluated the efficacy of the proposed scoring system in the smart engagement of the users in a DR event. To this end, we selected five DR days, *not* included in historical days to avoid *a priori* biases in DR flexibility assessments. Using different cut-off threshold values for S^n (0, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8), different groups of users were selected for participation in DR events. We have quantified the potential energy reduction for each user and each load type by measuring the differences between consumed energy at the peak timeframe $[t_1{:}t_2]$ (using ground truth) and the resultant time-series obtained by load shifting/shedding. We used numerical integration on the power time-series to calculate energy consumption.

We used load shifting for deferrable loads (i.e., EV, dryer, washing machine, and dishwasher). For AC, as a TCL load, we used partial load shedding through changing the temperature setpoint for simulation. In doing so, we adopted the assumptions and findings outlined in [71, 101] – a 25% power reduction during peak time by 1°C increase in temperature setpoint of AC. Although the findings of these studies were dependent on their context, we used their assumptions to be able to estimate the aggregate contribution of different segments of users in demand reduction. Sophisticated models for thermal behavior of these houses could be developed by using simulation tools such as energyPlus. However, such model require detailed information from the buildings and is out of the scope of this study. Using the load shifting/shedding framework described in section 2.3.2, we considered different scenarios as presented in Figure *2-5*.

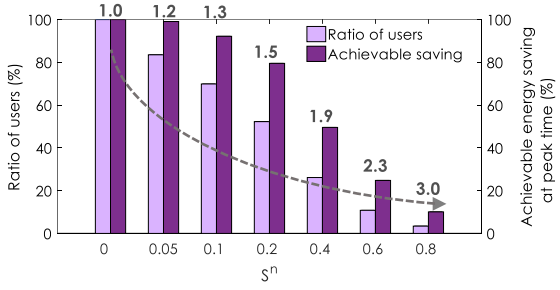
2.4.2.a. First scenario - maximum potential

Figure 9 illustrates the achievable energy saving for different S^n values and load types. The right and left vertical axes show the achievable energy reduction during the peak (from the entire community) and the percentage of engaged users, selected according to the S^n values, respectively. The horizontal axis shows different cut-off values for user segmentation based on the S^n values. In these graphs, $S^n = 0$ indicates the participation of the entire community and the maximum load shifting potential for activated loads without accounting for the human-appliance interaction patterns. As shown in Figure 9, for all load types, as the S^n cut-off increases, we generally observe that the demand reduction drops with a lower rate compared to the ratio of engaged households. Therefore, it is shown that we could identify and select a small portion of users to achieve demand reduction goals. For example, for EVs (Figure 9(a)), by selecting users with S^n higher than 0.8,

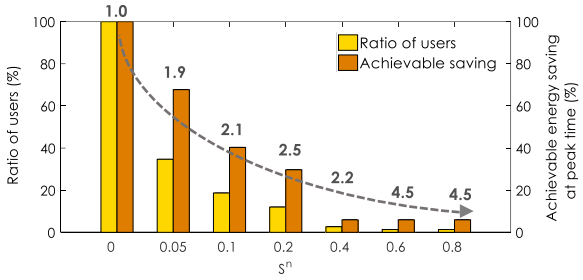
0.6, and 0.4, energy reduction of 11%, 35%, and 49% could be achieved, respectively. These values correspond to only 3%, 11% and 19% of all the households in the community, respectively.



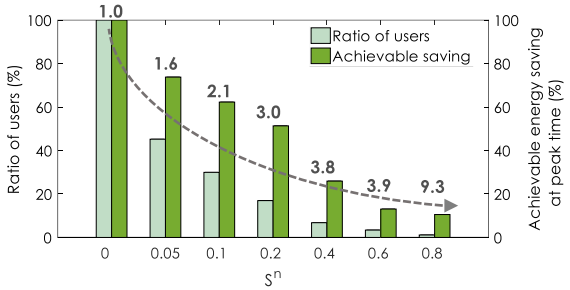
a) EV



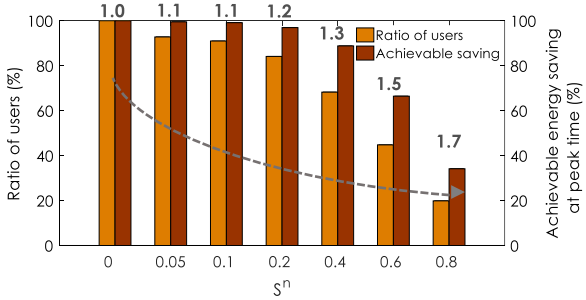
b) Dryer



c) Washing machine



d) Dishwasher



e) AC

Figure 2-9. Load shifting potential for scenario 1 (averaged over 5 days) across different subsets of users. The value on top of the double-bar shows the ratio of 'achievable saving' to 'ratio of users' for each case (a higher value is desired).

In order to provide a cross-appliance comparison of the energy reduction potentials, Figure *2-10* shows the individual contribution of users for different load types. The horizontal axis shows the user IDs, sorted based on ascending values of S^n, and the vertical axis shows the energy reduction potential during the DR timeframe, averaged across five DR days. As the results in Figure *2-10* show for different load types, if the loads have been activated during the peak, there is an increasing trend in saving potential as S^n increases. As can be seen, many users have not

operated/charged the deferrable loads on DR test days during the peak time. On the other hand, given the regularity of usage, AC provides the highest energy reduction potential after EV. Among the deferrable loads, EV provides the highest potential for energy reduction followed by the dryer. Washing machine and dishwasher have shown to provide the least potential for DR operation on the selected test days.

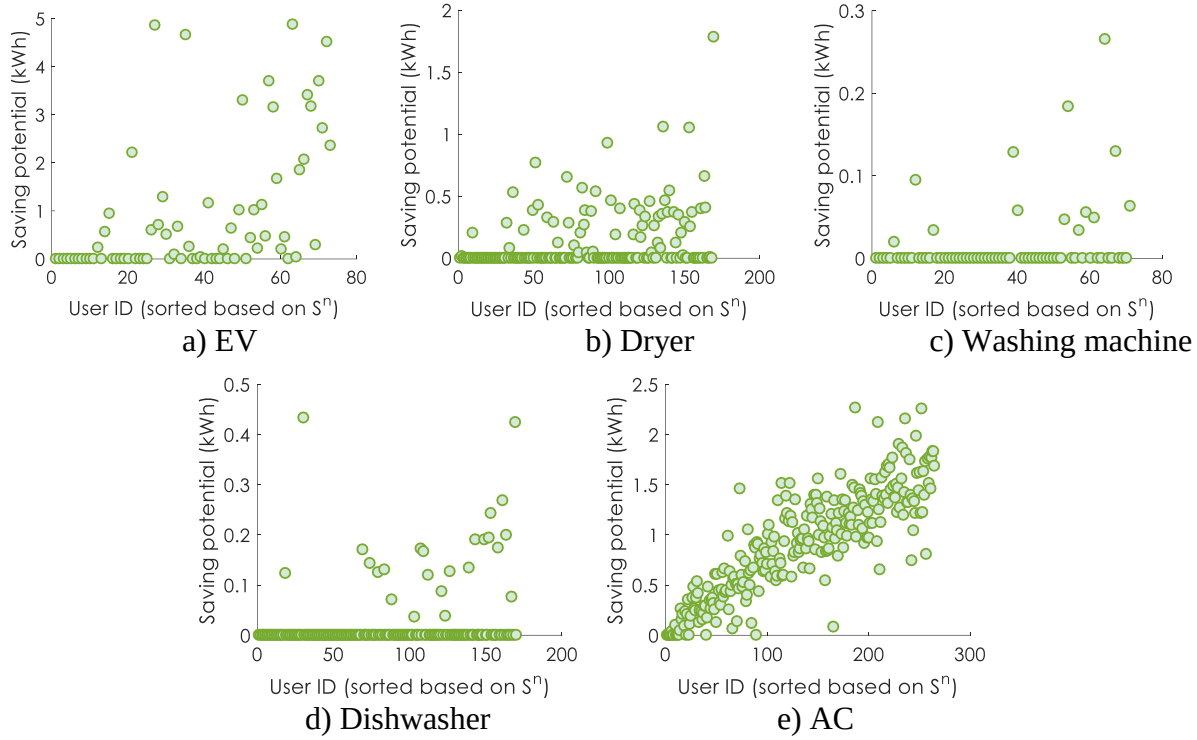
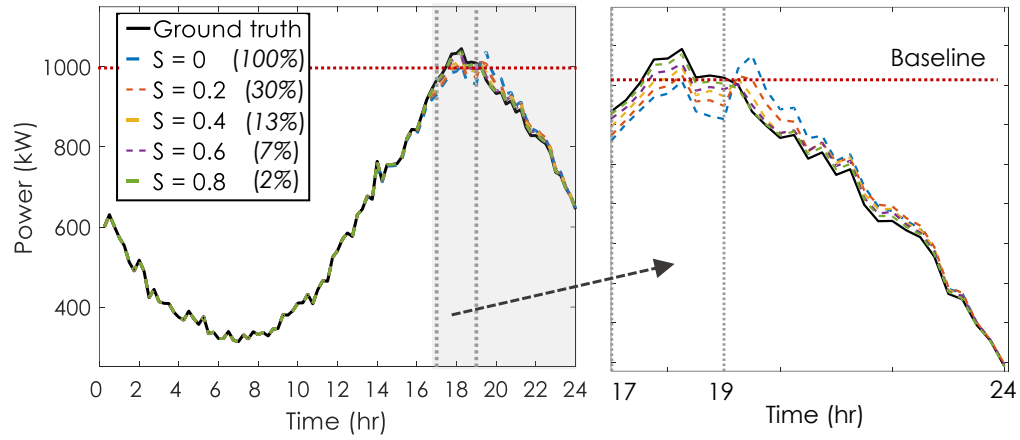


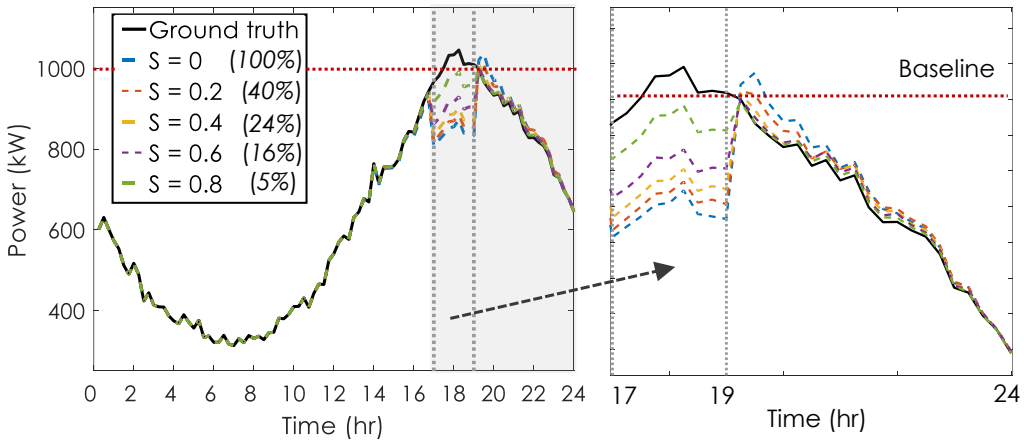
Figure 2-10. Demand reduction potential for scenario 1 (averaged over 5 days) for each user.

As noted earlier, unjustified and synchronized compensation through engaging all categories of loads immediately after the DR event could result in a rebound effect – the creation of a secondary peak. In order to demonstrate the applicability of the method for load selection to avoid the rebound impact, we have shown the aggregate power profile from the entire community on a DR test day in Figure *2-11*. Figure *2-11*(a) illustrates the results from the contribution of all deferrable loads while Figure *2-11*(b) also includes the AC. The ground truth line shows the actual aggregate power for the test day. The resultant load profiles from load shifting according to different segments of S^n are shown with dash lines. The percentage values in the legend of Figure *2-11* shows the average ratio of users engaged in DR. A hypothetical baseline of 1000 kW was considered. As shown in subplot (a), the cut-off value of $S^n = 0.2$, corresponding to the participation of 30% of users, can result in the balance between the primary and the secondary peak and achieving the

hypothetical baseline. On the other hand, if the setpoint control of AC will be considered in addition to deferrable loads (subplot (b)), the same objective can be achieved by selecting $S^n = 0.8$, corresponding to 5% of users. As the results show, engaging users according to their previous interaction patterns could help us achieve the targeted energy reduction without creating a rebound effect compared to engaging all users.



a)



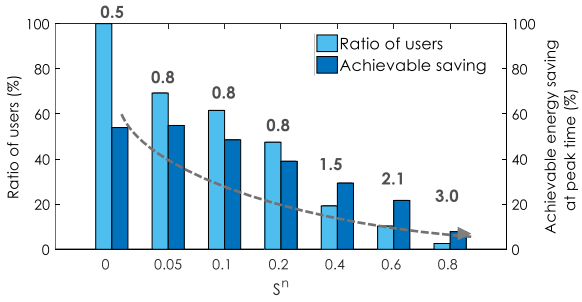
b)

Figure 2-11. Impact of load shifting/shedding (peak and secondary peak) across different subsets of users for (a) all deferrable loads and (b) all deferrable loads plus AC on a DR test day. Numbers in parentheses show the percentage of users engage (averaged over considered loads) in DR.

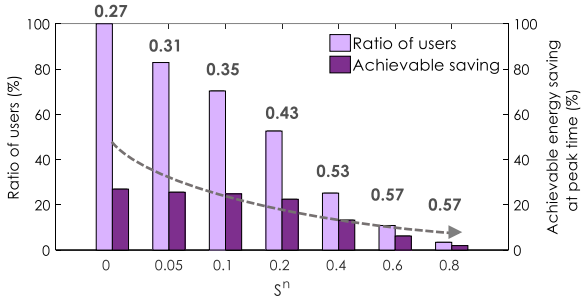
2.4.2.b. Second scenario - user compliance driven

Figure *2-12* shows the achievable energy reductions according to different S^n values for the second scenario, which represents a more realistic user engagement. Therefore, the maximum achievable reduction was observed to be a portion of reductions in the first scenario. The ratio between these two observations is close to the associated compliance factors. The user response to DR request could be negative, even if the activation indeed contributes to high peak demand, and therefore the maximum achievable saving will be limited. Nonetheless, the same trend that a high percentage of achievable energy reductions can be realized by a lower rate of targeted participation was also observed for this scenario.

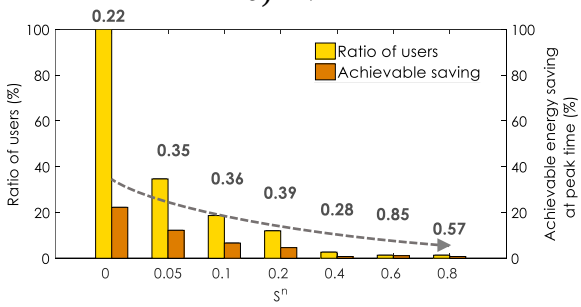
Similar to the previous scenario, Figure *2-13* illustrates the energy reduction potential for the 2-hour DR timeframe for each user. The results for this scenario also revealed the efficacy of the proposed scoring method in the targeted engagement of users in DR events. Given the lower probability of the user compliance for dryer and washing machine, the reduction of flexibility potential for these loads has been relatively higher while AC and EV manifest better opportunities for load shifting.



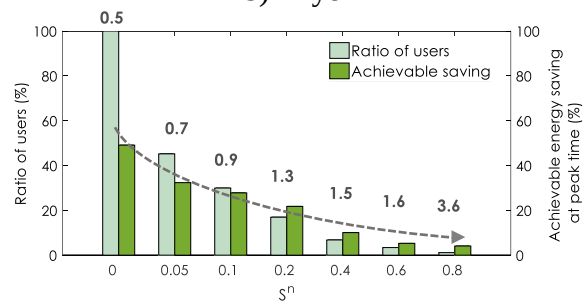
a) EV



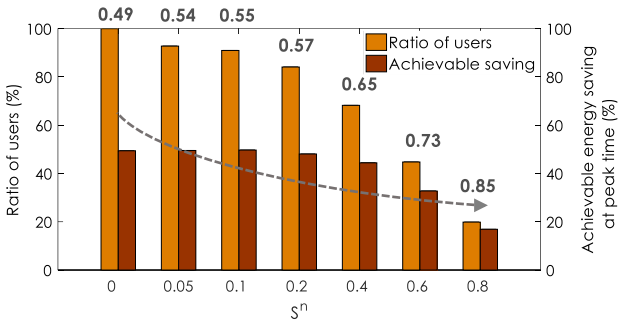
b) Dryer



c) Washing machine

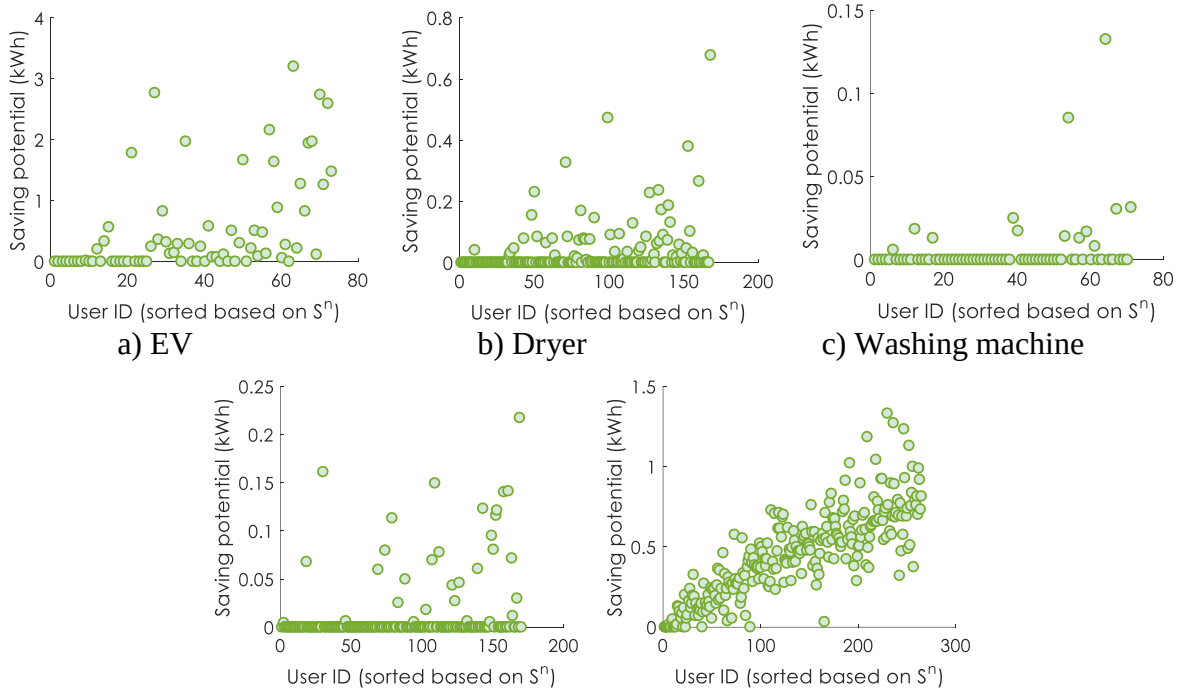


d) Dishwasher



e) AC

Figure 2-12. Load shifting potential for scenario 2 (averaged over 5 days) across different subsets of users. The value on top of the double-bar shows the ratio of 'achievable saving' to 'ratio of users' for each case (a higher value is desired).



a) EV b) Dryer c) Washing machine

d) Dishwasher e) AC

Figure 2-13. Demand reduction potential for scenario 2 (averaged over 5 days) for different users

Table *2-4* presents the numeric values of demand reduction for different scenarios. It is expected that some users might not operate their deferrable loads at all during the DR test events, and in this case, the associated energy reduction will be zero. Accordingly, the averaged values in Table *2-4* could be interpreted as the realizable demand reduction for different segments of users. Given the high variance among different user quantiles for each load type, the potential of the ancillary

services could be estimated and prioritized for different load types. For our case study community, EVs, ACs, dryers, dishwashers, and washing machines show the highest potentials, respectively. The energy saving potentials, presented in Table *2-4*, could be also used as weight factors in case of combining the potential score of individual appliances into a metric at the household level. In that case, the metric would characterize the overall flexibility of the user and reflects the overall evaluation of household potential for individual load flexibility.

Table 2-4. Average demand reduction potential over a 2-hour afternoon DR event based on different quantiles of users for different loads

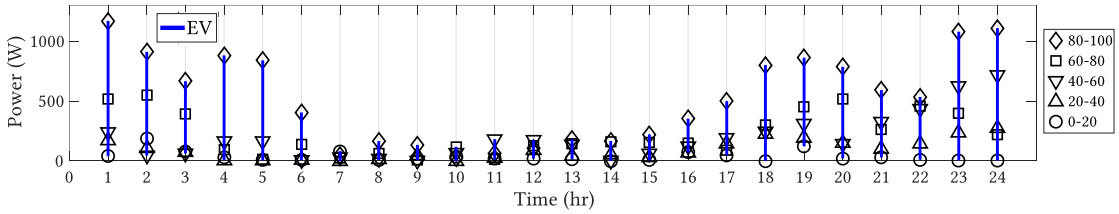
Load Type	User quantile based on S^n	Maximum potential (Wh)	User response-driven (Wh)
EV	80-100	2095	1260
	60-80	1065	550
	40-60	495	280
	20-40	700	435
	0-20	120	75
Dryer	80-100	260	70
	60-80	145	40
	40-60	140	50
	20-40	90	25
	0-20	20	5
Washing machine	80-100	40	15
	60-80	20	10
	40-60	10	5
	20-40	0	0
	0-20	10	0
Dishwasher	80-100	60	35
	60-80	20	15
	40-60	20	10
	20-40	0	0
	0-20	15	5
AC	80-100	1515	760
	60-80	1195	590
	40-60	1010	485
	20-40	620	310
	0-20	210	105

2.4.3. Diurnal variation of flexibility

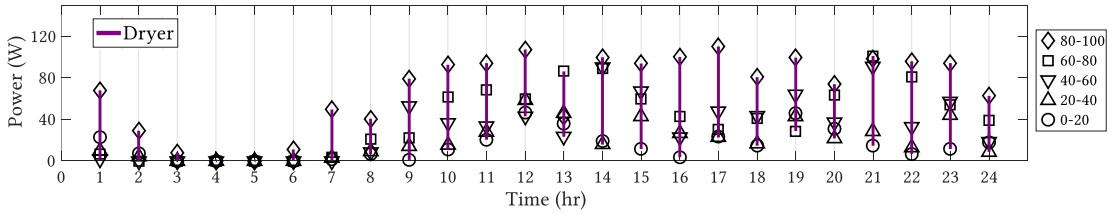
Defining the temporal flexibility of loads (as a function of the time of the day) could provide opportunities for the integration of renewables and maintaining the power balance in addition to the peak demand reduction. Accordingly, analyzing the historical patterns of temporal variations of load-specific demands provides insight into load targeting. To demonstrate the demand reduction potential of the considered load types with respect to different segments of users and

according to time-of-use, we presented the results in Figure *2-14* for maximum potential scenario and based on S^n values on a 1-hour basis. Each bar represents the range of average demand reduction of the associated quantile for a given hour, and each marker shows the average demand reduction for the associated user quantiles. A longer bar indicates high variation in usage style and vice versa. Moreover, a higher concentration of the markers on the lower parts indicate that only a small portion of users could provide a higher power reduction potential. For example, the top quantile of users with EV in the 80-100% quantile (associated with high S^n values) at 18:00 are highly suitable for shifting their demand, while the rest are not contributing as much.

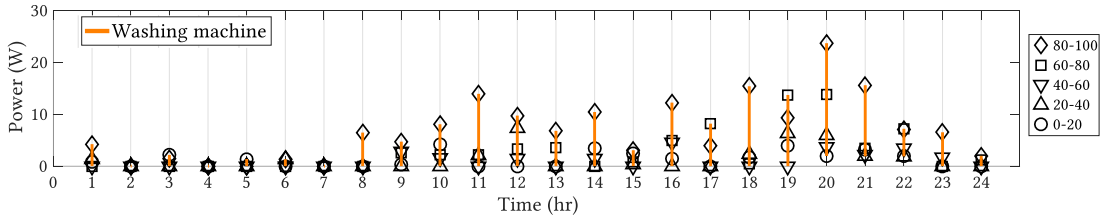
For EV, the diurnal charging pattern reflects a typical home to work commuting lifestyle as the charging time mainly starts in the late afternoon and stretches to early morning hours. For AC, the potential for demand reduction across the entire day is observed. The demand reduction potential reflects the occupancy pattern as well as the typical temperature variation pattern. In general, compared to deferrable loads that show a high uncertainty for demand reduction potential across different classes of the users, AC shows a lower variance and more predictable behavior. Wet appliances manifest a bi-modal distribution with a double peak around noon and the evening. The peak around noon indicates opportunities for solar power integration.



a)



b)



c)

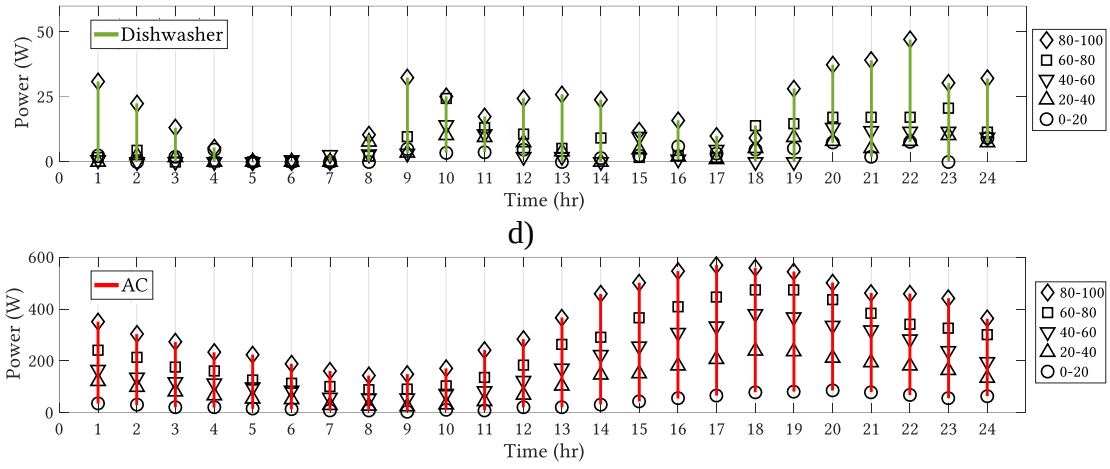


Figure 2-14. Average diurnal demand reduction potential for different user segments: a) EV, b) dryer, c) washing machine, d) dishwasher, and e) AC - The legend shows different quantiles of selected users based on their S^n score.

2.4.4. Rebound impact

In order to illustrate the impact of load shifting to off-peak time on different subsets of consumers, we performed an experiment to assess the impact of consumer segmentation on the rebound effect. To this end, for different subsets of consumers according to their S^n, the EV load signatures, if found on the DR test day timeframe, were shifted to the off-peak time. Figure *2-15* shows the aggregate power demand of the entire community (with 78 EV owners). As can be seen, the participation of the entire community (i.e., a cut-off threshold of S^n= 0 for selecting consumers) results in an undesirable off-peak demand, while selecting S^n as 0.2 or 0.4 (participation of ~20-40% of consumers) could balance the demand during DR time and off-peak time. Furthermore, to investigate the impact of the potential score on the optimal balance between DR time and off-peak time, the following equation was employed:

$$\emptyset = \sum_{i=0}^{2*(t_2-t_1)+1} |cost_i| \tag{8}$$

$$where \begin{cases} cost_0 = 0 \\ cost_{i+1} = P^n(t_1+1) + cost_i - P_b^n(t_1+1) \end{cases}$$

in which $P^n(t)$ and $P_b^n(t)$ are the aggregate power demand and the desired baseline demand at time t, respectively, and $cost$ is the effort for transforming the observed demand curve into the baseline. Both $P^n(t)$ and $P_b^n(t)$ are normalized such that their value at time t is divided by the

sum of values at the DR timeframe. The above equation is based on the Earth Mover's Distance concept [102] and measures the required work to transform the power demand $P^n(t)$ from DR peak to off-peak time into the uniform target baseline $P_b^n(t)$. Therefore, a lower value of $\emptyset$ is desired as it reflects a more uniform distribution of load during and after DR. Table *2-1* demonstrates the $\emptyset$ values for different subsets of consumers. Based on the ground truth curve shown in Figure *2-15*, a value of 350 kW was assumed for the $P_b^n(t)$ as the utility decision. As can be seen, a potential score value of $S^n = 0.2$, corresponding to the participation of 40% has led to the optimal balance of power demand between peak and off-peak time. In contrast, the full participation ($S^n = 0$) resulted in a rebound effect with a high concentration of demand at the off-peak time.

Table 2-5. Values of $\emptyset$ for different subset of consumers

S^n	0	0.4	0.4	0.6	0.8	Ground truth
Ratio of participating consumers (%)	100	40	18	9	4	0
$\emptyset(1e^{-2})$	26.3	6.4	9.8	20.6	23.4	30.0

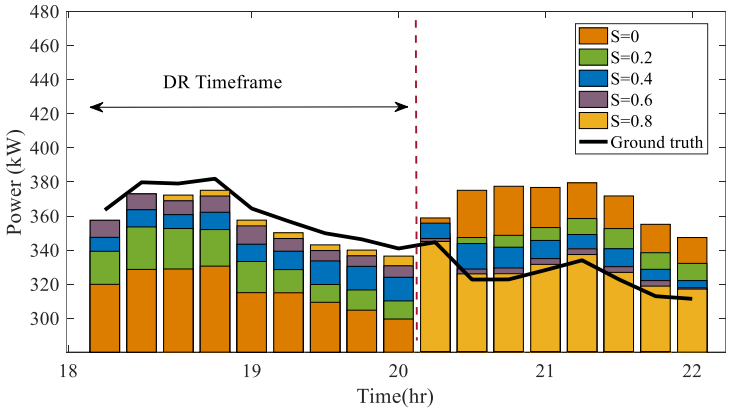


Figure 2-15. Comparison of the aggregate power demand of the community based on ground truth and different subsets of consumers for DR and subsequent off-peak timeframe

2.5. Implications and limitations

2.5.1. Large-scale potential for demand reduction

The findings in this study could provide insight for planners and utilities to identify and target potential candidates for automated DR. To this end, we have estimated the potential for Austin TX, as a demonstration of city-scale assessment. According to the U.S. Census Bureau, the number of households in Austin was about 360,000 in 2017 [103]. Although the appliance ownership rate in Austin could be different from our sample, we used the same rate as a basis for a rough estimate of demand reduction at a city-scale. Therefore, by assuming the same appliance ownership rate (Table 2-2) and the information in Table *2-4*, it is estimated that participation of only 20% of users with higher demand reduction potential, during a 2-hour DR event, could result in up to 35.2MW, 8.2MW, 1.0MW, 2.0MW, and 91.1MW for EV, AC, dryer, washing machine, and dishwasher, respectively. This suggests a total of 137.5MW reduction for all loads combined.

2.5.2. User segmentation and fairness

An important consideration in DR programs is *fairness* to ensure that a specific set of users are not targeted on a regular basis and avoid the risk of user fatigue [104], which could result in reduced willingness for participation. In our segmentation scheme, given that DR events might be selected at different time intervals for different days as operatorial decisions dictate, the S^n values for users could vary across different days. Therefore, the same user could be assigned to a different subset of targeted users for each DR event, which in turn helps to alleviate the fatigue problem. To demonstrate how users will be assigned to different subsets (sorted based on their potential score), we have considered three DR time intervals of 16-18, 17-19, and 18-20. For each interval, based on the S^n values, each user will be assigned to a different quantile (i.e., a subset), which contains the set of users such that $\frac{i}{100} < S^n < \frac{i+10}{100}, i \in \{0,10, \dots ,90\}$. Therefore, users will be assigned to 10 different quantiles.

The results are presented in Figure *2-16*. Each polyline spanning the x-axis represents one user across different DR events. Darker lines are associated with a higher frequency of occurrence. A horizontal polyline indicates that the user is always assigned to the same quantile at different DR times, while each inclined line indicates the allocation of the user to a different quantile as DR timeframe varies. In Figure *2-16* considerable variations across the community are observed, implying that users are allocated to varying levels of priority for participation for different DR

timeframes. Heuristic-based methods, which iteratively update the record of DR participation for each user at each event, have been used as an alternative method for improved fairness [67].

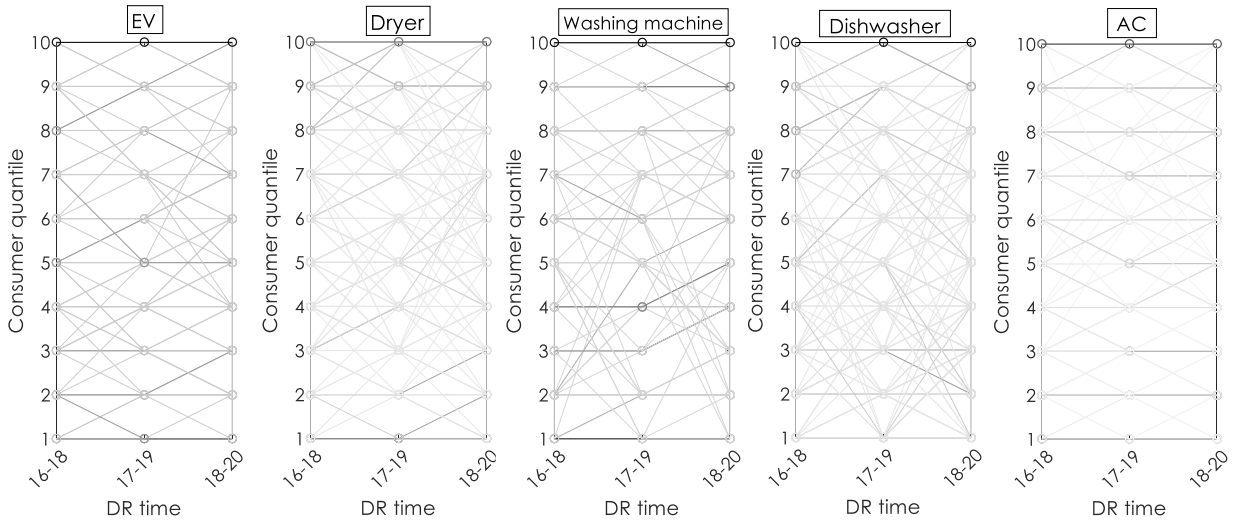


Figure 2-16. Association of different users to different quantiles based on varied DR timeframes. Each line represents the output for a user. Darker lines are associated with a higher frequency of occurrence.

2.5.3. Data availability

In our proposed segmentation scheme, data at the individual load level (i.e., appliance-level) is used for analysis. In the context of smart grid, smart appliances will be capable of providing such information and respond to DR signals for load shifting. Furthermore, considering the appliances mentioned in this study, recent efforts based on disaggregation methodologies have achieved reasonable accuracy for load disaggregation from smart meter data. For example, disaggregation of EV loads [105, 106], dryer, washing machine, dishwasher [107], or AC [108], with relatively acceptable accuracy (~70-100%) have been reported in recent studies. Accordingly, the proposed segmentation scheme, coupled with disaggregation on smart meter data could be used for incentive programs or for an economic feasibility assessment of offering smart appliance rebates to suitable candidates under budget constraints.

We evaluated our method based on a case study analysis from the Pecan Street project [68], which is currently the largest available energy consumption dataset at both aggregate and appliance levels. More specifically, we used a sample of more than 300 households to ensure that the analysis covers diversity in energy consumption styles and interactional behaviors of users. However, this approach is considered a generalized approach as it includes two main steps: (1) quantifying the

differences between users by using the proposed scoring metric, and (2) providing quantified energy demand variations. As long as the datasets that enable examining the complex human-building interaction at the level of individual loads characteristics are available, these metrics could be quantified.

2.5.4. Limitations

There are a number of limitations associated with this work as outlined here: **(1)** Daily routines and occupants' lifestyle is impacted by the day of the week (working days or weekends). Accordingly, consumption styles can be reflected differently on weekends compared to weekdays. For the computation of potential score in this work, we have not distinguished between such differences. Nonetheless, given larger duration and availability for the data, S^n values can be calculated separately on weekdays and weekends and separately evaluated for DR events. **(2)** We have looked at the flexibility based on potential demand reduction that can be realized by the participation of individual loads. However, the potential increase in power demand is also deemed as a flexibility attribute, which was not investigated in this work. **(3)** Looking at the results in Figure *2-14*, the increasing trend in demand reduction potential based on higher S^n values are not observed in all cases. This is due to the fact that the usage styles in the future (used as test events) are not always dependent on previously shown historical behaviors and uncertainties are involved. Nonetheless, the results in most cases appear to agree with our expectation, and regular or seasonal upgrade in calculating the scores could be effective. **(4)** We accounted for user compliance using previous community-level empirical studies. However, the average compliance factor for each load type might not be the best representation of the community, and varying levels of willingness in complying with the DR requests for each household need more sophisticated model compared to the implemented distribution function in this work. For example, it is possible that users with frequent and consistent usage pattern will be less willing to shift their loads. Therefore, the causality of behavioral patterns is an important factor to consider. Investigation of the association between user compliance and historically observed consumption behavior is among our future research directions. **(5)** We used the same timeframe for all the households with respect to the same benchmark calendar days for the evaluation. In our sample in this study, there were cases with missing data (e.g., hours of missing data for a day) in which we opted for eliminating the specific day from our analysis. In practice, missing data points, such as every 15-minute data, could be interpolated with adjacent points as long as such data points are scarce and not continuous.

However, continuous metering failures that limit the measured data could hinder the evaluation process for a specific customer. An interesting research direction is to develop a data-driven technique that considers a certain duration of data and measures when the behavior of a customer has changed. As a result, if the behavior has not changed over the considered duration, a practical solution could be referring to existing alternate days (e.g., same day of the week) in the historical dataset for the same customer as a replacement alternative. **(6)** In specific cases such as the presence of outmoded or anomalous appliances, higher electricity consumption could be incurred. However, such improper functionality does not necessarily imply increased/proper engagement of users for DR program. Nonetheless, a two-tier scenario can be considered: (I)- The customers can be suitable for energy-efficiency programs such as being targeted for appliance replacement and not necessarily DR programs that look at imposing temporal changes in the dynamic patterns of consumption. Recent data-driven methods for identifying outmoded or anomalous appliances for historical data includes reference [109] that looked into identifying potential households with outmoded and inefficient appliances such as HVAC or reference [110] that identifying anomalous appliances through examining their load signatures. Relying on such existing methodologies from the literature, it is possible to filter out potential customers with such specific cases in the data pre-processing step and filter them in the evaluation step. (II)- The findings can be used for incentivizing right customers for automated technology adoption. In the case of having outmoded appliances, the operation time could still be shifted. In other words, shifting the operation time does not depend on the technology in the appliance. Therefore, although users with outmoded appliances cannot be integrated into DR programs such as Direct Load Control (DLC), they can be targeted for adopting smart appliances.

2.6. Conclusion

In this study, we systematically investigated the demand reduction potential for different individual residential loads according to historical consumption styles. A data-driven scoring approach was used to characterize and rank the users based on the patterns of their interaction with different loads. The empirical assessment in the context of DR shows the applicability of the score for user segmentation for automated DR. The investigations were conducted for a variety of deferrable loads including EV, dryer, washing machine, dishwasher, as well as AC. In a comparative analysis of flexibility potential, we used two scenarios: (1) maximum potential of load shifting/shedding

and (2) user compliance modeling. The findings show that households manifest high variations in providing demand reduction potential and the proposed scoring approach could capture those variations. Furthermore, the scoring approach was shown to be effective in the targeted engagement of the users for effective demand reduction without a rebound effect. EV and AC were shown to provide a higher level of flexibility compared to wet appliances. A demand reduction projection for households in Austin, TX, proposes that justified selection of considered loads for DR during the afternoon timeframe could potentially provide around 140 MW power reduction, with only 20% participation of residential users in the community.

Chapter 3: A Machine Learning Framework to Infer Time-of-Use of Flexible Loads: Resident Behavior Learning for Demand Response

Afzalan, Milad, and Farrokh Jazizadeh. "A Machine Learning Framework to Infer Time-of-Use of Flexible Loads: Resident Behavior Learning for Demand Response" *IEEE Access*, 2020, DOI: https://doi.org/10.1109/ACCESS.2020.3002155

Abstract

Load shapes obtained from smart meter data are commonly leveraged to understand daily energy use patterns for adaptive operations in applications such as demand response. However, they do not provide information on the underlying causes of specific energy use patterns – i.e., inference on appliance time-of-use (ToU) as actionable information. To this end, we investigated a scalable machine learning framework to infer appliance ToU from energy load shapes in a collection of residential buildings. A scalable and generalized inference model obviates the need for model training in a given building to facilitate its adoption by only relying on available training data from previously observed buildings. We have investigated the feasibility of using load shape segmentation to boost ToU inference in buildings by learning from their nearest matches that share similar energy use patterns. To infer the appliance ToU for a building, classification methods are trained on subintervals of load shapes from matched buildings with known ToU. The framework was evaluated using real-world energy data from Pecan Street Dataport. The results for a case study on electric vehicles (EV) and dryers show promising performance by using 15-min smart meter load shape data with 83% and 71% F-score, respectively, and without in-situ training.

3.1. Introduction

In recent years, conventional centralized power systems are shifting to decentralized alternatives that integrate distributed energy resources (DER) such as solar panels, district resources, storage systems, and advanced technologies for smart metering and control. These changes provide opportunities and call for efficient adaptive and responsive operations – e.g., utilization of Demand Response (DR) programs for load balancing or energy exchange at neighborhood level. However, the successful implementation of adaptive operations in the residential sector requires a sound understanding of energy usage patterns such as load shapes – i.e., the variation of power demand over the span of a day. Detailed analysis of the usage patterns reveals temporal drivers of demand, which in turn enables efficient targeting of customers for customized energy programs and demand

control automation [37, 38, 93]. Building energy load shapes are commonly characterized through data-driven segmentation methodologies by clustering smart meter data to help engage consumers for DR programs [13, 87] by providing information on peak time and off-peak time demands. However, understanding the drivers of demand variations through an in-depth analysis of the human-building interactions (HBI) at the appliance (i.e., individual load) level will improve the efficacy of managing loads for demand-supply balance as shown in previous research [37, 111].

HBI assessments center on identifying the patterns of using different flexible loads (e.g., electric vehicles) in a household. These assessments provide a quantitative and statistical measure to evaluate the benefits for engaging end-users in adaptive management of loads (e.g., engagement in DR programs). By measuring time-of-use (ToU) of flexible loads and driving other metrics such as frequency and consistency of use, operations could be moved toward intelligent and distributed load operation/scheduling with reduced burden for end users [47]. In other words, information from HBI patterns, as supplementary information to the load shapes at the aggregate level, helps manage power systems more efficiently by engaging a sub-set of consumers that will result in higher gain in efficient operations. Leveraging smart meter data, machine learning (ML) tools have been used to characterize different energy usage patterns in residential buildings and neighborhoods and to infer variations in energy lifestyles (e.g., [16, 112, 113]). However, leveraging building energy load shapes has not been used for inference of the appliance ToU.

Determining the daily ToU of appliances can be carried out by using individual plug sensors at each device. Such an approach provides accurate measurements but calls for distributed installation of sensors, which is intrusive and may be prohibitively complicated and expensive. On the other hand, appliances' ToU can be implicitly inferred through inference models using pattern recognition algorithms from the power data at the aggregate (i.e., whole building) level. However, an inference model suited to a specific building calls for *a priori* assumption about appliance characteristics, high-resolution data, or in-situ training, which might be an obstacle for a scalable approach. Therefore, enabling a generalized and data-driven inference approach that does not rely on specialized instrumentation and in-situ algorithm training, or specific assumptions on an individual building, could facilitate scalable adaptive operations such as DR or DER management. Research has shown that the dynamic patterns of energy consumption, driven by residents' interactional behavior, are correlated with appliance ToU [114, 115]. Accordingly, in this study, we have investigated ToU inference models that leverage neighbor (i.e., previously observed)

buildings with similar energy use behavior for training by leveraging a behavioral learning framework. To this end, we have introduced a machine learning framework for inference of appliance ToU in buildings by accounting for resident behavior reflected in their energy load shapes from smart meter data. In other words, we have sought to:

- Propose a scalable approach for ToU inference of flexible loads by relying on only low-resolution smart meter data – i.e., 15-min resolution, which is the default sampling rate in practice.
- Enable a data-driven learning framework for ToU inference to identify the similarities between households by leveraging the aggregate energy load shapes. In contrast to methods that require specific instrumentation at the building level or *a priori* information on appliance models, we solely leverage the information in energy load shapes of a sample of neighbor buildings whose appliance ToU is available for training.

Therefore, in contrast to in-situ training for energy disaggregation, which requires user engagement or specialized sensors in each building, we have investigated the feasibility of a limited sample of known buildings with similar behavior, whose appliance ToU data is already known – i.e., a generalized training data set. In this way, the information granularity in energy load shape analysis could be increased through a scalable approach.

The rest of this paper is structured as follows. In section 2.2, the related literature and research background, including the research on energy disaggregation, have been presented. In Section 3.3, the proposed inference framework has been presented. In Section 2.4, the framework evaluation through a case study, as well as the results and discussions are presented. Section 3.5 includes the concluding remarks.

3.2. Related works

ToU inference in buildings is becoming increasingly important for mitigating the stress on the grid and supporting distributed energy management. To this end, several studies have explored the underlying causes of variation in appliance usage profiles at different times of the day from the user interaction perspective [37, 116, 117]. The study by Cetin et al. [116] investigates the energy use of appliance profiles including dryers, washing machines, and refrigerators in 40 buildings to understand the variation at different times of the day. Statistical results showed electricity use varies more amongst buildings during the peak time of the day compared to off-peak time. The

variations in appliance ToU amongst buildings have been studied to estimate cost saving under dynamic pricing scheme [117] and to quantify energy saving and load flexibility for different segments of users [37]. The results have shown highly variable energy demands and high load flexibility potential at different times of a day. By leveraging such information, it has been shown that the associating appliance ToU with load shape analysis and segmentation in the residential sector could lead to improved implementation of DR programs [71, 118].

To infer appliance ToU, studies have proposed different predictive models by learning form the variations on aggregate power time series as their data source [119-122]. Basu et al. [119] investigated the applicability of several classifiers, such as decision trees and Bayes networks, for ToU inference by using historical consumption information and human expert knowledge as the basis for in-situ training. Various studies have looked into inference of appliance usage by employing different deep learning architectures including long short-term memory (LSTM) [121], denoising autoencoder [123], and convolutional neural network (CNN) [124]. Barsim and Yang [122] developed a convolutional neural network architecture to extract individual profiles of appliances and evaluated it on five residential buildings. Kelly and Knottenbelt [121] investigated the applicability of three deep learning architectures including LSTM and denoising autoencoders and investigated their performance for detecting the use of major appliances including washing machine and dishwasher for five buildings. Lie et al. in [108] used the concept of transfer learning and leveraged a pre-trained deep learning model on an image recognition dataset for appliance classification. As a common feature, the applicability of such methods have been investigated on datasets with high-resolution data (e.g., with sampling rates of 1 Hz or higher – example relevant datasets with high-resolution data could be found in [125]), or alternative metrics such as voltage-current trajectory have been employed instead of real power measurement, which is commonly recorded by smart meters.

As an alternative to direct appliance ToU inference, classical energy breakdown (disaggregation) methodologies could be employed to extract the ToU of appliances. Through aggregate load event identifcation or reconstructing the time-series of individual appliances from energy breakdown and comparing thresholds with the values on the resultant time-series, ToU at different time of the day can be extracted. Energy breakdown methodologies have been widely studied – references [126-128] are well-known representative surveys in this field. In these studies, a wide range of parameters, important in driving the outcome, have been used across studies: (1) varying levels of

sampling rate in acquiring aggregate power data have been used, including high resolution data (@60Hz) [40, 129-131] and low-resolution (sampling every hour) data [132], (2) varying number of buildings for evaluation, including individual buildings [133], a few buildings (3 to 6) [41, 123], or 10 and more buildings [134, 135], and (3) different methods of training, including supervised methods that may require extensive parameter search [136], or supervised methods, capable of automated parameter tuning [40, 41, 130], and unsupervised methods using Hidden-Markov models [137, 138]. The scalability potential of these solutions varies from one to another [27] given their specific design and implementation. A scalable solution for ToU inference ideally accounts for the following factors: it avoids in-situ training – learning from the same environment that is the subject of prediction, and uses low-resolution (real-power) data that can be acquired with ubiquitous metering infrastructure such as smart meters and their default configurations.

Since this study has been motivated by DER integration and automated DR, we looked at the class of loads that are controllable and energy-intensive. Recent efforts have reported reasonable accuracies for energy breakdown methods. For example, disaggregation of EV loads [105, 106], dryer, washing machine, dishwasher [107], or AC [108] has reached accuracies in the range of ~70-95% in several efforts. However, the aforementioned studies have focused either on high resolution data (1-minute or higher sampling rates), required considerable model parameter selection, and reported the performance on small samples (i.e., a few buildings or a short duration of 1 day to a few days).

A number of studies focused on developing scalable solutions, for which learning the data representation from a sample of labeled data is carried out to identify the breakdown energy for another set of buildings. Using monthly aggregate data, Batra et al. [135] estimated the monthly energy breakdown of appliances for around 50 buildings through matching one building with similar ones in which the energy breakdown data is available. They used different features such as monthly energy usage, statistical attributes such as skew and kurtosis, and external attributes such as temperature and building size for classification with K nearest neighbors. The results show an improved performance over a benchmark Hidden Factorial Markov (HMM) model for a variety of appliances including HVAC, washing machine, dryer, with accuracies ranging between 40-75%. The authors later developed a Matrix Factorization approach to develop a scalable solution for energy breakdown, and they evaluated the performance of the method over 500 houses [139]. However, the scalable solutions in [135, 139] only estimated the monthly usage (i.e., one

consumption estimation per month), without providing insight into the variation of daily appliance ToU, which is essential for dynamic load management applications.

A summary of example related works discussed in this section is presented in Table 3-1. Summary characteristics of example related research efforts. Typically, these studies rely on higher-resolution data and in-situ training as common features, which hinder their adoption as a scalable approach for ToU inference. In contrast, in this research, we have investigated the feasibility of an ML framework for inferring ToU of major flexible loads from smart meter data with 15-minute resolution that accounts for the following features: (1) integrating a residential behavior learning component that leverages energy load shapes for identification of a training dataset to increase the efficiency of the machine learning training process, (2) inferring ToU for unseen buildings which have not contributed to the training process, and (3) investigation on appliance ToU is performed over hourly basis with application to dynamic load management (e.g., demand-response).

Table 3-1. Summary characteristics of example related research efforts.

Resolution	Method	Performance	Major appliances	Ref
6 seconds (real power)	Deep learning	F-score of 49%-72%	Washing machine, dishwasher	[121]
1 Hz (real power)		F-score of 41%-80%	Washing machine, dishwasher	[122]
1 Hz (real power)		F-score of ~80-90%	Dryer, dishwasher	[123]
1 Hz (real power)	Hidden Markov Model (HMM)	F-score of 84%	Wash/Dryer	[137]
1 minute	HMM	Energy accuracy of 52%-75%	Dryer, dishwasher	[138]
30 kHz (current-voltage)	Transfer learning	F-score of 87%-100%	AC, washing machine	[108]
1 min	Heuristic algorithm	Normalized mean square error of 19%	EV	[106]
1 min – 5 min	Independent component analysis (ICA)	F-score of 68-94%	EV	[105]
1 hour	Sparse coding	Accuracy of 55%	Dryer, dishwasher	[132]

3.3. Appliance ToU Inference Framework

We propose this framework to investigate a scalable approach for inferring appliance ToU from aggregate load shapes in buildings (called *target buildings*) with smart meter data using default resolution of 15-minute intervals. Target buildings are environments that are new to the framework

and do not contribute in model training. The premise of this framework centers on learning from buildings (called *training sample buildings*) with similar energy behavior patterns, reflected in the characteristics of their aggregate daily load shapes. The training sample buildings are instrumented with plug metering devices at individual load level in addition to commonly accessible smart meters. The training sample buildings exclude the target building and could be identified as the best matches (in terms of energy behavior patterns) with target building. The framework is as illustrated in Figure *3-1*. We have elaborated on the components of this framework in the following sections.

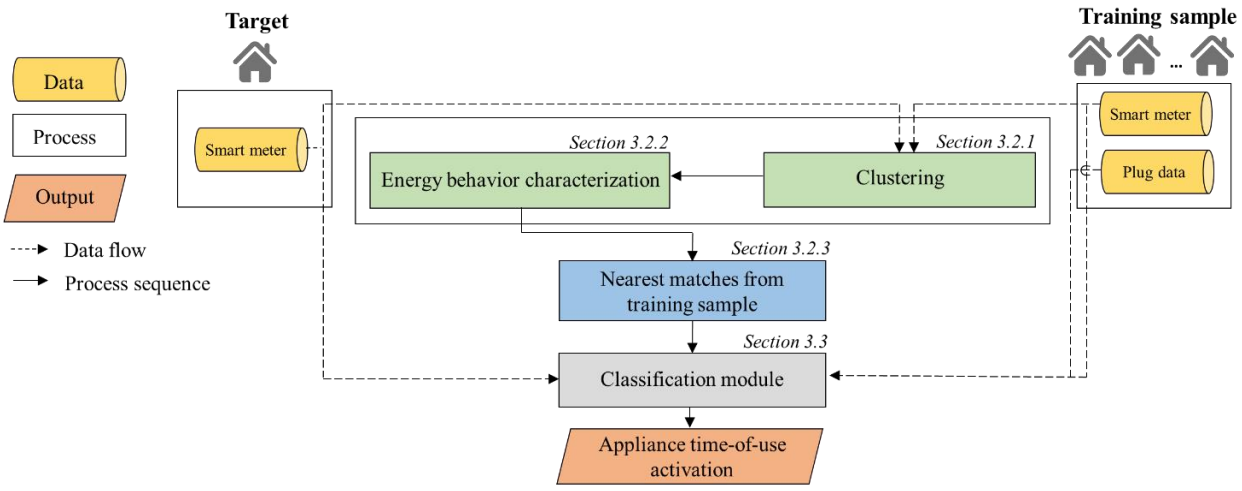


Figure 3-1. Schematic framework for appliance ToU inference.

3.3.1. Data requirements and processing

Figure *3-2* shows the required information flow for the framework. The data from the *training sample buildings* include (1) power time series at 15-min intervals (resembling the sampling rate by smart meters [140]), (2) appliance ownership information, (3) and power time series from plug meters (i.e., sub-metering at appliance level) with the same resolution as the aggregate-level power data. Considering the data collection campaigns and public release of large-scale energy data sets from residential buildings in the past recent years, data sets with such characteristics are available for hundreds of buildings [141]. On the other hand, the data sets in *target buildings*, in which the framework makes the inference, include (1) power time series from smart meters at 15-min intervals and (2) appliance ownership information. Accordingly, the process of identifying training sample for a given target with similar behavior (which we refer to as the matching process) calls for aggregate power time series and appliance ownership information, which are either available or could be acquired. Smart meters have been deployed in half of the United States and considered

as a scalable platform for data collection and analysis, and appliance ownership data can be obtained through a one-time inquiry from home owners or could be automated.

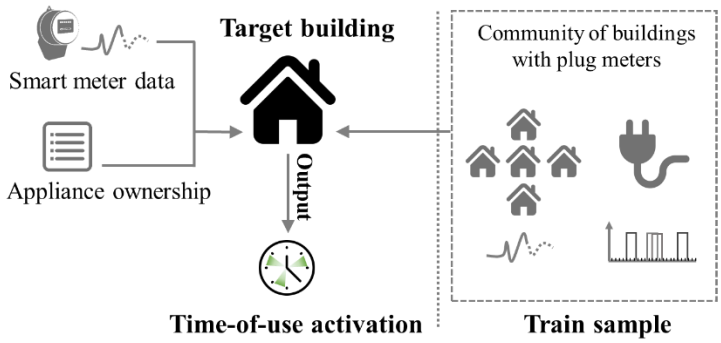


Figure 3-2. Data requirements and information flow for the proposed framework.

3.3.2. Characterization of energy behavior patterns for training sample

Identifying the training sample buildings for a given target building relies on the energy behavior patterns of these buildings. Therefore, the framework characterizes the energy behavior patterns by leveraging aggregate daily load shapes in order to find potential candidate buildings that are similar to the target. In this process, characterization is defined as understanding the variations of daily energy use patterns, which could be achieved by segmentation of daily load shapes through clustering techniques.

3.3.2.a. Segmentation of daily load shapes (clustering)

Let $P_{in}(t) = \{p_1, p_2, p_3, \dots, p_T\}$ be the daily load shapes in the form of power time series collected through a smart meter. Here, T is the total number of samples per day, $i \in \{1,2, \dots, I\}$ is the building index (I is the total number of buildings), and $n \in \{1,2, \dots, N\}$ is the index for historical days (N is the number of historical days). Therefore, considering M as the total number of load shapes ($M = I \times N$), the load shape library could be clustered into K clusters.

Different clustering techniques can be employed for the segmentation of energy load shapes, including K-means, hierarchical clustering, and customized methods for power time-series segmentation. In this framework, we have proposed a two-stage method based on Self-Organizing Map (SOM) clustering for creation of initial clusters, followed by extracting temporal features from clusters for merging similar ones. In the first step, SOM is applied to all the load shapes (M) to obtain K' clusters, in which $K' \gg K$. In the second step, a number of statistical features are extracted from clusters that provide quantitative metrics for describing the temporal shape or magnitude of power demand (such as the timing of peak demand) to collectively characterize a

cluster. Specifically, the feature set describes the energy consumption level (total consumption level), and the distribution pattern (number of relative peaks in each cluster), the timing of peak occurrence, and the intensity of the peak demand (magnitude of peak demand compared to adjacent points).

Figure *3-3* illustrates the extracted features from a cluster that characterize the information contained in one cluster. Using this approach, clusters with similar set of features are merged to further reduce the size of the cluster library. In this way, various load shapes with distinct temporal shapes are extracted, while each one presents a unique shape. More details about the proposed clustering approach could be found in [142]. Upon creating the clusters, each load shape (P_{in}) will be associated with a cluster index $k \in \{1, \dots, K\}$.

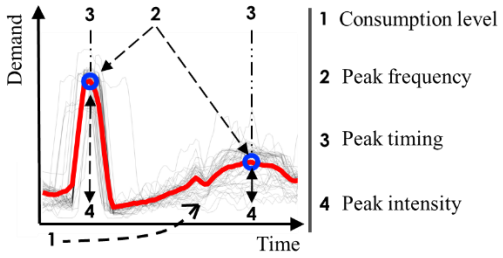


Figure 3-3. Example of a clustered load shape and its extracted features. Detail about feature extraction can be found in [142].

3.3.2.b. Energy behavior characterization

The outcome of the segmentation for a given building includes a number of representative clusters that summarizes different patterns of energy use across a historical period. To characterize the energy behavior of that building, one could consider the dominant cluster that includes majority of observations over the historical days. However, residential buildings show high uncertainty in energy consumption patterns [114], and they are typically expected to be represented by different clusters and change their pattern across different days. Therefore, it is more realistic to associate the behavior of each building with multiple clusters that are commonly observed. To this end, upon segmentation, each building will be characterized based on the frequency of clusters observed over historical data.

Considering the frequency of different clusters for each building, the energy behavior of each building *i* can be characterized in a feature vector π_i:

$$\pi_i = \{Pr_{i1}, Pr_{i2}, \dots, Pr_{iK}\} \quad (1)$$

in which k is the index of the clusters and $Pr_{ik}, k \in [1:K]$ is the probability (frequency of observing a cluster) of observing C_{ik} across historical days for a building. C_{ik} is the k^{th} cluster in the load profile data set of building i. Pr_{ik} is defined as:

$$Pr_{ik} = \frac{\| C_{ik} \|}{N} \quad (2)$$

In which $\| C_{ik} \|$ is the number of daily load shapes for building i in cluster k. The π_i represents the energy behavior of buildings according to their daily load shapes and the frequency of observations.

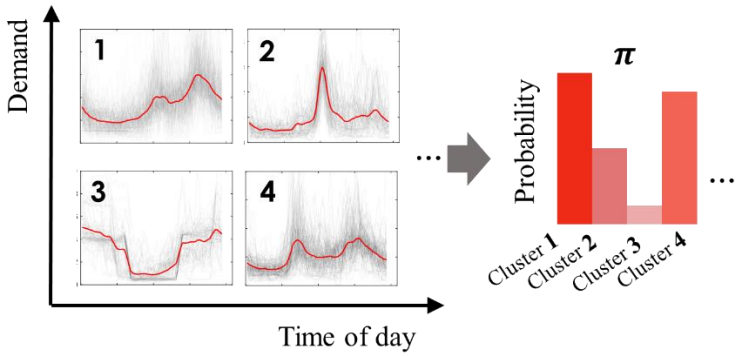


Figure 3-4. Representation of π according to the frequency of clusters.

Figure *3-4* illustrates an example of clusters and the associated feature vector (π) that is derived according the frequency of clusters over a period of historical data. It is noted that clustering is performed on the entire sample (both the training and target buildings), and π_i is characterized on all the buildings.

3.3.2.c. Identifying Nearest Matches to Target as the Training Sample

To identify the training sample buildings for ToU inference in the target building, we adopted the use of KNN algorithm for finding the nearest training sample buildings that match the energy behavior of the target. Considering the training sample size of $I' \ll I$, the nearest matches in a community with respect to the target are the $[1, \dots, I']$. Here, π_i vector is used as the feature vector for KNN algorithm with Euclidean distance as the similarity measure. In this work, we primarily investigated 10 as the number of neighbors for the KNN, while comparing it against other values in the result section.

3.3.3. Appliance time-of-use inference

Given the power time series of the daily load shape for target building i on day n, $P_{in}(t)$, the objective is to identify the use of a flexible appliance across a collection of time bins during the day – $\Omega_{in} = \{\omega_1, \omega_2, \omega_3, .., \omega_\Gamma\}$ in which Γ is the total number of time bins ($\tau \in [1: \Gamma]$). A time bin is a window of multiple data points with the length l, which represents a timeframe that could be selected for DER management or a DR event. The elements in vector Ω_{in} has the following binary form:

$$\omega_\tau = \begin{cases} 1 & if\ appliance\ was\ being\ used \\ 0 & if\ applaince\ was\ not\ used \end{cases} \quad (3)$$

Therefore, to infer a deferrable load status we have:

$$\gamma_{in\tau} = \begin{cases} 1 & if\ P\left(\omega_\tau = 1 \middle| P_{in\tau}(t)\right) > P\left(\omega_\tau = 0 \middle| P_{in\tau}(t)\right) \\ 0 & else \end{cases} \quad (4)$$

in which $\gamma_{in\tau}$ is the appliance ToU prediction for building i, on day n, and at the time bin ω_τ. $P_{in\tau}(t)$ denotes the $P_{in}(t)$ power time-series associated with the time frame τ, in which $t \in [\tau l - l + 1: \tau l]$

. $\gamma_{in\tau}$ is used as the appliance ToU predictor for presenting the results in this paper.

In order to prepare the ground-truth in the form of Ω_{in} (binary series across different time bins per day) for the training sample buildings, the continuous power time-series need to be converted into the binary form at each time bin. Considering $P^a(t) = \{p^a{}_1, p^a{}_2, p^a{}_3, \dots, p^a{}_T\}$ as the appliance power time-series ground truth, we employed the following equation for preparing the ground-truth Ω_{in} in our training sample:

$$\omega_\tau = \begin{cases} 1 & if\ \max\left(P^a(t)|_{t=l(\tau-l)+1:\tau l+1}\right) > P^n \\ 0 & else \end{cases} \quad (5)$$

in which P^n is the nominal power draw threshold of a load while being on.

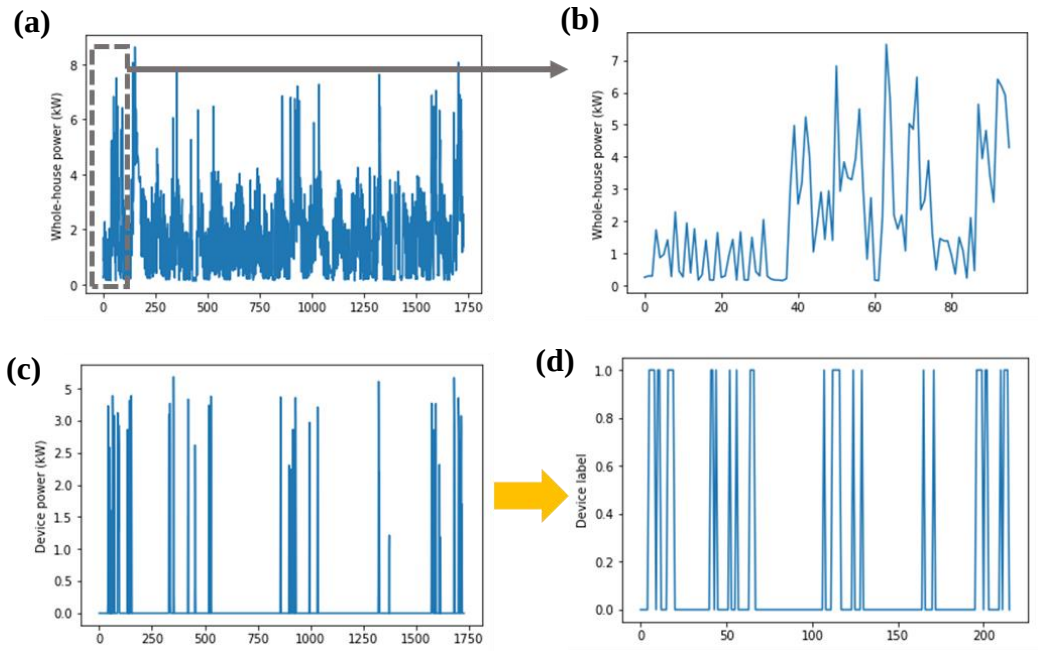


Figure 3-5. Example of daily load shapes over twenty days for one building and its corresponding dryer power time series and label for each time bin: (a) power consumption collected by using a smart meter at 15-minute interval, (b) magnified view of the first day from part (a), (c) power consumption of a dryer for the same time span in (a), and (d) extracted labels of dryer from part (c) at two-hour intervals.

3.3.3.a. Classifier component

To classify the appliance ToU at each time bin in the target building, we investigated the application of several classifiers. The processed historical data of daily load profiles $P(t)$ and Ω from the training sample buildings are used for training, and $P(t)$ from the target building is used for inferring the appliance ToU. Therefore, the classifier infers Ω from $P(t)$ by relying solely on the information from training sample with known ToU $[f: P(t) \rightarrow \Omega]$. Using the nearest matches from the training sample building (described in section 3.2), predictions for each appliance will be carried out separately. To this end, we investigated Dense Neural Network (Dense NN), KNN, Support Vector Machine (SVM), and Random Forest (RF) as four classification algorithms.

Considering that building load shapes are presented in relatively low dimensions and the dataset that is moderate in size (compared to common vision and text datasets), we opted for a two dense-layer NN. The architecture consists of 300 hidden units in the first layer with ReLU activation function, a sigmoid activation function in the second layer to map two status of 'On' and 'Off', and RMSprop as the optimizer. The second classifier, KNN, uses the class output of k nearest neighbors based on Euclidean distance for classification. 5 nearest neighbors for load shapes was used in the analysis upon empirical observation. The third classifier, SVM, defines decision

boundaries based on hyperplanes to separate the feature space into classes. The fourth classifier, RF, is an ensemble learning method that leverages multiple decision trees and select the mode of the classes of individual tress as the output.

3.3.3.b. Oversampling consideration for training

The appliance ToU data sets are imbalanced by nature. Intuitively, for most buildings, it is expected that "Off" class is dominant compared to the "On" class. However, for our problem, the minority "On" class has higher importance. Therefore, it is desired to reduce the number of false negatives (FN) compared to false positives (FP). Imbalanced datasets can adversely affect the performance of classifiers and lead to bias in favor of the majority class [143]. Solutions to mitigate this problem include oversampling or creating synthetic data for the minority class [144]. Here, we have used the Synthetic Minority Over-sampling Technique (SMOTE) for mitigating the imbalanced problem. SMOTE [145] creates artificial instances of the minority class that are close to the existing ones in the dataset. Synthetic samples are created by calculating the difference between observations from the minority class and their nearest neighbors (Δ). These differences are then multiplied by a random coefficient between 0 and 1 and added along each feature to create a synthetic observation. Using this approach, the extra synthetic data for 'On' examples is only added to the training sample of buildings to ensure it does not affect the dataset for the target building.

3.4. Results and Discussion

3.4.1. Dataset and performance metrics

We used a sample of 467 buildings in Pecan Street Dataport [141], primarily located in Austin, TX, in July and August as the case study for segmentation and creating clustered load shapes. After pre-processing and performing moving average filtering on 15-minute data to reduce noise, 26870 daily load shapes were obtained. Moving average with a 1-hour window was considered to reduce the impact of noise in energy load shapes.

To evaluate the framework, we considered EV and dryer as two instances of flexible loads that are suitable for DER management and DR applications. For each load, we considered 10 buildings as the target buildings (test set) that owned the same appliance (using ownership data) with 20 days of data for evaluating the performance to provide adequate instances for evaluation. For each target building, 10 nearest matches were selected as the training sample with 60 days of data. The

aggregate data and appliance data had a sampling rate at every 15 minutes. A value of $l = 8$, corresponding to time bins with a duration of $\tau = 2$ hours was selected to represent the common timeframe for running DR events [146].

For training the classifiers, we used two scenarios of training for the target buildings. In the first scenario, labeled as RBL (resident learning), we used the data from 10 nearest matches, selected by using the method discussed in Section 3.2. In the second scenario, we used the data from 30 buildings randomly sampled in the community without considering nearest matches. We deliberately selected a higher number of buildings for the second scenario to give more power by feeding more data for training and compare it with the alternative approach of using similar matches for training. Nonetheless, a sensitivity analysis for the training samples size by evaluating different combinations of 10, 20, or 30 buildings is presented later in Section 3.4.3. Standard metrics including precision, recall, and F-score were used for evaluation considering our imbalanced datasets and need for assessing the trade-off among TP, FP, and FN.

3.4.2. Building energy characterization

The two-stage clustering technique described in section 3.2.1 was employed on the smart meter data. An initial 60 clusters were generated on the entire dataset. Clusters were visually examined to ensure that they are compact and their centroids are representative of their associated members. In the second stage, features from clusters were extracted and those with similar features were merged. This was done to eliminate highly correlated clusters and to maintain those that reflect distinct ToU on load shapes. This process further reduced the cluster library size to 39 clusters, and they were considered as the final clusters for the rest of analysis. Figure *3-6* shows samples of clusters in the library and their frequency of occurrence. After this process, each daily load shape in the library (for buildings in both groups of training samples and target) was annotated with its cluster index and building ID.

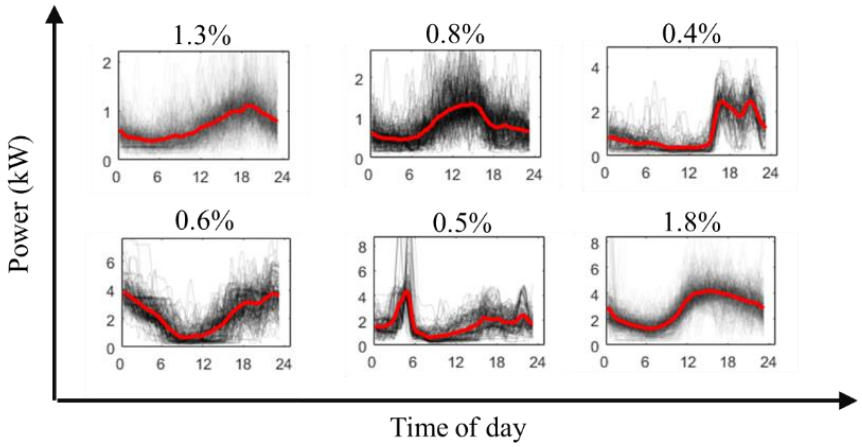
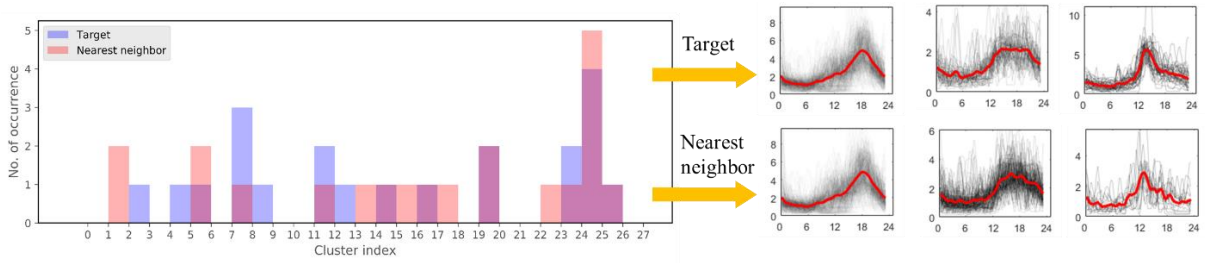


Figure 3-6. Examples of clustered energy load shapes and their frequency of occurrence in the community.

Using the results in the previous section, the energy behavior of the building, π_i (described in Section 3.2.2), was calculated. The nearest ten matches for each target were identified using KNN. Figure *3-7* demonstrates an example histogram of clusters in one of the target buildings and its nearest neighbor. Figure *3-7*(a) compares the histogram of clusters, and Figure *3-7*(b) shows the top three clusters with higher occurrence for two buildings. As shown, there is a high similarity in terms of energy use behavior across historical days for these two buildings.



(a) (b)

Figure 3-7. Household characterization and similarity search for one target building: (a) histograms of clusters for two buildings, and (b) top three daily load shape clusters.

3.4.3. Appliance time-of-use inference

Through using the statistical approach and different classifiers described in section 3.3, ToU inference of individual loads based on smart meter data was investigated. Figure *3-8* shows a visual demonstration of the EV charging prediction for one building over 20 days. The upper subplot is the smart meter data, collected at 15-minute intervals. The lower subplot compares the ground-truth and prediction of EV charging status at each associated time bin using Dense NN+RBL. As

can be seen, the detected charging instances are highly in agreement with the ground-truth, and most of the charging instances are correctly identified at its right time by load shape analysis.

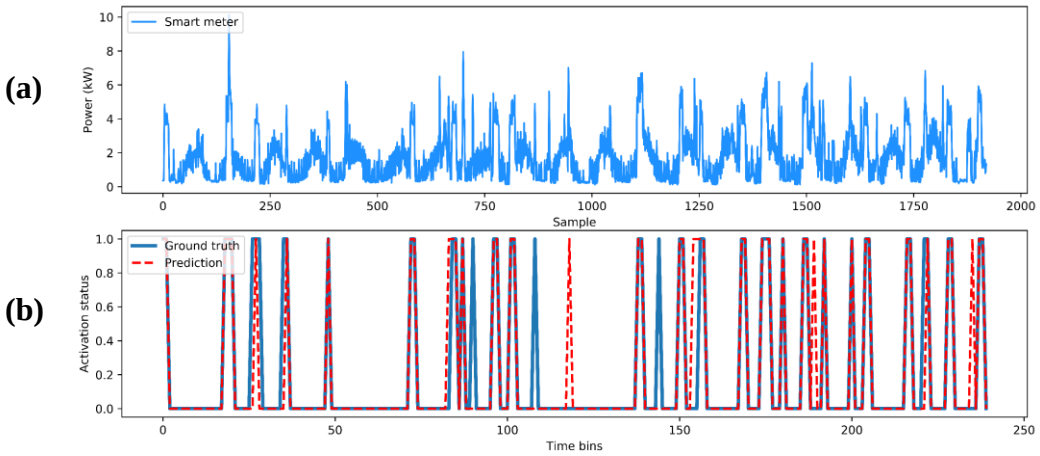
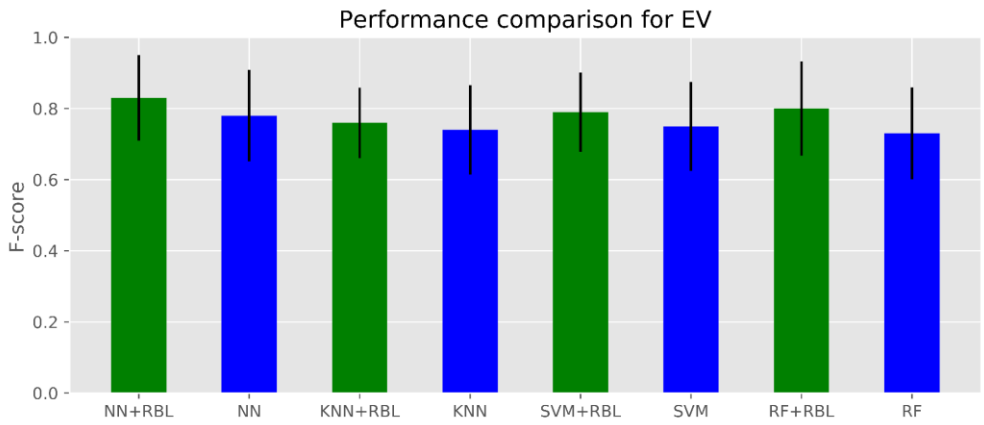
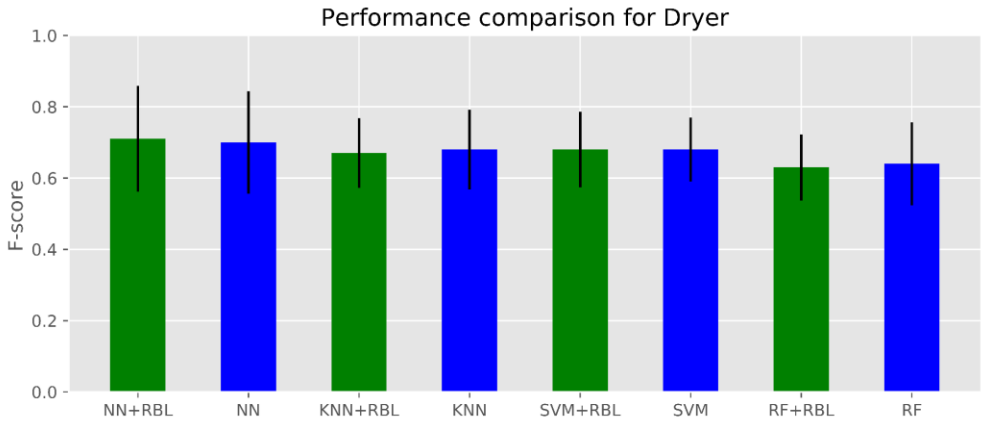


Figure 3-8. Visual demonstration for EV charging of one building for 20 days: (a) smart meter data, (b) ground-truth and prediction of EV charging.

To provide a comparison between different methods, *Figure 3-9* shows the F-score values, averaged over all target buildings. As noted, 'RBL' represents the 'resident learning' approach for selecting the training sample buildings. Based on the results shown in *Figure 3-9*, for both EV and dryer, the 'Dense NN+RBL' approach demonstrates the best performance, with an average F-score of 83% and 71% respectively. For EV, for all four classifiers, integrating the 'RBL' component improved the results (5% on average). For dryer, integrating the 'RBL' component improved the results for two classifiers, while the impact was negligible (1% improvement). It must be noted that three times more data was used for training based on random sampling compared to the RBL approach.



(a)



(b)

Figure 3-9. Performance comparison of different algorithms for (a) EV and (b) dryer.

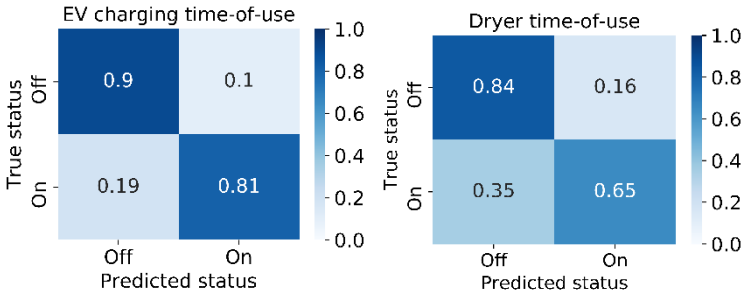
Table 3-2. Performance comparison of different algorithms for individual buildings (best results are shown in bold)

EV												
Method	**Dense NN+RBL**			**KNN+RBL**			**SVM+RBL**			**RF+RBL**		
Building ID#	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**
6072	**0.86**	**0.87**	0.86	0.82	0.84	0.80	0.83	0.84	0.83	**0.86**	0.85	**0.89**
9776	0.59	0.63	0.59	0.61	0.63	0.60	0.58	**0.64**	0.59	**0.62**	**0.64**	**0.61**
545	**0.91**	**0.88**	0.96	0.80	0.78	0.83	0.86	0.80	**0.97**	0.84	0.80	0.91
3036	**0.89**	**0.93**	0.86	0.79	0.87	0.75	0.85	0.90	0.81	0.87	0.91	0.84
1169	**0.91**	**0.90**	0.94	0.78	0.86	0.73	0.89	0.87	0.91	0.84	0.84	0.85
5749	**0.72**	**0.93**	**0.67**	0.67	0.86	0.63	0.65	0.80	0.62	0.67	0.84	0.63
9935	**0.66**	**0.78**	**0.68**	0.61	0.72	0.64	0.59	0.69	0.62	0.64	0.75	0.59
1629	**0.92**	0.90	**0.94**	0.87	0.88	0.86	0.88	0.86	0.90	0.89	**0.91**	0.89
4352	**0.96**	**0.97**	0.95	0.91	0.92	0.90	0.95	**0.97**	0.93	**0.96**	**0.97**	0.95
6990	0.84	0.87	0.82	0.78	0.83	0.75	0.78	0.81	0.77	0.84	**0.88**	0.81
Average	**0.83**	**0.87**	**0.83**	0.76	0.82	0.75	0.79	0.82	0.79	0.80	0.84	0.80
	Dense NN			**KNN**			**SVM**			**RF**		
6072	0.76	0.79	0.73	0.76	0.82	0.74	0.73	0.75	0.71	0.74	0.76	0.73
9776	0.58	0.61	0.58	0.54	0.61	0.57	0.55	0.63	0.58	0.54	0.60	0.54
545	0.87	0.83	0.93	0.87	0.84	0.90	0.83	0.78	0.92	0.82	0.76	0.96
3036	0.87	0.87	**0.88**	0.77	0.74	0.81	0.78	0.76	0.80	0.75	0.71	0.80
1169	**0.91**	**0.90**	0.94	0.86	0.84	0.88	0.88	0.89	0.87	0.89	0.85	**0.95**
5749	0.71	**0.93**	0.66	**0.72**	**0.93**	**0.67**	0.71	0.88	0.66	0.67	0.76	0.62
9935	0.52	0.68	0.63	0.50	0.66	0.61	0.49	0.57	0.54	0.50	0.63	0.46
1629	0.86	0.83	0.91	0.70	0.68	0.73	0.77	0.73	0.88	0.69	0.67	0.80
4352	0.84	0.78	0.94	0.89	0.87	0.91	0.93	0.90	**0.96**	0.91	0.88	0.95
6990	**0.85**	0.85	**0.85**	0.80	0.81	0.78	0.81	0.81	0.82	0.80	0.78	0.83
Average	0.78	0.81	0.80	0.74	0.78	0.76	0.75	0.77	0.78	0.73	0.74	0.76
Dryer												
	Dense NN+RBL			**KNN+RBL**			**SVM+RBL**			**RF+RBL**		
Building ID#	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**	**F-score**	**Recall**	**Precision**
6990	0.65	0.79	0.63	**0.70**	**0.81**	**0.67**	0.69	0.77	0.65	0.64	0.70	0.62
8282	**0.78**	**0.87**	**0.74**	0.65	0.69	0.64	0.72	0.77	0.70	0.70	0.74	0.67
3044	0.64	0.64	0.63	0.68	0.69	**0.68**	0.63	0.63	0.63	0.59	0.59	0.62
6139	0.71	0.70	**0.73**	0.66	0.67	0.65	**0.73**	0.75	0.71	0.64	0.63	**0.73**
7901	**0.91**	**0.86**	**0.98**	0.79	0.77	0.82	0.77	0.72	0.85	0.73	0.69	0.83
9356	**0.76**	**0.78**	**0.75**	0.64	0.65	0.63	0.70	0.72	0.69	0.59	0.61	0.65
4874	**0.85**	**0.83**	0.88	0.82	0.78	0.89	0.79	0.74	0.89	0.80	0.75	0.92
4505	0.74	0.75	0.73	0.67	0.69	0.66	0.75	0.74	0.76	0.62	0.62	0.63
5288	0.34	0.59	0.52	0.44	0.49	0.50	0.40	0.50	0.50	0.45	0.52	0.44
3367	0.71	0.74	0.70	0.68	0.69	0.68	0.67	0.66	0.68	0.55	0.56	0.61
Average	**0.71**	0.75	**0.73**	0.67	0.69	0.68	0.68	0.70	0.71	0.63	0.64	0.67
	Dense NN			**KNN**			**SVM**			**RF**		
6990	0.64	0.77	0.62	0.64	0.73	0.62	0.65	0.70	0.62	0.68	0.72	0.64
8282	0.69	0.76	0.67	0.65	0.69	0.64	0.62	0.70	0.62	0.59	0.60	0.60
3044	0.68	0.72	0.66	**0.69**	**0.75**	0.67	0.67	**0.75**	0.65	0.57	0.58	0.56
6139	0.68	0.71	0.67	0.63	0.65	0.62	0.71	**0.76**	0.68	0.57	0.57	0.63
7901	0.86	0.79	0.97	0.80	0.83	0.78	0.80	0.82	0.79	0.82	0.80	0.85
9356	0.68	0.76	0.67	0.61	0.69	0.62	0.60	0.67	0.61	0.51	0.55	0.47
4874	0.83	0.78	**0.95**	0.80	0.76	0.85	0.74	0.70	0.82	0.68	0.65	0.85
4505	**0.85**	**0.88**	**0.82**	0.79	0.83	0.77	0.76	0.83	0.73	0.83	0.87	0.80
5288	0.33	**0.64**	**0.54**	0.42	0.50	0.50	**0.48**	0.54	0.51	0.46	0.60	0.44
3367	0.74	**0.80**	0.72	**0.75**	0.79	0.73	**0.75**	0.79	**0.73**	0.66	0.66	0.67
Average	0.70	**0.76**	**0.73**	0.68	0.72	0.68	0.68	0.73	0.68	0.64	0.66	0.65

Table 3-2 provides the drill-down performance metrics for individual buildings. F-score, precision, and recall metrics were reported. For EV, out of 10 examples, 9 of them had improved F-score when classifiers were trained using the 'RBL' approach. For dryer, 6 out of 10 examples had improved F-score when the classifier integrates the 'RBL' approach for training. It was also

observed, that evaluation for EV achieved average Recall and Precision of 87% and 83%, respectively in the best case, while these values for dryer were 76% and 73%, respectively. In general, prediction on EV showed better performance compared to the dryer. Furthermore, the results in Table 3-2 shows that in most cases, the ToU can be predicted with high accuracy with Dense NN+RBL, while there were a few instances that turned out to be challenging in prediction (further discussed in Section 4.4).

Figure *3-10* shows the aggregate confusion matrix for all ten buildings over twenty days (200 days of observations) with two-hour time bins. For EV, there were a total of 2061 'Off' and 339 'On' instances, and for dryer, there were a total of 2070 'Off' and 330 'On' instances. As could be inferred from Figure *3-10*, for EV, out of 339 charging instances ('On' class), 273 of them were correctly identified. For the dryer, out of 330 'On' instance, 216 of them were correctly identified. Regarding the 'Off' instances, 1863 out of 2061 for EV, and 1740 out of 2070 instances for dryer were correctly classified. As the results show, the proposed approach has the potential to add a new layer of information on detecting the operational details for flexible loads by using daily load shape information from smart meter.



(a) (b)

Figure 3-10. Appliance ToU inference for (a) EV (including 2061 'Off' and 339 'On' observations) and (b) Dryer (including 2070 'Off' and 330 'On' observations).

To demonstrate the trade-off between true positive rate and true negative rate, Figure *3-11* shows the Receiver Operating Characteristic (ROC) curves across all buildings. For EV and dryer, the Area Under the Curve (AUC) show a high value of 0.93 and 0.86 (1 is perfect classification), corresponding to the probability of ranking a randomly chosen 'On' observation higher than a randomly chosen 'Off' observation. To adjust different levels of True Positive Rate (TPR) versus False Positive Rate (FPR), varying levels of thresholds for the classifier can be selected. For example, for EV, achieving TPR values of 0.7, 0.8, and 0.9 results in FPR values of 0.02, 0.10,

and 0.22, respectively. For dryer, the same TPR values of 0.7, 0.8, and 0.9 corresponds to the FPR values of 0.18, 0.25, and 0.36, respectively.

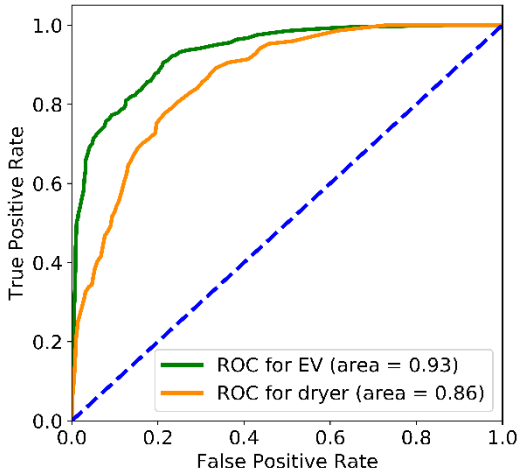
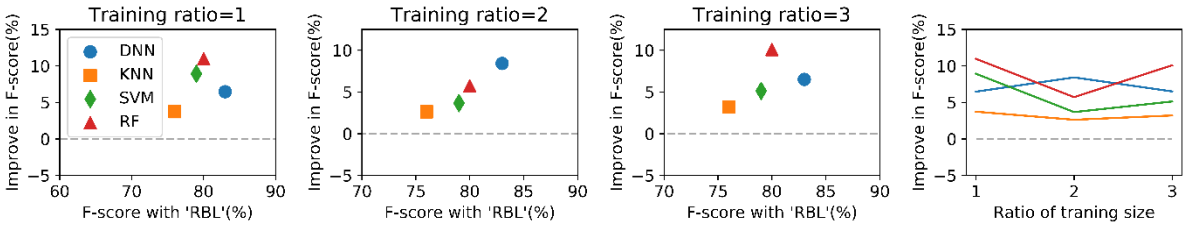
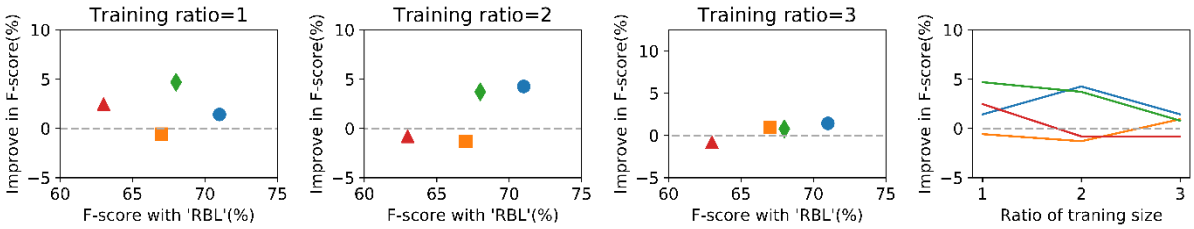


Figure 3-11. ROC curve across all target buildings.

- ***Sensitivity to the training size:*** To evaluate the impact of training sample size in our comparative analyses, we performed an experiment by changing the ratio of training sample for the scenario without "RBL" and compared it with the scenario that included "RBL". Specifically, in addition to the presented results in the previous section in which the ratio of training size in the scenario without "RBL" to the scenario with "RBL" was set to 3, we also performed the same experiments with the training size ratio of 2 and 1. Figure *3-12* presents the results for improvements in F-score by inclusion of RBL step in the framework. The F-score was averaged over the entire test set for different ratios of random sampling versus RBL for smart sampling. The first three subplots show F-score and improvement in F-score for different training ratios of 1, 2, 3. The x-axis represents the F-score using 'RBL' component, and y-axis shows the difference of F-score compared to the baseline of random sampling. The last subplot represents variation of F-score improvement across the training ratios. As can be seen, in most cases, the inclusion of "RBL" component improves the performance of algorithms since data points are placed above the baseline (the gray dashed line indicates no change). Besides, it confirms our assumption that setting the same number of buildings for both scenarios would further accentuate the improvement by including the 'RBL component. Additionally, results show varying sensitivities to the training size ratio. However, it does not considerably affect the performance.



(a)

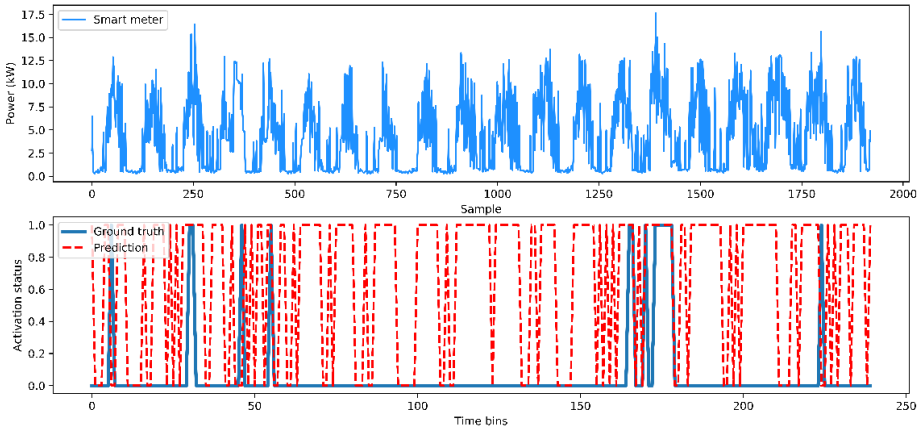


(b)

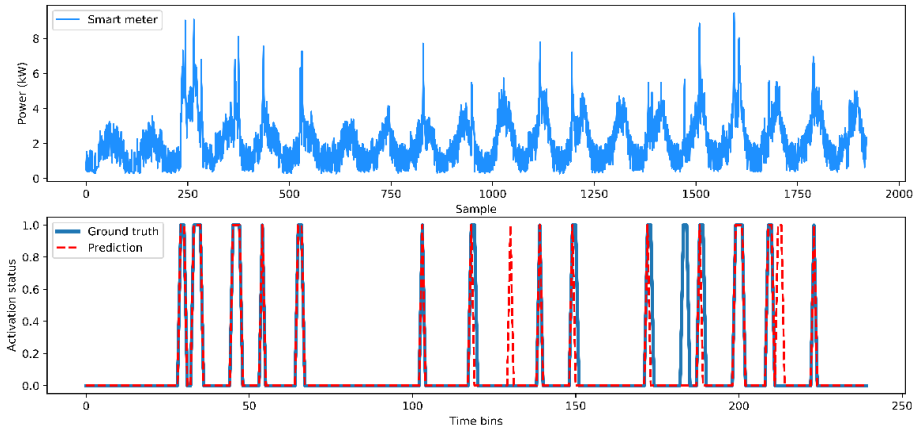
Figure 3-12. Comparison in F-score by changing the ratio of training size for (a) EV and (b) dryer.

3.5. Discussion and limitation

There are a number of limitations associated with this work as we elaborated here. We only focused on two classes of loads including EV and dryer that have high power draws. The usage of these devices can induce some noticeable change in load shapes pattern which make them favorable for our analysis. However, the identification of ToU for appliances with lower power draws will not make such changes in the load shape pattern, and therefore, might not be detected by the proposed approach. Nonetheless, we focused on two classes of deferrable loads that had the highest importance to DR applications in the residential sector [147, 148]. Moreover, as the results in Table 3-2 showed, although the framework showed relatively high detection rate in most buildings, there are a few instances that turned out to be challenging.



(a)



(b)

Figure 3-13. Comparison of dryer identification for two buildings with (a) worst performance and (b) good performance.

Figure 3-13 compares the ToU identification of dryer for a building with the worst performance (Building #9: ID 5288; F-score=0.34) with another building with a good performance (Building #5: ID 7901; F-score = 0.91). As could be seen, for the building with the high error value (Fig. 13(a)), the demand magnitude is considerably high at most times. Therefore, it could be considered an outlier in terms of energy consumption. Furthermore, there are many considerable sharp peaks in the load shapes, and the associated time bins were classified with dryer operation as false positives. This could be caused by simultaneous interference from other appliances that have similar load behaviors such as water heater or AC. To alleviate problems for such cases, we could add heuristics to detect outliers by using the known power draw information from typical appliances, and integrate contextual attributes such as occupancy information and outdoor temperature data for estimation of baseline loads. Additionally, outlier detection techniques can be applied to detect load profiles with discord in the building pool as a pre-processing or post-processing step [149, 150].

3.6. Conclusion

In this study, we have investigated the development of a scalable framework for inferring time-of-use (ToU) for major flexible loads in support of applications such as integration of distributed energy resources (DERs) and demand-response operations. The framework draws on the use of low-resolution real power data, sampled at 15-minute intervals through ubiquitously available smart meter infrastructure, and a training scheme that excludes in-situ training – i.e., inference is

carried out for buildings that are not part of the training data set. The framework uses a number of known buildings with specialized appliance-level metering instrumentation to provide the basis for generalized training (i.e., the training pool), and searches in the training pool to identify the buildings (i.e., the training sample) that their energy behavior matches that of a target building. The framework employs a segmentation component that uses clustering of daily load shapes and compares the frequency of observed clusters across buildings to match them as building with similar energy behavior. Upon identification of the training sample, a classification algorithm that uses the daily load shapes as the input parameter, detects the use of flexible loads across different times of a day. Various classifiers were employed to evaluate their performance in identifying the ToU. A case study on Pecan Street Dataport for EV and dryer was conducted. Using only the information from limited number of buildings in the training sample, the best average recall of 87% and 76%, and precision of 83% and 73% were achieved for EVs and dryers, respectively. The results show the feasibility and potential of the approach for adding a new level of information to the daily load shapes (i.e., the ToU) that reflects on the casualty of the observed energy consumption pattern on the daily load shapes. Future directions of this research will include the use of machine learning algorithms for automated feature extraction, accounting for simultaneous operation of loads with high power draw through heuristic augmentation of the framework, incorporating contextual information such as occupancy or building area in addition to identification of other flexible load types.

Chapter 4: An Automated Spectral Clustering for Multi-scale Data

Afzalan, Milad, and Farrokh Jazizadeh. "An automated spectral clustering for multi-scale data." *Neurocomputing* 347 (2019): 94-108. DOI: https://doi.org/10.1016/j.neucom.2019.03.008

Abstract

Spectral clustering algorithms typically require *a priori* selection of input parameters such as the number of clusters, a scaling parameter for the affinity measure, or ranges of these values for parameter tuning. Despite efforts for automating the process of spectral clustering, the task of grouping data in multi-scale and higher dimensional spaces is yet to be explored. This study presents a spectral clustering heuristic algorithm that obviates the need for any input by estimating the parameters from the data itself. Specifically, it introduces the heuristic of iterative eigengap search with (1) global scaling and (2) local scaling. These approaches estimate the scaling parameter and implement iterative eigengap quantification along a search tree to reveal dissimilarities at different scales of a feature space and identify clusters. The performance of these approaches has been tested on various real-world datasets of power variation of appliance signature with multi-scale nature. Our findings show that iterative eigengap search with a PCA-based global scaling scheme can discover different patterns with an accuracy of higher than 90% in most cases without asking for *a priori* input information.

4.1. Introduction

Clustering, the practice of partitioning data into different groups with similar observations, has a variety of applications in knowledge discovery for unknown phenomena in different fields such as object recognition [151, 152], cyber-physical systems [17, 153], or bioinformatics [154]. Spectral clustering [155-157] is a data analytics technique that has gained popularity in recent years. Due to its capability of high-quality clustering and handling non-convex clusters that are typically challenging for other methods [157], spectral clustering has been implemented in different domains like computer vision and speech separation with promising performance [158-161]. An overview of the literature reveals that spectral clustering has been *adopted and adapted* for different application domains. In this study, we have explored automated spectral clustering for feature spaces with multi-scale and higher dimensional attributes. Our vision has stemmed from the need for self-configuring algorithms in cyber-physical systems that need to adapt their behavior in different settings.

Spectral clustering decomposes the eigenvectors of a Laplacian matrix derived from an affinity matrix (i.e., similarity matrix) of the data and transforms the data into a new dimension, where it can be grouped with k-means or other algorithms that minimize a distortion metric. The affinity matrix in this context demonstrates the pairwise similarity between data points and is used to overcome the difficulties due to the lack of convexity in the data distribution. While considered as an unsupervised method, the algorithm calls for the determination of the number of clusters and a scaling parameter (that defines the behavior of the affinity matrix), which require algorithm tuning and *a priori* data provision. These values are commonly provided based on data-driven parameter tuning (i.e., model selection) techniques or the general knowledge of a domain and therefore, make the autonomous application of algorithms more challenging. Accordingly, research efforts have been made on automated (also known as self-tuning) spectral clustering with a focus on particular problems.

The main body of work in automated spectral clustering has focused on challenging two-dimensional and image segmentation problems (e.g., [162-165]). Automated spectral clustering for multi-scale high dimensional data (mainly time-series) is another domain of research that has been less explored. Clustering in higher dimensions could be a challenging task [166]; however, it has interesting applications in domains such as energy or power consumption pattern analysis, gene expression groupings, or speech separation. In other words, with emerging (and fast-growing) technologies for autonomous systems and smart environments, mainly in the form of cyber-physical systems, the need for self-tuning and context-aware algorithms that do not require human intervention for cluster analysis is increasing. The challenges for clustering of this type of data include: (1) Different groups of data can reside on different scales creating a multi-scale nature for the feature space. Consequently, larger scale components can mask the distinction of complex patterns in the smaller scales, and (2) the presence of noise in the acquired data could add to the complexity of the clustering process. Accordingly, in this study, we have proposed a heuristic algorithm for automated spectral clustering of multi-scale higher-dimensional data in order to obviate the need for *a priori* information (i.e., number of cluster k, and scaling parameter σ) so that the algorithms could configure their behavior by learning from the data.

Our proposed approach is built on the eigengap metric by introducing a new heuristic algorithm that couples eigengap with data-driven estimation of scaling parameter and a search framework that accounts for the multi-scale nature of the feature space. The performance of the proposed

heuristic has been evaluated on real-world labeled datasets with multi-scale nature in a higher-dimensional space and compared to the performance of commonly used internal validation techniques that call for a threshold as the stopping criterion (i.e. the number of cluster optimization). The proposed method was initially motivated by the task of energy disaggregation, which is the practice of dividing the aggregate power series into individual appliance components with considerable power draw values (i.e., different scales). While there are recent well-known clustering studies for electricity energy monitoring (e.g. [137, 167]), they presume the number of appliances and only handle appliances with high power draws. However, due to our interest in the automation of cyber-physical systems, we are seeking to perform clustering without parameter-tuning or *a priori* information (i.e., number of clusters) provision.

The rest of the paper has been structured as follows: In section 4.2, a background on automated (i.e., self-tuning) spectral clustering as well as clustering of high dimensional data is presented. Section 4.3 presents the proposed heuristic by introducing methods for estimation of scaling parameter and the number of clusters (K) that formalize our proposed framework for automated clustering. Section 4.4 discusses the datasets and their properties and then proceeds with presenting the results, evaluation, and efficacy of the heuristic algorithms. Finally, the conclusion summarizes the work and its findings.

4.2. Related Works in Spectral Clustering

Spectral clustering has gained popularity due to their ease of implementation and efficiency in clustering [168, 169]. Therefore, in recent decades, several clustering algorithms have been proposed and used for different applications. The focus in these algorithms has been on the application of the similarity matrix spectrum for dimensionality reduction and feature space transformation to introduce convexity. One of the well-known algorithms in this field is the one proposed by Ng, Jordan, and Weiss (referred to as NJW) [157]. In addition to the efforts in the formalization of spectral clustering algorithms, a number of studies have focused on expanding the algorithms into instances, which are capable of self-tuning or automated identification of natural partitions (or groups) in the data. Natural in this context refers to the clusters (or groups) that represent the actual/physical separation in the data.

4.2.1. Automated Spectral Clustering

As a widely adopted technique, Zelnik-Manor and Perona introduced a self-tuning spectral clustering algorithm [162] (built on NJW) that accounts for multiple scales in the feature space and automatically identifies the number of clusters using an optimization technique over a range of possible numbers for clusters. As part of this algorithm, they have proposed a novel similarity measure that integrates a data-driven scaling parameter by considering the distance of each point with some of its nearest neighbors. Scaling parameter refers to a parameter that controls the width of the neighborhood in the similarity metric. The number of clusters was estimated through examining a range of possible group numbers, recovering the rotation that best aligns the eigenvector of the matrix obtained from the data, and minimizing a cost function for possible rotations. The algorithm's performance has been evaluated on a number of reference 2D problems (identified as benchmarks) as well as image segmentation problems, with promising performance. A major part of the efforts in the field of automated spectral clustering has focused on the problem of image segmentation and thus the aforementioned study (i.e., [162]) has been used as the benchmark for comparative analyses. In a class of these studies, it has been argued that the eigenvector selection is a crucial task for clustering because not all of the largest vectors are informative for natural segmentation of the data. These studies mainly sought the task of automatic determination of the number of clusters under noisy and sparse data. Different methods have been proposed to account for eigenvector selection. Identifying the relevance of the eigenvectors according to their contribution in determining the number of clusters [163] and eigenvector selection through direct entropy ranking or a combination of elements in the ranking [164] are examples of these methods. Other studies have proposed alternative solutions to address the problem of automated clustering in image segmentation. For example, [165] uses non-normalized information of eigenvectors (rather than using a unit space for feature representation) and [170] performed iterative cluster and merge in order to address the problem of image down-sizing (which can lead to losing fine details).

Unlike the aforementioned efforts that have proposed solutions for challenging 2D datasets and image segmentation, [171] proposed a kernel spectral clustering for a large-scale network without parameter input. To this end, entropy was used to detect the block-diagonal of the affinity matrix that was created by the projections in the eigenspace. The efficacy of the proposed approach was studied through synthetic data and real-world network datasets. While these existing approaches

[163-165, 170, 171] were developed to tackle spectral clustering in an automated manner, they are either designed to solve the problem for multi-scale 2D and image segmentation or network data, which is different in nature from data with multi-scale higher dimensional attributes as sought here.

4.2.2. Spectral Clustering in Higher Dimension

Spectral clustering for higher dimensional feature spaces has also been the subject of some studies (e.g. [172-177]) to address different challenges. One of the examples of higher dimensional spaces in the real-world application is the time-series data. High level of noise and uneven sequence of length in data representation were among the challenges that have been taken into account. A class of studies has coupled spectral clustering and hidden Markov models (HMM) to benefit from structure and parametric assumptions of HMMs. These algorithms were evaluated on real-world datasets of motion capture, handwriting time-series sequence, sign language, and noisy sensor network data (e.g., [172], [178]). The inevitable challenge of noise in real-world data has led to studies on spectral clustering approaches that are robust to noise. Examples of techniques that focused on robustness to noise include using a mapping approach based on regularizations into a new space to separate the noise points in a new cluster [173], and proposing a partitioning criterion (discriminative hypergraph) which considers the intra-cluster compactness and inter-cluster separation of vertices [179]. The performance of these studies was evaluated on datasets including digit numbers with 256 features and gene expression data. In another class of studies with higher dimensional features, clustering of large-scale datasets (both in the number of features and instances) were explored since they are computationally expensive [174-177, 180]. These approaches typically integrate sparse coding-based graph or apply approximation methods to reduce cost while the performance might be deteriorated. Among the works that attempted to enhance the performance, we can mention the application of a landmark-based spectral clustering [176] that selects representative data points so that original data points are the linear combination of these landmarks and utilization of a sparse matrix and local interpolation to improve the approximate outputs [177]. These studies had a focus on the efficiency and improved performance of the algorithm or been applied toward a specific application solution. Therefore, they have considered parameter selection and prior knowledge of the domain.

Considering the existence of real-world data with higher dimensional attributes, our study focuses on a heuristic spectral clustering algorithm that can robustly reveal different groupings for

a class of multi-scale data for *autonomous* systems that need to adapt to different contexts.

4.3. An Automated Spectral Clustering Heuristic

The fundamentals of spectral clustering methods have been extensively described in the literature (e.g., [157], [168], [162], [181, 182]). Our heuristics is built on the NJW spectral clustering algorithm [157]. A brief description of the NJW algorithm is followed to expand on it for our extended algorithm.

Assuming the data set $S = [s_1, s_2, s_3 \dots, s_n] \in R^{n \times m}$ with K clusters, the NJW algorithm steps are as follows:

1) Develop the affinity matrix $A \in R^{n \times n}$, defined by:

$$A_{ij} = \begin{cases} \exp\left(-\frac{\|s_i - s_j\|^2}{2\sigma^2}\right) & i \neq j \\ 0 & i = j \end{cases} \quad (1)$$

where σ^2 is the scaling parameter of the model.

2) Using D, a diagonal matrix with the summation of the elements on the i-th ($i \in [1, 2, \dots, n]$) row of A as $D(i, i)$, the Laplacian matrix is defined as

$$L = D^{-\frac{1}{2}} A D^{\frac{1}{2}} \quad (2)$$

3) Compute $v_1, v_2, \dots, v_K$ the K largest eigenvectors of L, and form the matrix $V = [v_1, v_2, \dots, v_K] \in R^{n \times K}$

4) Form matrix $Y \in R^{n \times K}$ by renormalizing each row of V as

$$Y_{ij} = {V_{ij}} \Big/ {(\sum_j V_{ij}^2)^{\frac{1}{2}}} \quad (3)$$

5) Cluster each row of Y as a point in R^K via K-means algorithm.

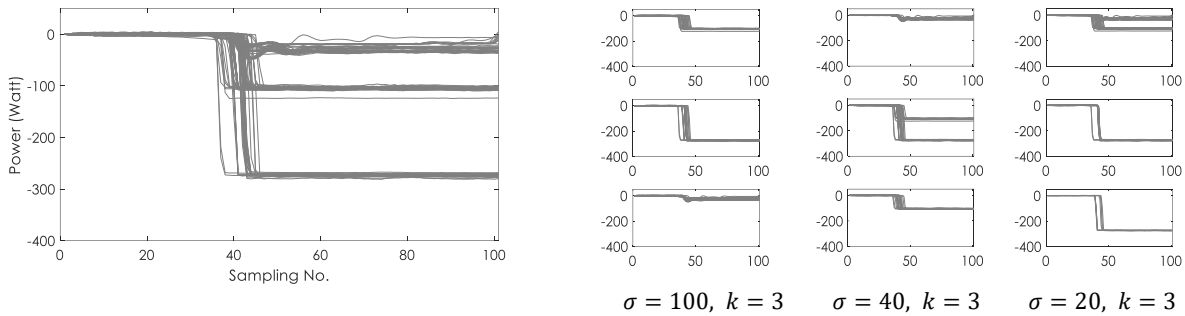
6) Original point s_i belongs to cluster *k*, if and only if row *i* of the matrix Y is assigned to *k*.

As the above steps state, the algorithm calls for input information. This information includes (1) the number of clusters (*K*), similar to other clustering algorithms that either need *K* (e.g., the well-known K-means) or other input parameters such as thresholds as stopping criteria or model parameters (e.g., hierarchical clustering or mean-shift [183], [184]), and (2) a scaling parameter (σ^2) for forming the affinity matrix. These parameters could be estimated by human knowledge for specific problem domains or through internal validation, which also requires a threshold identification as an input parameter. Specifically, as proposed by Ng et al. [157], scaling parameter

can be automatically fine-tuned by running the algorithm several times and selecting an optimal value from a range so that least distorted clusters of the rows in Y are obtained. However, identifying this range calls for knowledge of the data, which contradicts the self-configuration objective.

4.3.1. Estimating Scaling Parameter, σ^2

The scaling parameter (σ^2), shown in Eq. (1), defines the width of neighborhoods which subsequently affects the calculation of the affinity matrix. In other words, it is a reference distance, below which two points are evaluated as similar and beyond which dissimilar [162], [185]. Ng et al. [157] describe the scaling parameter as the parameter that controls how rapidly the affinity falls off with the distance between two observations (i.e., data points). Therefore, selection of this parameter characterizes the dissimilarity in the feature space and thus the structure of clusters. In order to illustrate the effect of scaling parameter on the clustering, we have presented the outcome of spectral clustering on a power dataset, with high dimensional data points (which are subsections of a power time-series) in *Figure 4-1*. The data in this figure represent selected feature vectors for a problem of energy disaggregation, which uses signal processing and machine learning algorithms to identify the contribution of individual appliances on the aggregated power time-series. The time-series data is collected through one sensor on the main circuit panel in a building to avoid extensive instrumentation. Thus, this figure also shows the challenges of clustering in energy disaggregation. The spectral clustering outcome was visualized in *Figure 4-1*(b) for different σ values to demonstrate the sensitivity. NJW algorithm was used with three as the number of clusters. While in all the cases the number of clusters is set to the correct number (*k=3*) as depicted in *Figure 4-1*. (a), variation of σ can affect the performance. As in this case, $\sigma = 100$ is a suitable estimation while $\sigma = 20$ or $\sigma = 40$ leads to false prediction. The performance is sensitive to the scaling value over different datasets, indicating the importance of correct inference for the automated approach. Therefore, in our proposed heuristic, we have adopted the following methods for data-driven scaling factor estimation.



(a) (b)

Figure 4-1.Sensitivity of σ on clustering output for a power time-series data. (a) is the feature vectors for three appliances and (b) shows the clustering results for three different values of σ. ($\sigma = 100$ gives the right prediction).

4.3.1.a. PCA-based Scaling Parameter

Scaling parameter identifies the boundaries of the similarity neighborhood. A larger σ indicates the similarity with more distant data points, whereas smaller values highlight the neighboring points. Therefore, in order to estimate the scale of neighborhoods, we have adopted the application of principal component analysis (PCA), which utilizes an orthogonal transformation to map the original variables into new space with uncorrelated variables. Given the higher dimensionality of data, we employ PCA in our approach to ensure that we focus on components of the feature space that account for the most variance in the data. In this study, through observations, we consider the one-time standard deviation of major principal axes (that accounts for the maximum variance in the whole data) in order to estimate the σ. This assumption allows us to form an approximate boundary threshold for distinguishing similar and dissimilar points based on the distribution of data points. Since only the first few components constitute the most variance, the number of considered principal axes is selected such that at least 95 percent of total variance is granted. This ensures reducing the input while also accounting for the whole variability of data.

Let us consider a set of data points S with n observations and m features as:

$$S = [s_1, s_2, s_3 \dots, s_n], \qquad S \in R^{n \times m} \tag{4}$$

The eigenvectors that correspond to the highest eigenvalues of the covariance of S are associated with the highest variance. The projection matrix U is formed by stacking the eigenvectors of corresponding eigenvalues sorted in the descending order. Using U, the sample data is transformed into the new space as follows:

$$P = S \times U, \qquad P \in R^{n \times m} \tag{5}$$

Each principal component is derived by selecting the corresponding column from P. We employ the variance information from principal components for evaluating the scaling parameter. Therefore, the scaling parameter (σ^2) will be estimated as follows:

$$\sigma^2 = \frac{\sum_{i=1}^{y} w_i v_i}{\sum_{i=1}^{y} w_i} \tag{6}$$

where w_i is the data-driven weighting factor from $\boldsymbol{w}$. w_i denotes the ratio of the variance for the $i-th$ principal axis to the summation of variance from all the principal axes (contained in P):

$$\boldsymbol{w}^T = [w_1, w_2, \dots, w_m],\ w_1 > w_2 > \cdots > w_m \tag{7}$$

We select y major principal axes (in Eq.6) such that

$$\frac{\sum_{i=1}^{y} w_i}{\sum_{j=1}^{m} w_j} > 0.95 \tag{8}$$

Here, y is the smallest integer that satisfies the above inequality. The above inequality is used to consider almost the whole variability of data from PCA without considering all the principal axes. Typically, the first few principal axes account for the most variance in the data. An empirical analysis for Eq. (8) is provided in section 4.3.2.

Also, $\boldsymbol{v}$ contains the variance of data points that are projected along the principal axes (each axis contains n points, i.e., the total number of observations in the data).

$$\boldsymbol{v}^T = [v_1, v_2, \dots, v_m],\ v_1 > v_2 > \cdots > v_m \tag{9}$$

4.3.1.b. Local Scaling for Self-tuning

As a data-driven approach for estimation of the scaling parameter, Zelnik-Manor and Perona [162] suggested that, instead of considering a global parameter for the whole affinity matrix, a local scale for each point allows the point-to-point distance self-tuning, which can further be used to compute the affinity for pairwise points. As proposed by [162], instead of using Eq. (1) for calculating the affinity matrix, a local scale parameter is defined by each point, and Eq. (1) is re-written as

$$A_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_i \sigma_j}\right) \tag{10}$$

where

$$\sigma_i = \|x_i - x_k\| \tag{11}$$

x_k is the k'th nearest neighbor of x_i. Through empirical observations, a value of k=7 was suggested [162] that works for a range of applications. Local scaling provides a highly representative measure of scale for each data point. However, comparing to methods that use a global scale, this

improvement in estimation comes with a higher computational cost since it calls for a KNN search for each data point in the process of forming the affinity matrix.

4.3.2. Iterative Eigengap Search Heuristic

Determining the number of clusters is a challenging problem even for human users and selecting the "right number of groups" is subject to different interpretation. In the case of spectral clustering, a commonly used heuristic is the eigengap that measures the stability of the eigenvectors in the Laplacian matrix. Based on the matrix perturbation theory, the subspace spanned by the first i eigenvectors of the Laplacian matrix L is stable if and only if the eigengap measure in Eq. (12) is large:

$$\delta_i = |\lambda_i - \lambda_{i+1}|, i = 1, \dots, N-1 \quad (12)$$

where N is the total number of observations, $\delta_1, \dots, \delta_{n-1}$ are the eigengaps, and $\lambda_1, \dots, \lambda_n$ are the eigenvalues of the Laplacian matrix L (defined in Eq. 2).

The value of eigengap for subsequent eigenvalues can indicate the place to pick the number of clusters. By examining the eigenvalue measures, the number of clusters can be estimated through

$$K = argmax_i(\delta_i) \quad (13)$$

where δ_i is calculated through Eq. (12).

Eigengap heuristic is a technique that mainly works in well-separated feature spaces [10] but is not capable of properly partitioning the feature space in the presence of multi-scale data or mixed background as is the usual case in real-world data. In order to extend the capabilities of eigengap metric for multiscale data analysis, we propose to use an iterative eigengap search to reveal the complex topology of the feature space.

4.3.2.a. Iterative Eigengap Search with global scale

In a multi-scale feature space, the data in the larger scale could mask the dissimilarities and therefore the microstructure of the smaller scales. However, if an algorithm explores the structure of feature space in different scales, the dissimilarities could be revealed and therefore, the eigengap metric could be utilized for identifying the number of clusters. This is the rationale behind our proposed Iterative Eigengap Search (IES) that partitions a feature space through searching along a tree-like structure. While at each iteration the eigengap might not find the final groupings of data points, it will segregate the data in different scales and consequently accentuates the dissimilarities at each scale. Therefore, the algorithm could refine the clusters through visiting each node (i.e.,

cluster) from the previous iteration by passing it to the spectral clustering algorithm. The process of refining the clusters will be carried out until eigengap cannot reveal finer structures in a leaf node. In other words, the stopping criterion is when all the nodes of the tree have been visited and all the leaf nodes contain only one cluster according to the eigengap measure.

Fig. 2 illustrates the conceptual process of Iterative Eigengap Search tree. In this tree structure, the root indicates the whole dataset, which is passed through eigengap heuristic (Eq.(12)) to determine an initial K (Eq. (13)), and then clustered with NJW algorithm with a PCA-based global scaling parameter. Through empirical observations, we found that the K should be sought in the first half of the vector of eigenvalues. Each of the produced nodes is processed with eigengap heuristic to be clustered again. As schematically shown in *Figure 4-2*, there are 6 nodes with estimated k=1 (after being analyzed with eigengap) that are accepted as final clusters, which all together form the data in the root node. The final clusters are thus the ones at the leaf nodes. At each level of the tree and for each node, the σ measure is updated with respect to the content of that node to identify a scaling factor for that specific subset of data points.

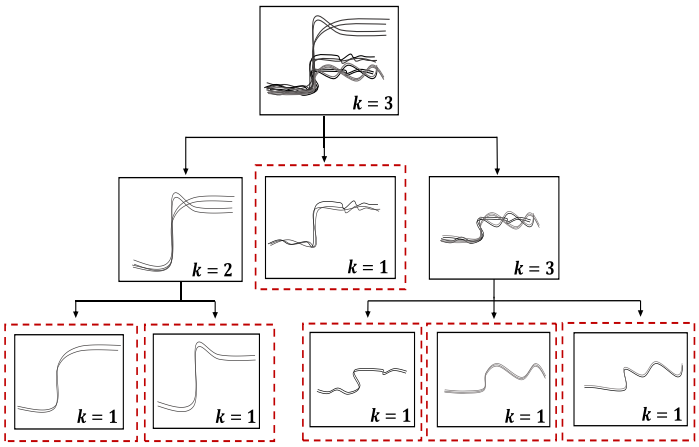


Figure 4-2. Framework of Iterative Eigengap Search (IES) for discovering patterns in different scales (groups with k=1 are accepted as the final clusters).

As described in section 4.3.1.a, we used the inequality in Eq. (8) to consider the major principal components. Considering the fact that the first major components typically account for the most variance in the data [186], we select the first major components such that at least 95 percent of variance is granted. As an empirical demonstration, we have considered all the datasets, later described in section 4.2, and measured the amount of variance and the percentage of major components for each generated node in the IES, as shown in *Figure 4-3*. As can be seen in *Figure 4-3* (a), the inequality results in an amount of variance that is close to 1 in our heuristic. On the

other hand, as shown in *Figure 4-3* (b), in most cases, the number of principal components is limited to a small portion of the features to achieve the objective described in Eq. (8). Therefore, we used the cut-off threshold of 0.95 to account for almost the whole variance by limiting the number of principal components.

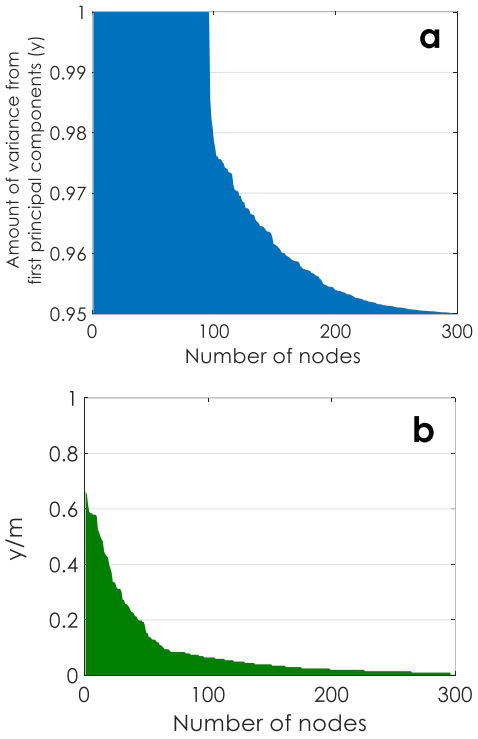


Figure 4-3. Empirical analysis for the selection of major components in the generated nodes in IES: (a) Amount of variance granted, (b) ratio of selected major components.

4.3.2.b. Iterative Eigengap Search with local scaling

In this alternative of the heuristic, we have adopted the local scaling parameter that identifies the scale by quantifying the distance between nearest neighbors in our search tree framework. As proposed by the original work by Zelnik and Penora [162], we have utilized 7 nearest neighbors for the σ estimation. The selected number of neighbors was suggested in [162] based on comprehensive analysis of both high dimensional and low dimensional data. This approach considers the impact of point-to-point distance in forming the affinity matrix such that multiple scales of data are accounted for. The local scale is used in the Iterative Eigengap Search to identify the structure in the feature space. Given that local scaling has already considered the multi-scale

nature of data, the results through one iteration could be considered as the clustering output, which we call eigengap with local scaling (ELS).

Figure 4-4 presents the pseudo-code for the Iterative Eigengap Search. The tree search is carried out similar to a depth-first search algorithm and therefore a stack data structure that uses the LIFO (last-in-first-out) feature is used to store the data (and the subsequent sub-sections).

Algorithm 1. Iterative Eigengap Search

Input: Feature matrix as D
Output: Final clustered data as C and total number of clusters as K.
1: Store D in stack S.
2: Initialize cluster number $K = 0$.
3: **while** $S \neq \emptyset$
4: Return σ (or σ_i) through **function Sigma** with **input:** *S.peek*(), and σ *type*.
5: Return number of clusters k and clustered data c_i(i=1 to k) through **function HSC** with **input:** *S.pop*() and σ
6: **if** (k<2)
7: K ++
8: Assign K as cluster label to c_1 and store the result in C.
9: **else**
10: **for each** i (i=1 to k)
11: Push c_i in S.
12: **end**
13: **end**
14: **end**
15: Return C and K.

peek(): Method that selects the top item in stack (without removing)
pop(): Method that selects and remove the top item in stack

Function Sigma

Input: *S.peek*() and σ *type*
Output: σ (for global scaling with PCA) or $\sigma_{i(i=1\ to\ N)}$ (for local scaling)
1: **Switch** σ *type*
2: **Case** Global scaling with PCA
3: Perform PCA on *S.peek*().
4: Form $\boldsymbol{w}$ and $\boldsymbol{v}$ (Eq. **7** and Eq. **9**).
5: Find $\boldsymbol{y}$ as the major principal axes from Eq. **8**.
6: $\sigma = \sqrt{\frac{\sum_{i=1}^{y} w_i v_i}{\sum_{i=1}^{y} w_i}}$
7: Return σ
8: **Case** Local scaling
9: Find $\sigma_{i(i=1\ to\ N)}$ from self-tuning method.
10: Return σ_i
11: **end**
end function

Function HSC. Spectral clustering by NJW and eigengap heuristic

Input: *S.pop*() and σ (or σ_i)
Output: Number of clusters k and clustered data c_i(i=1 to k)
1: Form the affinity matrix A and diagonal matrix D from *S.pop*() and calculate L
2: Compute eigenvalues $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ and eigenvectors $V = \{v_1, v_2, \dots, v_n\}$ of L
3: Sort λ in descending
4: Find k as $k = argmax_i\ (abs(\lambda_i - \lambda_{i+1}))$, $i = 1\ to\ n/2$
5: $X \leftarrow$ Concatenate k columns of V based on λ in descending.
6: $Y \leftarrow$ Normalize rows of X
7: Perform k-means on Y.
8: Return number of clusters k and clustered data c_i(i=1 to k).
end function

Figure 4-4. Pseudo-code for Iterative Eigengap Search Heuristic.

In this work, we have adopted NJW, which uses the normalized Laplacian matrix to extract the structure of the data as the standard spectral clustering (SC) for our automated clustering method. In the past recent years, different variations of SC have been proposed that showed improvement over NJW from specific perspectives including improved eigenvector selection [164, 187], alternate affinity matrix generation [188], and reduced computational cost [189, 190]. Nonetheless, we have adopted NJW as a seminal well-established algorithm. Considering the nature of the proposed framework for automated clustering, other variations of spectral clustering could be replaced instead, as long as they employ a graph Laplacian matrix (e.g., [168, 191]) that enables the use of eigengap heuristic and the scaling parameter in their similarity estimation.

4.4. Algorithm Evaluation

4.4.1. Evaluation Metrics

In this study, the algorithm performance has been explored through external validation and was compared with internal validation techniques. Clustering validation is a domain which determines the goodness of clustering output [181]. While external validation relies on the external data such as the class labels, internal validation only searches for the information in the data to check the goodness of partitioning, and can also be employed to find the optimal number of clusters [192]. Both data types used in this study are labeled, which enables us to use external validation. However, we are also checking the performance against commonly used minimization of the sum of squared error in cluster dispersion to contrast the algorithm performance against conventional methods of automated clustering.

In internal validation, different metrics typically consider the compactness (high intra-cluster similarity) and separation (low inter-cluster similarity) to estimate the quality of partitions. These metrics can be used as a measure to find the optimal number of clusters. To measure the dispersion (or tightness) of clusters, the sum of the squared error (SSE) [193], [194] can be measured as:

$$SSE_k = \sum_i \sum_{x \epsilon C_i} \| x - \bar{x}_i \|^2, i = 1,2, \dots, k \quad (14)$$

where x are the data points in cluster *i,* $\bar{x}_i$ is the centroid of cluster *i*, and *k* is the total number of clusters.

The SSE is measured for a set of clustering outcome for a range of *k* values to form an "elbow curve". The optimal number of cluster is decided based on the rate of dispersion by identifying a threshold for change between subsequent values on the elbow curve.

For the external validation, as the data is fully labeled, we have adopted precision, recall, and F-measure of the confusion matrix. Based on a majority vote, we assign a dominant label to a cluster and form the confusion matrix.

4.4.2. Dataset Description

Power consumption datasets

This category of data is focused on power time-series and the power draw of appliances in a typical building. As appliances change their operational states (e.g., going from off to on), the power draw changes. Clustering has applications in non-intrusive electricity consumption disaggregation, which uses minimal sensing in a building unit coupled with machine learning frameworks. More details on the need and challenges for automated clustering in this field of problems could be found in [130]. In order to shed light on the nature of this data type, *Figure 4-5* shows a sample of raw time series data over a 3-hour period with 60Hz resolution. The red circles in this figure illustrate the events, when operational states of appliances were changed.

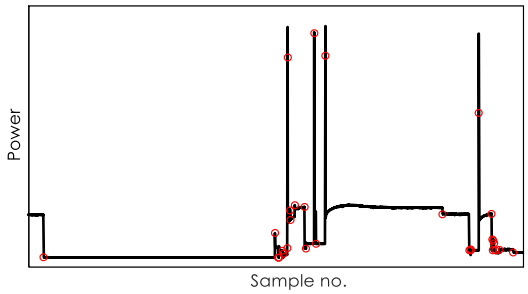


Figure 4-5. A sample of aggregate power variation time-series.

Events are detected using the Generalized Likelihood Ratio (GLR) event detection algorithm. Therefore, the dataset contains the noise due to performance of automated feature extraction algorithms as well. The transient in power draw in the vicinity of these events are defined as the appliance signatures and are used as feature vectors, rendering this problem as a feature-based time series clustering according to [195]. In this study, we have used the transient signature (comprised of real and reactive power) for one second after each event and 2/3 of a second before each one. This dataset has been collected and labeled in three occupied apartments over the course of two weeks [125], in which we used the data from the first apartment for our analysis. The dataset is fully labeled under human supervision with the data from ground truth sensors. The labels represent appliances operational states, and each appliance could have several operational states. These labels have been used for external validation.

Power data has a highly multi-scale nature, which has been visualized in *Figure 4-6* for one example dataset. This dataset contains 16 labels (i.e., different classes). In this figure, feature vectors for all instances have been plotted (only real power section of the vectors was presented). Going from left to right, feature vectors in the larger scales were recursively removed and thus the dissimilarities in smaller scales have been revealed. Differences in scales stem from differences in appliances' power draw. In the smallest scale, the dataset contains 7 clusters that are completely masked when the scaling parameter is not estimated according to that scale. The challenging task of clustering in this problem arises from the fact that automated clustering can simply overlook distinguishing patterns in the small-scale region. In this study, we have used four power datasets, for which in *Figure 4-7*, the average of variations for all the events of particular labels were plotted. The wide range of power variations for a multitude of labels in all the datasets demonstrates the multi-scale characteristic of this type of data.

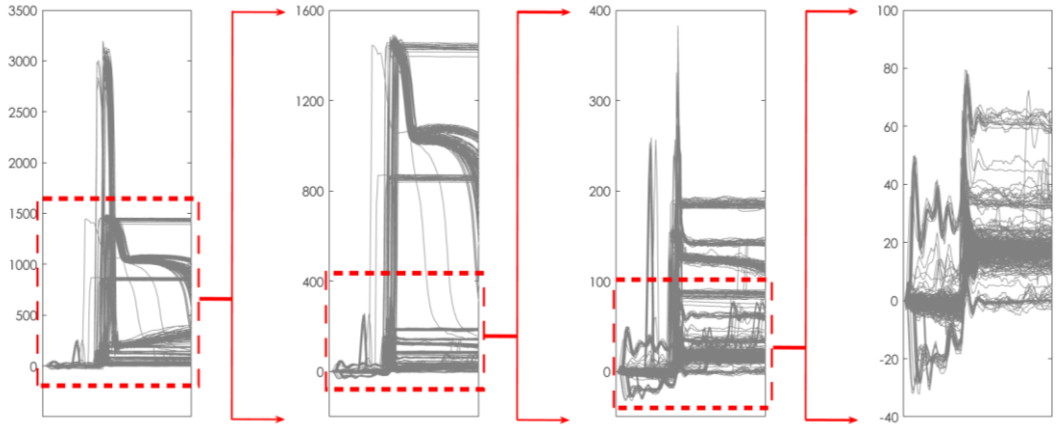


Figure 4-6. Visualization of a power dataset (time series signal) that shows the effect of multiple scales for the clustering problem; the presence of different clusters are magnified from left to right. In the right frame, 7 groups exist while in the left frame their presence is entirely concealed due to the presence of other patterns with high measurement difference.

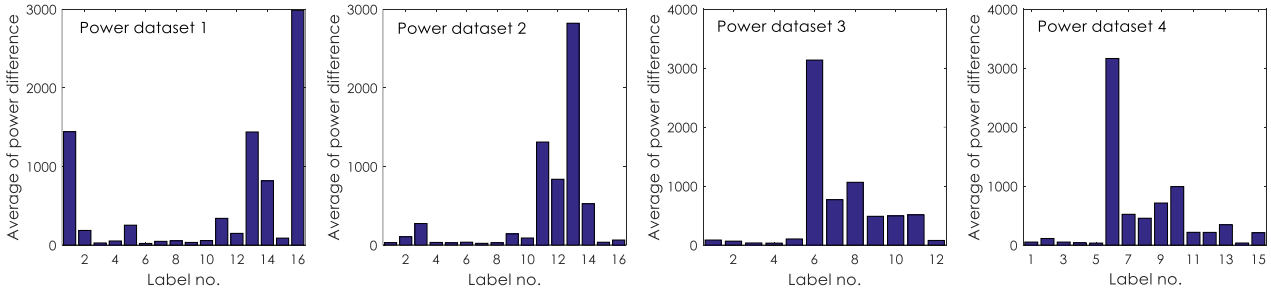


Figure 4-7. Average of measurement difference for different labels.

Table 4-1. Description of datasets.

Dataset	No. of data points	No. of features	No. of classes
Power dataset 1	756	202	16
Power dataset 2	498	202	16
Power dataset 3	1454	202	12
Power dataset 4	2235	202	15
Cell cycle dataset	384	17	5

4.4.3. Performance Assessment

In this section, we provide details on qualitative and quantitative assessments. In the former, the effect of the proposed algorithm on the quality of clusters has been visually described. The latter evaluates the performance by using metrics including accuracy, F-measure and computational time. A comparative performance assessment has been also included.

4.4.3.a. Qualitative Performance Assessment

The qualitative assessment is presented for selected datasets that help illustrate (accentuate visual variations) the challenges and performance of the algorithm. *Figure 4-8* and *Figure 4-9* visualize the clustering output for power dataset 1. *Figure 4-8* shows the clustering outcome with Iterative Eigengap Search (IES) with global scaling after the first iteration on the search tree. *K* is initially estimated as 3, and 3 child nodes are generated. As expected, conventional eigengap is not able to identify the structure of the feature space and resulted in low-quality clusters. *Figure 4-9* shows the outcome of the clustering for the iterative search of eigengap, in which 44 clusters were identified at the leaf nodes (where *k*=1). As *Figure 4-8* and *Figure 4-9* demonstrate, the Iterative Eigengap Search starts with a coarse level separation of clusters in the first iteration and refines the result iteratively to provide high-quality clusters as the outcome.

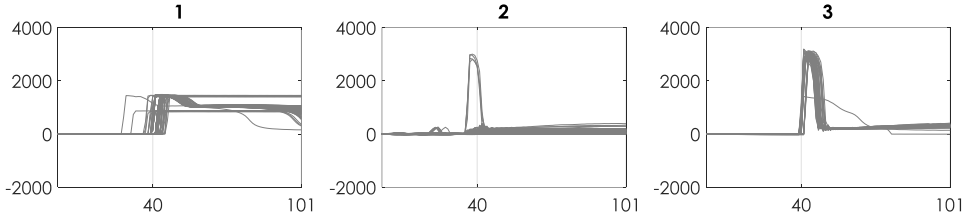


Figure 4-8. Clustering outcome after first iteration with the coarse-level division on power dataset 1.

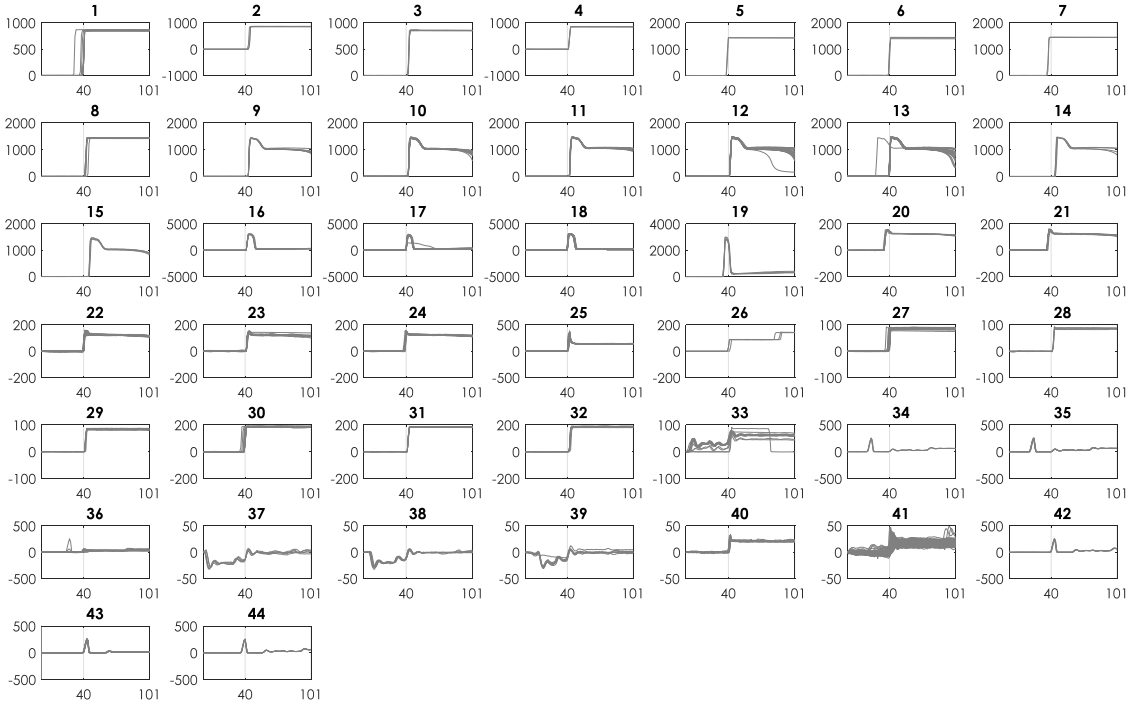
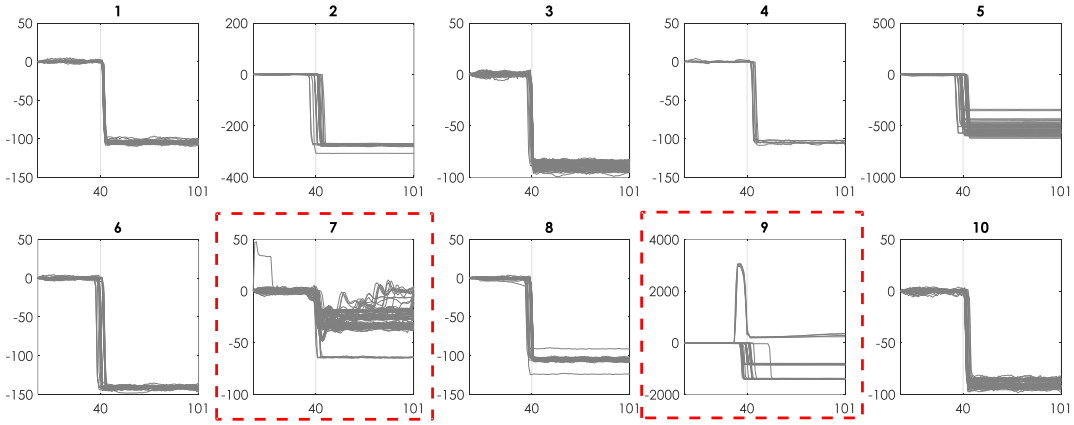
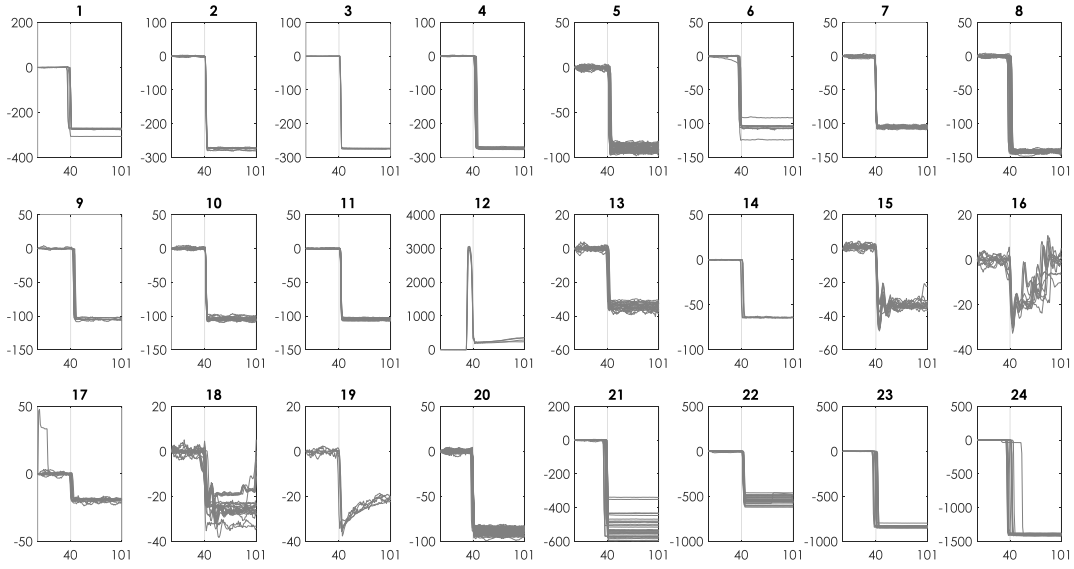


Figure 4-9. Clustering outcome using Iterative Eigengap Search with global scaling on power dataset 1.

Figure 4-10 shows the clustering outcome for Iterative Eigengap Search with local scaling on power dataset 2. In *Figure 4-10* (a), the clusters in root node were presented. As shown in this figure, a reasonable estimation for *k* is achieved, but there are few clusters (highlighted with dash lines), which potentially could be improved. *Figure 4-10* (b) presents how the iterative approach modifies the clusters. Local scaling results in a higher number of clusters in the first iteration with a shallower tree structure.



(a)



(b)

Figure 4-10. Clustering outcome using Iterative Eigengap Search with local scaling on power dataset 2: (a) cluster outcome after first iteration (clusters with potential for improvement were highlighted) and (b) cluster outcome on the leaf nodes.

The quantitative impact of this difference, both in terms of accuracy and computational time (given local scaling has a higher computational cost), will follow.

4.4.3.b. Quantitative performance assessment

Given that the labels for data points are known, we carried out external validations to quantify the algorithm performance in comparison to state-of-the-art and conventional internal validation. For internal validation, by considering a range for the number of clusters in ascending order, the SSE was obtained for each dataset using the concept described in Eq. (14). *Figure 4-11* shows the elbow curves for all power datasets.

In all cases, PCA was used to estimate the scaling factor (σ). Since spectral clustering algorithm implements K-means in the last step, the results are affected by the random initialization of centroids, and thus the structure of the elbow curve for the subsequent number of k's would be affected. To avoid this bias, we assigned fixed initial seeds values (i.e., the same specific data points for initialization of K-means) for all k values in forming the elbow curve. As shown in *Figure 4-11*, the noisy structure of data brings about inconsistencies in descending trend of SSE values as k increases, though the general pattern of flattening for measurement is preserved, except for case (d). For case (d), since the elbow curve within the considered range of k is not presented, estimating k based on this plot has not been considered. Upon forming the elbow curves, we have manually selected the number of clusters by visual evaluation of the rate of decrease in values for the internal validation.

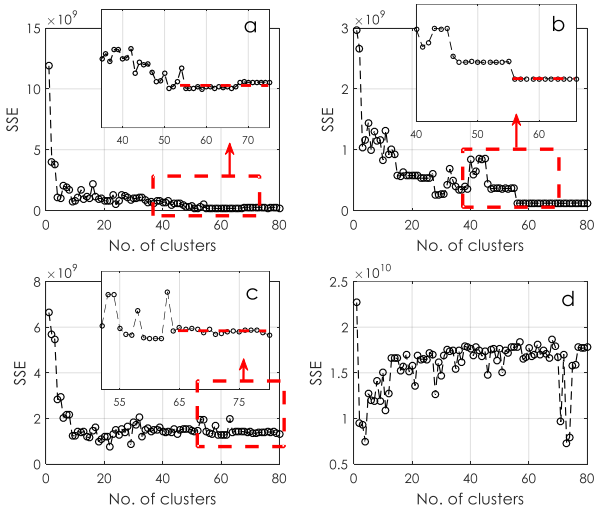


Figure 4-11. Elbow curves for a) power dataset 1, b) power dataset 2, c) power dataset 3, and d) power dataset 4.

To associate the cluster label with the ground truth, we form a matrix that relates the cluster number (assigned by the algorithm) for each observation to its corresponding ground truth label and call it the *association matrix* henceforth. As an example, *Table 4-2* shows the association matrix for power dataset 1. The association matrix is mapped to a confusion matrix for the performance quantification. In forming the confusion matrix, clusters are labeled based on the majority vote. In order to provide insight on labeling clusters with the majority vote, let us consider cluster number 23, which contains 20 feature vectors (i.e., data points) in total. It could be seen that 19 data points are from class 14501 and one instance is from class 18001. Considering the dominance of class 14501, it is assigned as the label for cluster 23.

Table 4-2. Association matrix between generated clusters and ground truth label (Power dataset 1).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
11101									8	21	18	27	43	4	4		1																											
11102																														20	8	9													
12900																																				2			1		177			
12901																																	14											
12902																																										5	14	6
12903																																						5	8	7				
12906																																		3	7	2								
14101																																				12								
14201																																				1				13	8			
14301																																				8								
14401																						1			30																				
14501																				7	14	70	19	29																				
16201					2	8	2	3																																				
16301	10	3	3	4																																							
18001																							1			3	24	4	7				2										
18002																8	19	23	4																									

Ideally, the number of generated clusters will be the same as the number of ground truth labels. However, as the number of generated clusters exceeds the number of ground truth labels, clusters with the same label will be merged to form the confusion matrix. For example, the contents of column 20 to 24 in *Table 4-2* are cumulated and the groups are merged since they all represent class 14501. *Figure 4-12* shows the output of mapping to form the confusion matrix for power dataset 1.

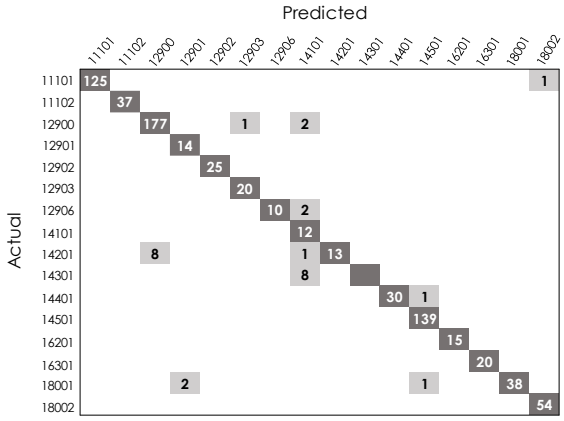


Figure 4-12. The equivalent confusion matrix mapped from association matrix (power dataset 1).

The final clusters after assigning the associated label and (manual) merging of similar clusters is shown in *Figure 4-13*. Except for class 14301 that has similar signature representations with 14101, all other classes were retrieved and preserved through the clustering process. It must be noted that the abovementioned process of manual merging was performed only for visualization of clusters in association with the classes in the physical environment. Nonetheless, the quantitative performance assessment of the clustering process was carried out with respect to the direct output of the clustering algorithms through the association matrix (*Table 4-2*) without any cluster merging.

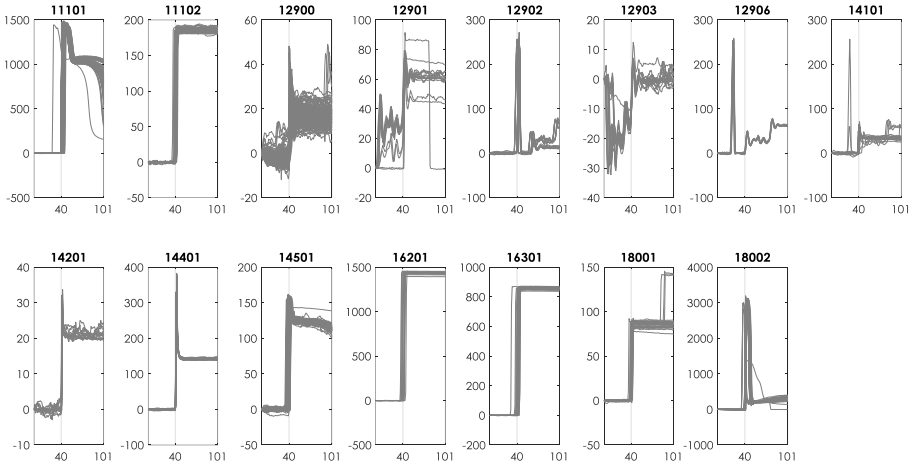


Figure 4-13. Cluster representation after (manual) merging for power dataset 1. Each label represents an appliance transition state.

To demonstrate the effectiveness of the Iterative Eigengap Search (IES) that is the focus of this study, for comparison, we provided the outcome of the following spectral clustering algorithms: (1) similar to our work, the ZP self-tuning technique [162], and MEG-CD [196] are automated spectral clustering methods, (2) FUSE spectral clustering [197] is specifically focused on multi-scale datasets; (3) and NJW [157] and CPQR [198] are conventional spectral clustering algorithms that call for parameter inputs. In addition, the results for the internal validation, which is commonly used for the validation of clustering methods, as well as the legacy eigengap heuristic have been presented.

Calculated from the association matrix, *Table 4-3* presents the performance metrics of different methods on all the datasets. Five-fold cross-validation was used for evaluation, and the average is reported for the accuracy, precision, recall, and F-measure metrics to avoid bias in the quantification of performance metrics. As the values in *Table 4-3* indicate, Iterative Eigengap

Search with global scaling shows the best performance. Followed by that is the Iterative Eigengap Search with local scaling with fewer number of clusters, which is more compatible with the natural separation of patterns in the feature space. As noted, Iterative Eigengap Search does not require any range for the number of clusters. For the ZP self-tuning [162] and MEG-CD [196], although considered as automated clustering, they required a range of initial values to optimize over the number of clusters. Therefore, a range of 2 to 80 clusters for power datasets, and 2 to 20 for the cell cycle dataset were considered for ZP self-tuning [162] and internal validation. A range of 100 to 3000 with intervals of 100 for σ was considered for MEG-CD [196]. We chose this range based on our empirical observations on the dataset. Since ZP self-tuning [162] underestimates the number of clusters, it does not result in an accurate outcome, specifically for the power datasets. Similarly, legacy eigengap heuristic leads to a smaller number of clusters and consequently low performance due to its incapability of accounting for the multi-scale nature. MEG-CD [196] performed better in terms of estimating the number of clusters but the inaccurate estimation of σ (shown by the internal validation) led to a low performance in clustering. For NJW [157], since both K and σ are estimated manually, K is assumed to be equal to the number of classes (input information), and σ was selected such that the clusters with smallest distortion are obtained. The results show that while K was manually selected to its true value, the performance in all the cases falls behind the Iterative Eigengap Search. For the recent multi-scale clustering algorithm, FUSE [197], the IES outperforms as well. Also, CPQR [198] showed to be less accurate compared to the IES in all cases, but the approach has the highest computational efficiency among all the methods. The last column in *Table 4-3* shows the total analysis runtime. It must be noted that for the NJW algorithm [157] ZP self-tuning [162], and MEG-CD [196], we had to define a range for σ and K, which consequently affects the reported runtime. Based on the extent of familiarity with the problem, a higher or lower range can be defined, which can significantly change the reported time.

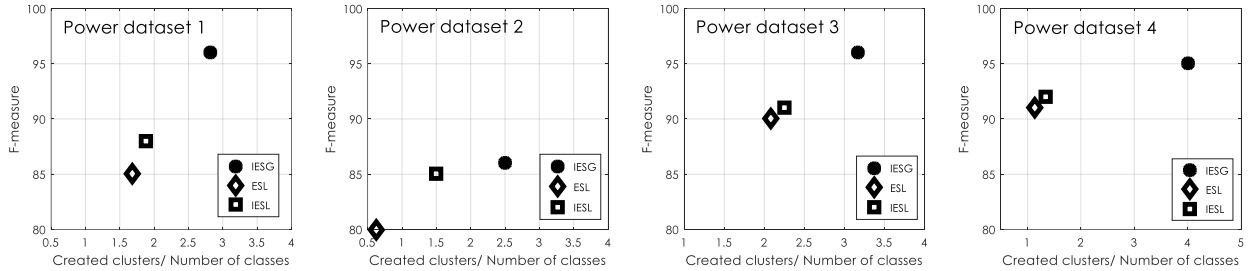
Table 4-3. Performance quantification of different methods (Performance metrics are averaged over 5-fold cross-validation). The first three methods (indicated in bold) are proposed.

Method	Dataset	σ selection	K	Accuracy[a]	Precision[a]	Recall[a]	F-measure[a]	Runtime (s)
IES with Global Scale (leaf node clusters)	Power data1	PCA	44	**0.97**	**0.97**	**0.97**	**0.96**	15
	Power data2		40	**0.89**	**0.85**	**0.89**	**0.86**	9
	Power data3		38	**0.97**	**0.96**	**0.97**	**0.96**	24
	Power data4		60	**0.96**	**0.95**	**0.96**	**0.95**	78
One step IES with Local Scale (ESL)	Power data1	Local Scaling	27	0.88	0.84	0.88	0.85	6
	Power data2		10	0.84	0.78	0.84	0.80	4
	Power data3		25	0.93	0.89	0.93	0.90	8
	Power data4		17	0.93	0.89	0.93	0.91	13
IES with Local Scale (leaf node clusters)	Power data1	Local Scaling	30	0.90	0.87	0.90	0.88	55
	Power data2		24	0.89	0.84	0.89	0.85	43
	Power data3		28	0.93	0.89	0.93	0.91	50
	Power data4		20	0.94	0.90	0.94	0.92	47
Internal validation	Power data1	PCA	55	0.93	0.89	0.92	0.91	89[b]
	Power data2		56	0.87	0.83	0.87	0.85	40
	Power data3		45	0.94	0.94	0.94	0.93	201
	Power data4		N/A[c]	N/A	N/A	N/A	N/A	1201
Legacy Eigengap	Power data1	PCA	3	0.44	0.25	0.44	0.30	1
	Power data2		3	0.37	0.17	0.3	0.22	<1
	Power data3		6	0.60	0.36	0.59	0.44	5
	Power data4		5	0.51	0.26	0.51	0.35	15
NJW [157]	Power data1	Least Distortion	16[d]	0.82	0.73	0.82	0.76	30[b]
	Power data2		16	0.66	0.61	0.66	0.61	20
	Power data3		12	0.89	0.82	0.89	0.85	115
	Power data4		15	0.54	0.30	0.54	0.38	520
MEG-CD [196]	Power data1	Local Scaling	12	0.47	0.29	0.47	0.34	7[c]
	Power data2		6	0.32	0.13	0.32	0.17	4
	Power data3		7	0.58	0.35	0.58	0.43	34
	Power data4		3	0.51	0.26	0.51	0.35	111
CPQR [198]	Power data1	Local Scaling	16[d]	0.80	0.76	0.80	0.77	2
	Power data2		16	0.83	0.76	0.83	0.80	1
	Power data3		12	0.86	0.75	0.86	0.80	2
	Power data4		15	0.91	0.84	0.91	0.87	4
Self-tuning ZP [162]	Power data1	Local Scaling	3	0.47	0.23	0.45	0.30	445[b]
	Power data2		2	0.40	0.16	0.40	0.23	225
	Power data3		4	0.81	0.67	0.81	0.73	584
	Power data4		4	0.92	0.85	0.92	0.88	850

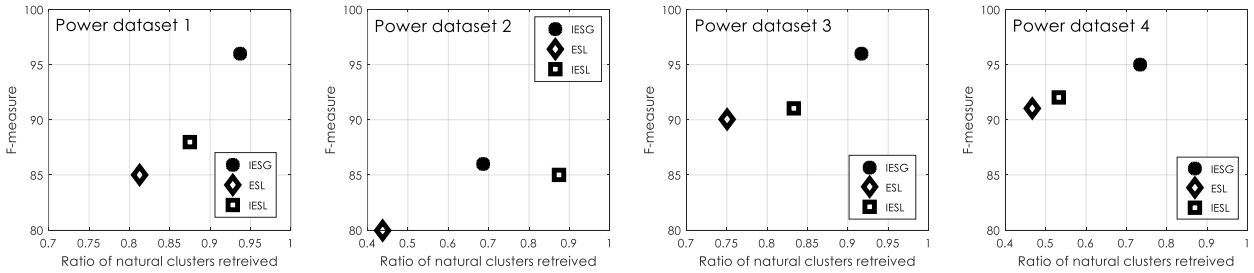
[a] These values are calculated based on the weighted average of each label.
[b] Results of these columns with these methods are affected by the range of the considered parameters.
[c] Not applicable since the elbow curve structure is not formed in the identified range.
[d] K for this approach is manually set to the number of classes since K needs to be known in advance.

To provide a more accurate context for comparing the clustering outcome, two important factors of clustering quality were taken into account: (1) the ratio of the generated clusters to the number of class labels and (2) the ratio of the class labels retrieved after clustering. The former indicator shows how close the number of clusters is to the number of ground truth labels. Ideally, this value is equal to 1 when all the ground truth observations of each class are contained in one distinct cluster. The latter indicator denotes the percentage of class labels that possess a separate cluster after forming the confusion matrix. A value of 1 indicates the ideal case. However, the similarity between different classes and their significant unbalanced distribution can reduce this value (e.g., in a case where instances of a very small class are put in a cluster that also contains a ratio of a very large class, the majority vote selects the larger class). *Figure 4-14* presents the variation of

these indicators versus F-measure for power datasets only. Each subplot in *Figure 4-14* represents one of the variations of the proposed heuristic method. As shown in *Figure 4-14* (a), IES with *local scaling* maintains a better balance between the 1st indicator and the performance. On the other hand, IES with *global scaling* results in better performance for all the cases at the cost of generating a larger number of clusters. Regardless of the number of clusters, the application of PCA for estimation of the global scaling factor results in improved performance. Considering the 2nd indicator, as shown in *Figure 4-14* (b), IES with *global scaling* outperforms in recalling class labels with high F-measure values (three out of four cases), which could be interpreted as the ability to retrieve natural patterns.



(a)



(b)

Figure 4-14. Variation of cluster quality indicators versus F-measure for (a) generated number of clusters, (b) ability for retrieving natural patterns. For the legend of this plot, IESG denotes IES with global scale (leaf node clusters); ESL denotes one step IES with local scale; IESL denotes IES with local scale (leaf node clusters).

In order to provide insight on the computational cost of these techniques, *Figure 4-15* presents the run-time for different methods. All the analyses were carried out through MATLAB implementation. As shown in this figure, eigengap search (only one iteration) with local scaling is the most computationally effective approach but it sacrifices the efficacy of results (as discussed from *Figure 4-14* (a) and (b) and *Table 4-3*). As expected, IES with *local scaling* generally takes more time compared to IES with *global scaling*. The result of the comparable self-tuning approach

[162] is excluded here to avoid bias since the publically available code was partially implemented with C++, which is known to be more efficient compared to MATLAB. However, since it considers a range of number for clustering as the post-processing step, the results are more computationally expensive unless a narrow range based on the knowledge of the domain is selected.

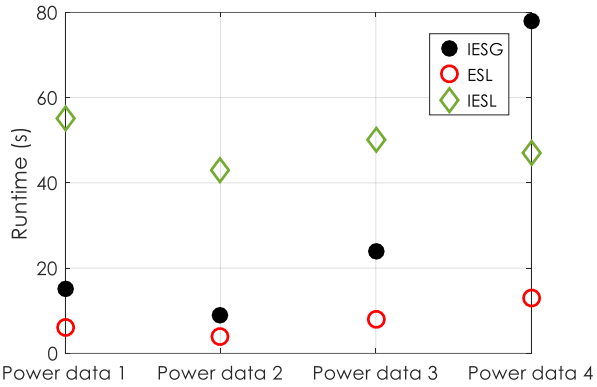


Figure 4-15. Comparison of analysis runtime.

4.5. Conclusion

We have proposed and evaluated an Iterative Eigengap Search (IES) heuristic for automated spectral clustering of multi-scale and higher dimensional feature spaces. The proposed heuristic does not require *a priori* assumption for the number of clusters (K) or scaling parameter of affinity measures (σ) including a range of values for the number of clusters. The algorithm iteratively searches for eigengaps at different scales of the feature space along a tree structure to partition and refine generated clusters with eigengap heuristic. The scaling parameter is estimated through data-driven methods using (1) a PCA-based global scaling factor or (2) using a local-scaling factor that quantifies local scales by measuring the distance of each data point with its nearest neighbors. The scaling parameters are updated at each node of the tree to reveal the dissimilarities in the local structure of a feature space. We have evaluated the performance of the proposed heuristic on several real-world appliance signature power datasets with multiple classes. The datasets are of higher dimensions with multi-scale and heterogeneous nature. The performance of the IES has been compared against several well-known fundamental spectral clustering methods and an internal validation approach that seeks to minimize the dispersion of clustering outcome. The performance assessments showed that the IES heuristic outperforms comparable approaches in

terms of accuracy (an average of 90% for most of the evaluated cases) and capability of finding (recovering) natural partitions in a feature space.

Chapter 5: Self-Configuring Event Detection in Electricity Monitoring for Human-Building Interaction

Afzalan, Milad, Farrokh Jazizadeh, and Jue Wang. "Self-configuring event detection in electricity monitoring for human-building interaction." *Energy and Buildings* 187 (2019): 95-109. DOI: https://doi.org/10.1016/j.enbuild.2019.01.036

Abstract

Monitoring the temporal changes in the operational states of appliances is a key step in inferring the dynamics of operations in smart homes. This information could be leveraged in a variety of energy management applications including energy breakdown of individual loads, inferring the occupancy patterns, and associating the energy use to occupants' activities. The operational states of appliances could be identified through detecting and classifying the events on power time-series. Despite the advancements in the field of event detection, they often require in-situ configuration of model parameters to achieve a higher level of performance according to each new context. In order to address such limitation, in this paper, we have proposed a self-configuring event detection framework for detecting the changes in operational states of appliances. The framework seeks to autonomously learn the contextual characteristics of the loads from the environment and adapt the event detection parameters. The proposed unsupervised framework couples an automated clustering for identifying the recurring motifs, which are representations of the appliances' transient power draw signatures in a given environment and a proximity-based motif matching for detecting the events. The framework was evaluated on EMBED dataset, a publicly available fully labeled electricity disaggregation dataset, collected from three apartments with different categories of the appliances. The evaluations demonstrate that the proposed event detection framework outperforms the conventional event detection in detecting the operational states of different classes of loads across different environments. The proposed framework could also facilitate human-building interactions in training smart home applications by populating motifs to infer the operations of appliances and activities of occupants.

5.1. Introduction

Buildings account for 74% of total electricity consumption in the US with a share of more than half for residential buildings [199]. Therefore, enabling efficient consumption of the electricity in buildings remains as a major sustainability objective. Understanding the energy consumption of

appliances and their operational schedule in a building and the interactions between occupants and the appliances could bring about a number of advantages that pave the way for achieving sustainability goals in the form of both demand-side and demand-response energy management [31, 200]. Among these goals, one could point to providing detailed energy information to occupants for increased awareness (e.g., [27-33]), characterizing the energy impact of occupants' activities (e.g., cooking a meal or adjusting the air conditioning setting), understanding the habitual patterns of occupants in use of appliances for smart and autonomous operations [201, 202], inferring the occupancy of building units for occupancy-driven energy management [203-208], and enabling utilities to identify and target critical loads for grid reliability at high peak demand time [83, 209]. At the center of all these applications is the understanding of load dynamics in buildings that reflect the individual appliance operations. Individual load operations could be monitored through direct sensing of individual appliances. However, in pursuit of scalable appliance-level analyses, electricity disaggregation [128, 210-212], which relies on data measurement at the aggregate level at one sensing node, could be used as a cost-effective alternative to break down the aggregate load into end-use individual loads.

By inferring when an appliance changes its operational state, which is interpreted as an event on the aggregate power time-series, we could potentially identify its energy consumption, the user interaction with the appliance, and the activities of users (inferred from interacting with a series of appliances). Therefore, identifying the events on power time-series comprises an important step in characterizing the energy performance of individual loads and human-appliance interactions. In this context, an *event* denotes a change of the appliance operational state, which could be the on/off switching (e.g., in case of a lighting load) or an operational state-transition (e.g., different cycles of a washing machine). Assuming a time-series signal $P(t)$, collected at the aggregate level, an *event detector* aims to identify the timestamps associated with the change points $T = \{t_1, t_2, \dots, t_N\}$. These timestamps will be mapped to a set of labels (corresponding to the classes that represent appliance names) through a training step. This information could be further analyzed by pattern recognition methods for processing the data into either energy breakdown estimation of appliances [213, 214], inferring the trend of household's appliance use to determine the drivers of consumption and predicting future demand [116, 215-217], or inferring the activities of occupants [64, 218, 219].

Depending on the resolution of electricity consumption data, different algorithms could be used for event detection. Increasing the resolution of the data could help improve the accuracy of the classification algorithms [210] in inferring load identities. This improvement is due to the information gain from the presence of transient data (i.e., a momentary increase in the power before reaching a steady state of power draw), which reveals more information on the dynamics of the loads [211] (Figure *5-1* illustrates an example power time-series with trainset power draw data). However, increasing the resolution brings about an increase in noise interference, which results in challenges for event detectors, such as the increase in false positive detection [220]. To tackle the challenges in detecting events on high-resolution data, advanced event detection techniques were adopted and devised in the literature (e.g., [221-224]). As a common feature, these algorithms rely on a number of algorithm parameters that drive the detection statistics and the performance and thus need to be configured for different environments through a training (i.e., tuning) process. Although the configuration could be carried out through experimental efforts and across different environments, the diversity in the appliances' technologies could pose a challenge in achieving scalable performance in different environments. Therefore, there is often a need for reconfiguration of the algorithms to ensure that the better set of parameters for each new environment is set.

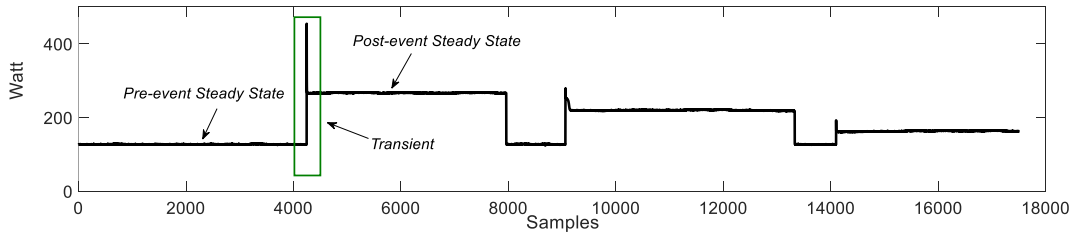


Figure 5-1. Sample real power time series with a 60Hz resolution that shows the information gain from transient power draw

To contribute to the scalability of the electricity disaggregation solutions, in this study, we have proposed a framework to move towards self-configuring event detection techniques. The proposed framework is centered on enabling event detectors to learn from the data in an environment and adapt to the characteristics of the environment and its unique loads. The framework is built on populating an initial dataset for learning and leveraging the recurring motifs, reflected in the shapes of appliances' transient power draw. In its high-level concept, the framework consists of the following steps:

- Populating an initial dataset of appliances' signatures (i.e., vectors representing transient power draws) from a buffered power time series by utilizing a conventional event detector (with generalized configured parameters).
- Characterizing the recurring motifs by using an unsupervised automated clustering on detected events in the buffered data. The extracted motifs in the library represent the characteristics of appliances' signatures in the environment.
- Shifting the event detector to a motif-based matching method: events will be identified, where the shape of the power time series matches one of the motifs. Motif-matching step involves an outlier detection in comparing the power-time series shape with the representative motifs in the library.

Therefore, this study contributes to the body of literature by proposing an event detection framework that seeks to automatically learn the algorithm parameters based on the context of an environment to avoid parameter-tuning in each new environment. Due to the nature of the motif detection step, which employs an automated clustering (without assuming the number of clusters), the proposed framework could also act as a classification step (mapping the events to their associated appliances). Moreover, our evaluation of the framework over the EMBED dataset [125], which is the most comprehensive labeled dataset with three building units, puts the analysis among the most comprehensive assessments.

The rest of the paper is structured as follows. In section 5.2, we presented a literature review of the related studies. In section 5.3, we presented the framework and discussed its components, following by section 5.4, in which the results of evaluations have been presented. The paper was concluded in section 5.5.

5.2. Background and Related Works

Disaggregation methodologies have been applied in different capacities with a focus on residential buildings (e.g., [211, 225]) or commercial and office settings (e.g., [226-228]). Several recent efforts have focused on improved feature selection for appliance classification [229], employing deep learning methods [123, 124], using alternative metrics such as current or voltage instead of the commonly used power metric for load monitoring [124, 230], and leveraging the impact of device interaction (mutual operational status) for improving the disaggregation accuracy [231, 232]. employing interactive visualization and user-interaction (expert knowledge) to interpret

disaggregation results [233]. Here, our focus is on the event detection methodologies. Event detection is a commonly used approach in the analysis of time-series data and it has applications in various domains. Similarly, event detection has been used for the analysis of the power time-series (or similar electricity consumption metrics). The approach has been often adopted in the electricity disaggregation, also known as non-intrusive load monitoring, specifically in a category of efforts, described as event-based methods. In this category, the abrupt changes on the time series are identified as events that are associated with changes in the operational states of appliances. In early studies on electricity disaggregation, the efforts were focused on developing heuristics for detecting the events of on/off appliances. Hart [234], in his seminal research, proposed a detector for on/off transitions by segmenting the normalized power values into steady or changing states. In this heuristic, using low-resolution power data, the steady states were identified by measuring the power variations against a threshold for a pre-defined number of samples on the power time series (e.g., three sample points). The segments of the time series that violate this rule are considered to be changes (i.e., events). This approach has been also used in the recent state-of-the-art disaggregation studies (e.g., [235]). In another heuristic method [236], separate profiles for different appliances were collected as the training and a set of rules were defined to find the on/off transitions by comparison with the pre-defined power ranges. The approach called for profile initialization of the appliances in-the-field.

In order to benefit from the data reflected in the transient power draws, research studies shifted to use the data from the high-resolution power time series. Increasing the resolution of the power data poses more challenges to event detection due to the presence of noise. In order to address these challenges, improved event detection algorithms were proposed. The application of the generalized likelihood ratio (GLR) test was introduced by Luo et al. [221] to improve the performance. In this method, events are detected by measuring whether samples before and after an event are coming from two different Gaussian distributions. The algorithm calls for configuring several parameters including two window sizes to calculate the parameters of probability distribution functions before and after each event and a threshold for likelihood ratio to compare against the detection statistics. Variations of the GLR algorithm for enhanced performance has been also proposed (e.g., [129]), which call for more parameters and thus configurations. Other research efforts have also introduced variants of event detection algorithms for high-resolution data to improve detection accuracy. Some of these efforts are based on the goodness-of-fit χ^2 test-based algorithm on power

time series [133, 237], Window-with-Margins event detection method [238], and adaptive event detection [239] that detects the time limits of each transition interval. Similar to the GLR algorithm, these methods also call for a number of parameters that need to be configured for high-performance event detection. Alternative methods of event detection for alternative electricity measurement metrics have been also proposed. For example, high-resolution voltage time series were used by Patel, Robertson, Kientz, Reynolds and Abowd [201] for detecting events associated to switch on/off or changes of the cycle in appliances. They proposed a specialized event detection algorithm that uses a threshold for identification of the events on the voltage noise time series which required data acquisition systems with very high sampling rates.

Since extensive parameter-tuning or training imposes a barrier to the wide adoption of disaggregation technologies, a few recent studies have focused on unsupervised approaches [240, 241]. Wild et al. [240] proposed an unsupervised event detector based on the kernel Fisher discriminant analysis (KFDA) using current harmonics. The event detector requires two sliding windows with the predetermined length to calculate the test statistics and a bandwidth parameter for the Kernel function. Similarly, in [241], an event detector based on a two-step clustering from graph signal processing has been proposed. The approach collects the subsequent sampling points with power measurement difference above a threshold and then applies adaptive thresholds to refine the clusters until all of them have a coefficient of variation below a specific range for the quality control.

These aforementioned event detection methods either require a training process ([201, 236, 237, 242]) or a parameter-tuning ([129, 221, 234, 243, 244]) step, which calls for the reconfiguration of parameters for new environments and therefore pose a challenge on generalizability and wide adoption. In order to tackle such limitations, in this study, we have sought to propose an event detection approach that leverages the recurring appliance signatures, obtained from an environment to adapt to the characteristics of each unique environment. The approach allows the system to automatically identify detection parameters according to the characteristics of the new deployed environment and could potentially facilitate appliance event labeling through reduced user-system interaction.

5.3. Self-Configuring Event Detection Framework

The proposed framework seeks to shift the detection logic from searching for abrupt changes on a time series to a motif-based detection approach. Therefore, the framework utilizes a search for the most probable transient (power draw) shapes on a time-series that are associated with appliances' state transitions, which we refer to as *motifs*. The concept of using motifs has received attention in the time-series data mining domain (e.g., [245]). In this work, we leverage and deploy the recurring motifs (i.e., time-series subsequences) that represent a collection of signatures from a specific operational state of a given appliance. In doing so, the framework uses the processed power time-series as the representative metric of aggregate electricity consumption. Figure *5-2* illustrates the components and process map of the framework. There are two underlying steps: (1) self-training stage for motif processing on the buffered data, and (2) the proximity-based event detection stage. In what follows, the descriptions for different components have been provided.

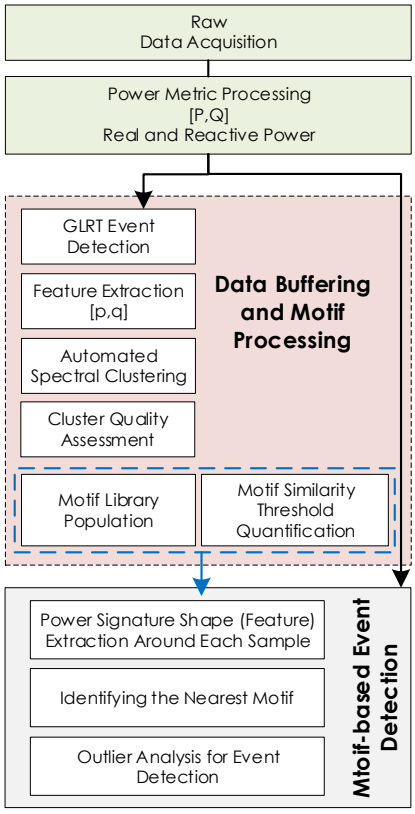


Figure 5-2. The framework for the motif-based event detection

5.3.1. Data buffering and motif processing

5.3.1.a. Conventional event detection for initial learning

The first step of this framework focuses on identifying the recurring appliance signature motifs in a given environment. Considering that the motifs represent the changes of the appliances' operational states, the number of these motifs is significantly lower compared to those of steady-state segments. Therefore, the framework leverages a conventional event detector to identify the occurrence of transient events that represent the potential motifs and prepare the dataset for contextual learning. As the nature of motifs implies, this framework leverages the information in the transient state in high resolution data. Although any event detection algorithm for high-resolution data could be integrated into this framework, in this study, we have utilized the GLR event detector [221] as a common method, which uses a statistical test to identify events. The algorithm evaluates likelihood ratio (detection statistics) between Gaussian distributions ($S{\sim}N(\mu,\sigma^2)$S~N(_, ^{2})), assigned to samples of data before and after each data point on the power time-series:

$$L(n) = \ln \frac{P\left(s_i \middle| \mu_1, \sigma_1^2\right)}{P\left(s_i \middle| \mu_0, \sigma_0^2\right)} \quad (1)$$

where s_i is the *i*-th signal sample point, and μ_0, σ_0^2, μ_1, and σ_1^2 are mean and standard deviation in two windows before and after each data point. Time-series sample points with a likelihood ratio higher than a predefined threshold are marked as events. Accordingly, the algorithm calls for configuring the size of two windows before and after each sample point for estimating parameters of the distributions and a threshold for event detection. In some implementations of this algorithm [129], additional mechanisms for reduced false positives have been used. These mechanisms in turn call for configuring a few more parameters.

In the proposed framework, the conventional event detector will be used without specific efforts on tuning these parameters. In other words, we use parameters from the literature that are the outcome of a configuration process for a specific dataset. Therefore, the event detector, in this step, is prone to identifying wrong events (false positives), or missing events (false negatives). Accordingly, the initially collected events from the environment include a mix of both correct and wrong detections. Nonetheless, the impact of such inaccurate detections is tackled by the motif-mining approach, which has been described in the upcoming sub-sections.

5.3.1.b. Feature Extraction

Upon identifying the initial events, a sequence of data samples that surround an event is extracted to form the feature vectors ($\boldsymbol{fv}$) representing each appliance signature motif. Two windows of pre-event samples ($w_{pre}\mathrm{w_{pre}}$) and post-event samples ($w_{post}\mathrm{w_{post}}$) are used to extract the features. Different harmonic components of real and reactive power time series could be used in the feature extraction stage. In this implementation of the framework, we have focused on the fundamental frequency component of real and reactive power time series as the representation signature motifs in the vicinity of the events:

$$\boldsymbol{fv} = \{\boldsymbol{p}_1, \boldsymbol{q}_1\} \tag{2}$$

where p_1 pand q_1 qare real and reactive power (fundamental frequency) components, respectively. Figure *5-3* illustrates some examples of transient power signatures (the real component only) representing changes in appliances' operational state. These feature vectors are normalized to have a value of zero at the point of the event so that the aggregate nature of the power time-series does not affect the comparison between two vectors.

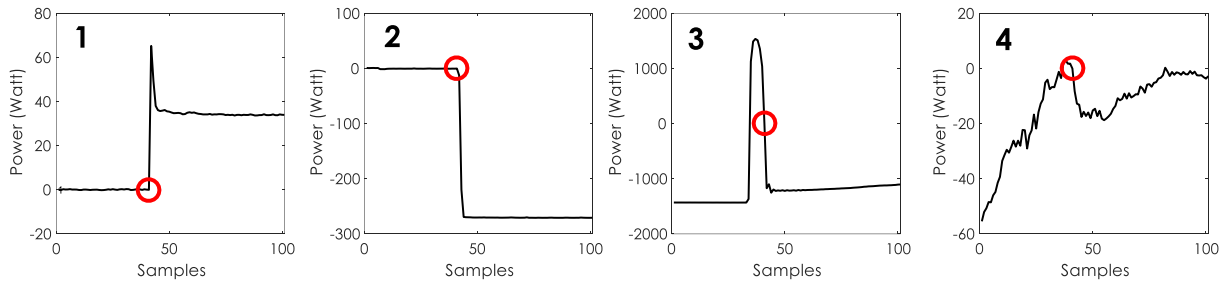


Figure 5-3. Examples of extracted features (the real power component) on the buffered data

5.3.2. Identifying Recurring Motifs through Automated Clustering

The next step includes self-learning of the recurring motifs for enhanced event detection. A critical component to achieve this goal includes automated clustering of the appliance signatures to infer the motifs. Although clustering algorithms are categorized under the unsupervised learning techniques, they commonly call for *a priori* parameters. For instance, K-Means clustering, hierarchical clustering, and mean-shift clustering require different parameters including the number of clusters, a threshold for pruning the tree, or the kernel bandwidth, respectively. Although these values could be set using the domain knowledge, the use of algorithms that need additional hyperparameters contradicts the objective of self-configuration. Therefore, we have developed autonomous clustering algorithms that obviate the need for input parameters (e.g.,

[130]). In this study, we have adopted our proposed heuristic for automated spectral clustering with application to electricity disaggregation [40]. The spectral clustering algorithm uses the eigenvalues of a Laplacian matrix from the data for dimensionality reduction and clustering in fewer dimensions [246]. Assuming a data set $X = [x_1, x_2, x_3 \dots, x_n] \in R^{n \times m}$ in which n is the number of data points (i.e., all feature vectors) and *m* is the number of features, a similarity matrix is defined as:

$$A_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) & i \neq j \\ 0 & i = j \end{cases} \quad (3)$$

In which σ^2 is the scaling factor. Through creating a diagonal matrix with the summation of all the elements on the *i*-th row of A as D_{ii}, a Laplacian matrix is defined as:

$$L = D^{-\frac{1}{2}} A D^{\frac{1}{2}} \quad (4)$$

Assuming *K* cluster, spectral clustering uses the top *K* eigenvalues of *L* and performs clustering on the associated normalized eigenvectors of *L* in a lower dimensional space using k-means approach. The clustering outcome commonly calls for two input parameters: (1) the number of clusters and (2) a scaling factor that depends on the context of the analysis. In order to enable automated clustering, we have proposed heuristics to identify these two parameters. For the former, we have introduced the concept of *iterative eigengap search* (IES) that partitions a feature space using a search tree structure to reveal eigengaps at different scales of the feature space. In order to learn the scaling factor from the data itself, we have proposed a principal component analysis- (PCA-) based quantification of scaling factor at different scales of the feature space. Scaling factor in this context is a parameter that controls the width of the neighborhood [168], and the value of σ defines the reference distance between connected data points within the same scale. In other words, σ is a contextual reference distance that defines if two data points should be considered similar. The outcome of the clustering process is the groups of feature vectors that represent similar operational states of appliances in the target environment. Since collecting the cluster library is a core part of the framework, we have briefly discussed the methodology in this section (more details on the proposed automated clustering could be found in [40]).

Iterative Eigengap Search (IES) Heuristic: Selecting the right number of clusters (*K*) is challenging and subjective in many application domains. In the case of load monitoring, *K* corresponds to unique and observable appliance state transitions (i.e., appliance signatures), in a

given environment, for which the number of appliances as well as the number of transition states for finite state machines is not known. In spectral clustering, eigengap, a well-known heuristic, could be used for automated evaluation of the number of clusters (K) [168, 191]. Let λ_i be the eigenvalues of a Laplacian matrix (calculated based on pairwise similarity distance of feature vectors in a given dataset). Eigengap δ_i is estimated through:

$$\delta_i = |\lambda_i - \lambda_{i+1}|, i = 1, \dots, n-1 \quad (5)$$

where n is the total number of feature vectors. The number of clusters (K) can be estimated by:

$$K = argmax_i(\delta_i) \quad (6)$$

In Eq. 6, the aim is to select the largest gap between *k*-th and (*k+1*)-th eigenvalues derived from the Laplacian matrix for the selection of the number of cluster. The eigengap heuristic justification has been outlined in the literature based on perturbation theory and spectral graph theory [168, 247]. Although eigengap is a measure for estimating the number of clusters, it typically works well in datasets with well-separated feature spaces [168]. However, for the appliance signature features with a multi-scale and noisy nature, the resultant number of clusters is not accurate. Specifically, the differences between the signatures of appliances with smaller transient power draws are masked by signature with larger power draws resulting in clustering the signatures from different appliances and states into one cluster. Therefore, we proposed the *Iterative Eigengap Search (IES)* to search for the eigengap at different scales of the feature space and overcome the challenges of scale effect in dissimilarity quantification.

The IES uses a recursive search on a search tree structure as schematically illustrated in Figure 5-4. At the first step, the entire dataset of feature vectors (root node in Figure 5-4) is processed by spectral clustering (NJW algorithm [246]) and eigengap heuristic. The process continues recursively by clustering the data at each node of the tree. This process is continued until the eigengap cannot further segregate the data in any of the leaf nodes (i.e., K is estimated as 1 based on Eq. 6). Once the clustering of all nodes is completed, the clusters at the leaf nodes will constitute the final clusters. As noted in developing the similarity matrix, a scaling parameter is required to identify the boundaries of the similarity neighborhood. In order to estimate the scale of neighborhoods in an automated manner, we employed the PCA, which deploys orthogonal transformation to map the original features into a new space with uncorrelated variables. Given that PCA sorts the transformed variables based on the maximum variance within the data, we employed PCA to account for the most variance from the major principal axes as the estimation

of σ^2. This will enable us to define an approximated boundary threshold for the formation of the similarity matrix and to distinguish the similar and dissimilar data points according to their distribution at each generated node of the tree. Therefore, the value of the scaling parameter will be different at each node and will be updated to consider the similarity neighborhood of that specific subset of data points. By using this approach, although the eigengap might not initially predict the right number of clusters, it can facilitate segregating the appliance signatures into different clusters that (likely) reside in different scales, which are characterized by different power draws.

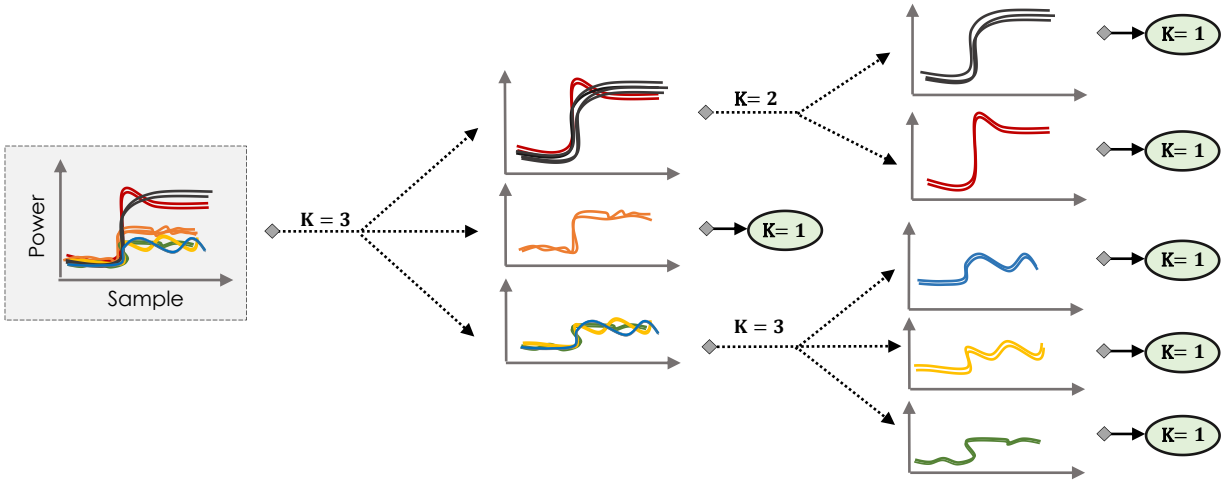


Figure 5-4. The visualization of the IES for appliance signature clustering: The root nodes include the entire dataset (with 6 clusters) which is iteratively partitioned with eigengap. The leaf nodes (K=1) are accepted as the final clusters.

Cluster quality assessment: the task of clustering aims at obtaining high intra-cluster similarity (dense clusters) and low-inter cluster similarity (well-separated clusters). However, a number of challenges may affect the quality of appliance signature clusters which in turn may affect the event detection performance. First, due to the uncertainty in appliance signature characteristics (e.g., power draws) and the inherent challenges involved in clustering, some of the clusters may contain signatures that are distant from each other. Second, as noted earlier, the conventional event detectors are prone to detect false positives, whose feature vectors will be processed during the clustering. Accordingly, it is required to mitigate the impact of such undesired effects by examining the quality of clusters before extracting the recurring motifs. Without such consideration, some of the motifs might not represent an appliance state change in the environment. To this end, we have used a dispersion measure for cluster quality assessment [248].

Assuming m feature vectors in a cluster, the dispersion measure for cluster k (DM_k) is calculated as follows:

$$DM_k = \max dist(\boldsymbol{fv_i}, \boldsymbol{fv_j}) , i, j \in 1{:}\, m \quad (7)$$

where $dist(\boldsymbol{fv}_i, \boldsymbol{fv}_j)$ is the pairwise Euclidean distance between $\boldsymbol{fv}_i$ and $\boldsymbol{fv}_j$. According to our empirical observations, a cluster might less likely represent a recurring motif if the following condition holds [248]:

$$DM_k > S_k + \sigma_k \quad (8)$$

in which S_k and σ_k are the mean and standard deviation of the $dist(\boldsymbol{fv}_i, \boldsymbol{fv}_j)$ across all the feature vectors in cluster k. In evaluating the cluster quality, feature vectors were normalized by dividing each element by the absolute maximum value in each feature vector. Accordingly, clusters which satisfy Eq. 8 are eliminated from the motif extraction step. Other variations of Eq. 8 can be employed ($DM_k > S_k + t * \sigma_k$). However, the adjustment of t could either result in accepting more clusters, some of which might be affected by cluttered observations that do not reflect a real appliance state change (in case of t>1) or reducing the number of clusters (in case of t<1). Therefore, through empirical assessment, we used Eq. 8 based on its efficacy in selecting the clusters with useful information.

Motif library population: to extract the motifs that represent an appliance state transition, we have used the centroid vector (mean of the feature vectors) within each cluster (M_k). These motifs (M_k) play an important role in characterizing activities, human-appliance interactions, and energy consumption in an environment. The number of the motifs depends on the number of appliances, the complexity of their operational states, and the timeline of the initial data buffering. Figure *5-5* illustrates examples of several extracted motifs on a sample dataset. The corresponding appliance label for each motif has been provided as well.

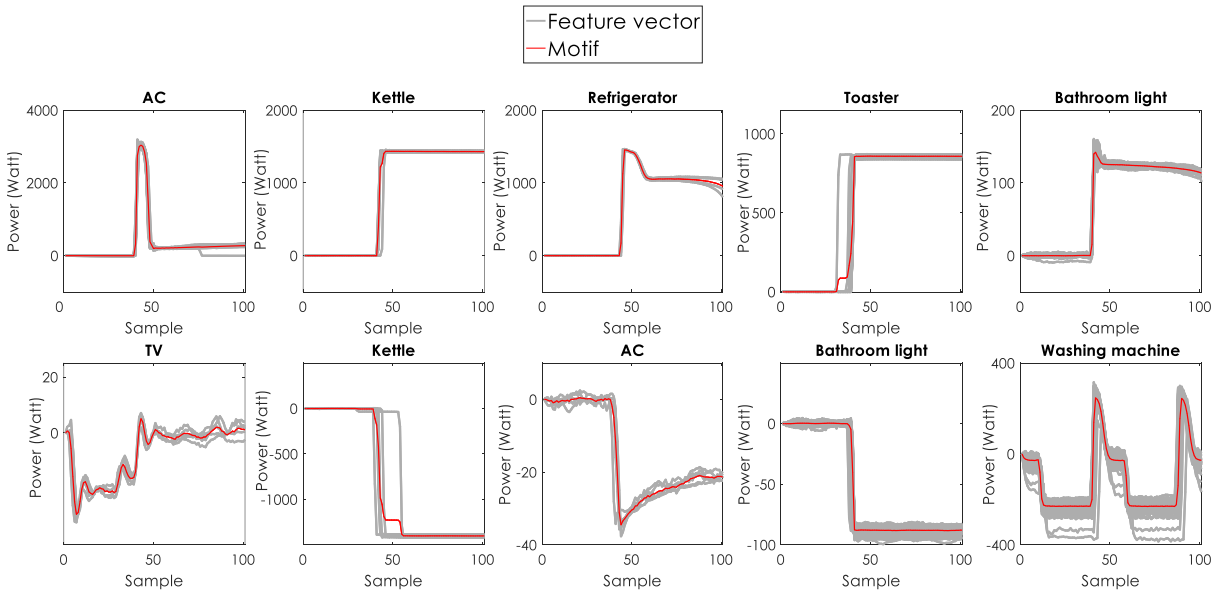


Figure 5-5. Examples of feature vectors, clusters, and their associated motifs (only real power component was shown). Each feature vector contains 100 samples (corresponding to a duration of ~1.7s).

Given that the motif identification process could be carried out over a few days and get updated on a regular basis, the potential false negatives from the conventional event detectors will not affect the motif extraction as numerous events representing each motif are often observed in an environment.

5.3.3. Proximity-based event detector

As the description of the framework implies, this approach uses a proximity-based technique such as nearest neighbor for motif-based event detection. Therefore, the approach is a semi-supervised classification problem. However, to control the false positives, an outlier detection step is required for accepting or rejecting an event considering that the nearest neighbor classifier will associate every new observation with one of the motifs. Therefore, the motif-based event detection process is as follows:

- For each sample point on the power time-series, extract a feature vector (P_i) following Eq. 2. We call this feature vector the *power signature shape*.
- Identify the closest motif to the power shape using a 1NN algorithm.

– Run an outlier detection to accept or reject the power shape as a viable event.

As the process shows, the proposed framework could combine the process of event detection and classification. Therefore, upon detection of an event, if the clusters are labeled (as shown in Figure *5-5*) with the name of appliances in the physical environment, the load identity is also revealed that in turn could be used for other applications such as human-appliance interaction monitoring, occupancy detection, or energy consumption assessment.

In extracting the power signature shapes, similar contiguous windows of pre-event samples ($w_{pre}\mathrm{w}_{\mathrm{pre}}$) and post-event samples ($w_{post}\mathrm{w}_{\mathrm{post}}$), as used for motif extraction, are employed. By sliding these two windows along the time-series, for each sample point, a representative power signature shape (P_i) is extracted in the form of a feature vector. These feature vectors are normalized to have a zero value at the intersection of the feature extraction windows. The index for the closest motif (k_i^*) to each power signature shape (P_i) is identified using the 1-NN algorithm such that

$$k_i^* = \arg\min \mathrm{dist}(P_i, M_k)\,, k \,\epsilon\, 1{:}\,K \tag{9}$$

Outlier detection: As noted, the objective of this outlier detection is to evaluate a binary hypothesis to decide if P_i is coming from the same distribution that motif k_i^* represents. There exist different methods of outlier detection that could be used for this purpose such as comparing the distance of a new observation from the motifs (e.g., using Mahalanobis distance that accounts for the shape of signatures as well) or two-class support vector classifiers. All these methods either call for a threshold value or training process to be used for hypothesis testing. Processing the motifs through clustering, not only provides the recurring motifs but also provide an insight into the stochastic nature of the power draw at different times. The variations of the feature vectors in each cluster provide the ground for learning the outlier detection thresholds from the environment itself. Given that the clustered feature vectors represent a dense set of data points, the clustered motifs do not include outliers. This fact enables us to identify a range of acceptable thresholds for each cluster to measure the similarity of power shape P_i with the corresponding motif k_i^*. Therefore, we have formulated the event detection in this context as an outlier implementation that uses a distance metric for evaluating the hypothesis. To this end, we have used the following criteria for hypothesis testing and detection statistics. A sample point i is flagged as an event if the following similarity criterion is met:

$$\boldsymbol{D}_i < \mu_{k_i^*} + f\sigma_{k_i^*} \tag{10}$$

where $\boldsymbol{D_i}$ is the distance between the motif (k_i^*) and the power signature shape (P_i); $\mu_{k_i^*} + f\sigma_{k_i^*}$ indicates the *similarity threshold* for each motif and is learned from the content of each cluster; $\mu_{k_i^*}$ and $\sigma_{k_i^*}$ are the mean and standard deviation of the distances between the motif and each feature vector in the cluster k_i^*, and f is the parameter that controls the flexibility of the boundaries in each cluster. In this outlier detection, we have adopted the Frechet distance [249], which is defined as follows:

$$\boldsymbol{D_i} = \boldsymbol{d}\left(P_i, \mu_{k_i^*}\right) = inf\ max_{\alpha,\beta\ t\in[0,1]}\{\| P_i\left(\alpha(t)\right) - \mu_{k_i^*}(\beta(t)\ \|\} \tag{11}$$

In which $\|.\|$ denotes the L_2 norm and α, β: [0,1]→ [0,1] span over all continuous increasing functions. The Frechet distance (FD) measures the similarity between time-series segments by considering the order and location of data points in the shape of the signatures, and therefore suitable in our implementation as it considers the continuity of shapes into account for distance measurement. This will make FD a more effective measure for our purpose compared to the widely used distances such as Euclidean or Mahalanobis distance that employ one-to-one mapping for calculating the distance. The pseudocode of the proposed event detection algorithm is as presented in Figure *5-6*.

For each sample point i,

$P_i = S(i - w_{pre} - 1 : i + w_{post} + 1)$

$M_{k_i^*} \leftarrow 1NN\ (P_i, \boldsymbol{M})$

$\delta_i = |S_{i+1} - S_i|$

$\psi_i^* = FD(M_{k_i^*}, P_i)$

If $\delta_i > \tau_\delta$ and $\psi_i^* < \mu_{k_i^*} + f\sigma_{k_i^*}$

$E_p \leftarrow i$

End

End

Figure 5-6. Pseudo code for the motif-based event detection algorithm

To visually demonstrate the underlying steps involved in the proposed event detector, Figure *5-7* shows the process for one sample point that was detected as an event. Part (a) reflects the data buffering and motif processing to populate the motif cluster library and identify the parameters for

outlier detection for each cluster (self-training process). The outcome of this part is used for the event detector in part (b), applied to all the samples for the power time-series. In this figure, steps I and II indicate the feature extraction on the buffered data. The collected features are passed to clustering (step III). In step IV, the quality of clusters is assessed. The centroids of remaining clusters were collected as motifs (step V). By calculating the statistics (average, μ_k, and standard deviation, σ_k) for observations within each cluster, similarity thresholds (step VI) for each motif is calculated (defined in Eq. 10). In part (b), for each power sample (step VII), a power signature shape (step VIII) is extracted. Using 1NN, the closest motif to the power signature shape is determined. In step (X), the distance of the closest motif (obtained in step IX) is compared with the associated similarity thresholds (from step VI). Since the equation holds in this case, the power shape for sample *i* is considered similar to the motif shown in step (X) and marked as an event. If the motifs were labeled as well, the load identity, i.e., kitchen light turn-on in this case, is revealed as well.

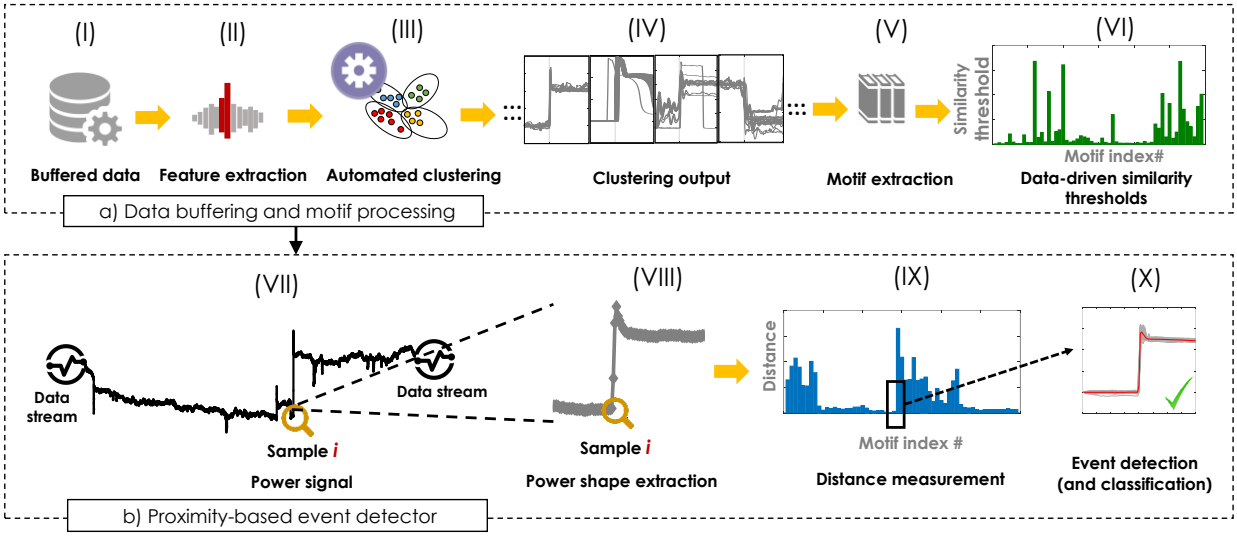


Figure 5-7. Schematic diagram for the appliance self-configuring event detector using real data

5.4. Evaluation and Results

5.4.1. Dataset Description

To evaluate the performance of the proposed approach, we applied the algorithm to a real-world dataset of electricity disaggregation. The EMBED dataset [125] includes the aggregate power time series, collected from three apartments at the main circuit panel of each unit, for different periods

varying from two to four weeks in Los Angeles, CA. The data has been fully labeled by leveraging ground truth sensors (both electricity and light) that were installed at the consumption node, and contains the timestamps as well as the corresponding appliance labels and sub-labels for different operational states of appliances. In our analysis, we have used the processed power time-series (real and reactive) at 60Hz, which enables capturing the transient shapes between steady states. In the US, the electricity infrastructure uses a split-phase system that feeds appliances on two major circuits (i.e., phases). Therefore, in our analysis, we have pointed to performance on different phases (A and B). Table 1 shows the characteristics of appliances operations in different apartments. Details of the data collection and post-processing could be found in Jazizadeh et al. [125] and therefore were not presented here.

Table 5-1. Properties of the EMBED dataset [125]

Dataset	No. of events	No. of appliances	No. of classes*
Apt 1	~4400	16	66
Apt 2	~9100	20	62
Apt 3	~7800	18	68

* No. of classes represent the number of appliance state transitions

5.4.2. Performance evaluation

To characterize and quantify the performance, we have utilized the commonly used evaluation metrics [211, 250] of precision, recall, and F-measure. In order to simulate the data buffering process in forming the benchmark motifs, we have divided the power data into two train and test subsets using a 75-25 ratio, respectively. It is intuitive to have a separate portion of the data for populating the motif library to avoid over-fitting and matching the power signature shapes to a motif that was already extracted from the same section. A pre-event window (W_{pre}) of size 40 (i.e., two third of a second) and a post-window (W_{post}) of size 60 (i.e. one second) were used in feature extraction. In classification studies, it has been observed that the detection performance is not highly affected by the size of the windows. A general rule of thumb in identifying the size is to ensure that signatures are not overlapping and the transient information perseveres. Moreover, considering the fact that the size of the window for both benchmark motif identification and event detection is the same, the window size effect will be minimized. In buffering the data, GLR event detection algorithm was adopted from [129]. Leveraging the assessments in the literature, a general

set of parameters for the GLR algorithm was used [251]. No specific effort was made to ensure that the GLR algorithm is tuned to its better level of performance. The data for the training subset was passed through the clustering algorithm to create the clustered motifs. An empirical value of $f = 50$ was considered for characterizing the thresholds for each cluster in Eq. 10. This value was identified through a sensitivity analysis over different values of f. A vector of $f = \{1, 5, 10, 25, 50, 75, 100\}$ was tested for the data set from Apt 1-Phase A. Figure 5-8 shows the F-score results for different f values.

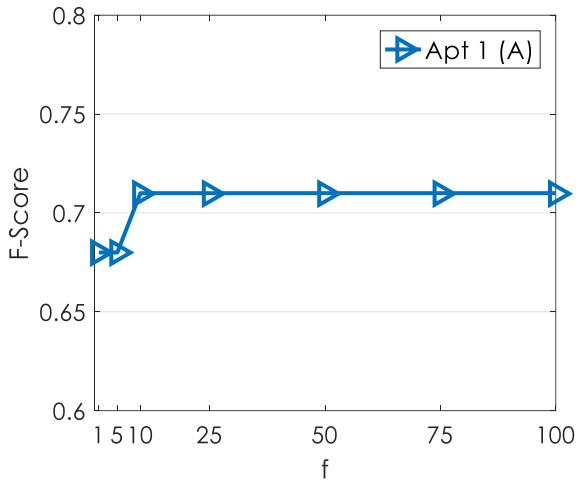


Figure 5-8. Sensitivity analysis for identifying the f value on phase A of Apt 1

5.4.3. Event detector evaluation

The performance of the proposed framework against the conventional GLR algorithm was assessed on the test data subset. The true positives (TP), false positives (FP), false negatives (FN), precision, recall, and F-score were used as the quantified performance metrics. True negatives (TN) are not presented since events are sparse on the power time-series, and the number of events is significantly lower compared to the number of instances on steady-state segments. Therefore, any event detector can achieve a very high number of TN, and the metric is not able to justify the performance. In order to quantify the metrics, a ±3 sample points (less than 0.1 second) was considered as the tolerance in the comparison between ground truth and predicted events. Table 5-2 shows the results of the evaluation. In this table, for the motif-based approach, events for the associated appliance classes that have been operated at least once in the training stage were considered for the evaluation. The higher values (i.e., better performance) for precision, recall, and F-score have been highlighted in bold.

As the results show, the proposed motif-based event detector had a promising performance without the need for parameter-tuning or a priori information for the clustering. Particularity, by using the motif-based event detection, an average F-score of 0.81 across all datasets was obtained that is higher compared to the benchmark GLR performance of 0.67. Moreover, the average value of precision (trade-off between true and false detection), and recall (trade-off between true and wrong detection) is 0.78 (compared to 0.66 for GLR) and 0.86 (compared to 0.76 for GLR), respectively. The motif-based approach maintains a better balance for these metrics, considering the standard deviation across all datasets. Figure 5-9 summarizes the performance metric results.

Table 5-2. Evaluation results for the proposed event detection, compared to GLR

Event Detection Method	Dataset	# of events	TP	FP	FN	Precision	Recall	F-Score
Motif-based (Self-configuring)	Apt 1 (Phase A)	337	229	61	108	**0.79**	**0.68**	**0.73**
	Apt 1 (Phase B)	1075	981	19	94	**0.98**	**0.91**	**0.95**
	Apt 2 (Phase A)	665	659	63	6	**0.91**	**0.99**	**0.95**
	Apt 2 (Phase B)	1016	764	381	252	0.67	0.75	0.71
	Apt 3 (Phase A)	1734	1497	717	237	**0.68**	0.86	0.76
	Apt 3 (Phase B)	304	296	164	8	**0.64**	0.97	**0.77**
	Average					**0.78**	**0.86**	**0.81**
	Standard deviation					0.14	0.12	0.11
GLR	Apt 1 (Phase A)	337	219	106	118	0.67	0.65	0.66
	Apt 1 (Phase B)	1075	529	23	546	0.96	0.49	0.65
	Apt 2 (Phase A)	1128	659	469	469	0.58	0.58	0.58
	Apt 2 (Phase B)	1016	925	423	91	**0.69**	**0.91**	**0.78**
	Apt 3 (Phase A)	1734	1608	881	126	0.65	**0.93**	0.76
	Apt 3 (Phase B)	304	299	451	5	0.40	**0.98**	0.57
	Average					0.66	0.76	0.67
	Standard deviation					0.18	0.21	0.09

As the visual illustrations of the performance metrics, shown in Figure 5-9, the motif-based approach outperforms GLR with higher precisions across different environments. This could be mainly associated with the fact that the motif-based approach reduces the false positives in event detection. However, comparison of the recall values shows an interesting trend. In some cases, the GLR shows a better recall as it is generally more sensitive the changes. However, this increase

comes with a cost of reduced precision. This trade-off is better balanced by the proposed motif-based event detection.

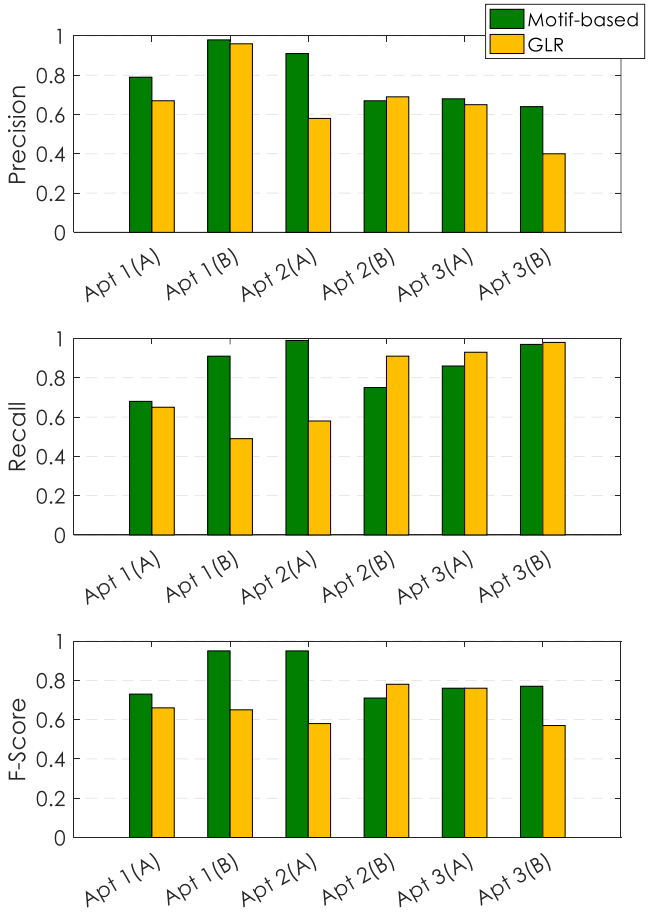
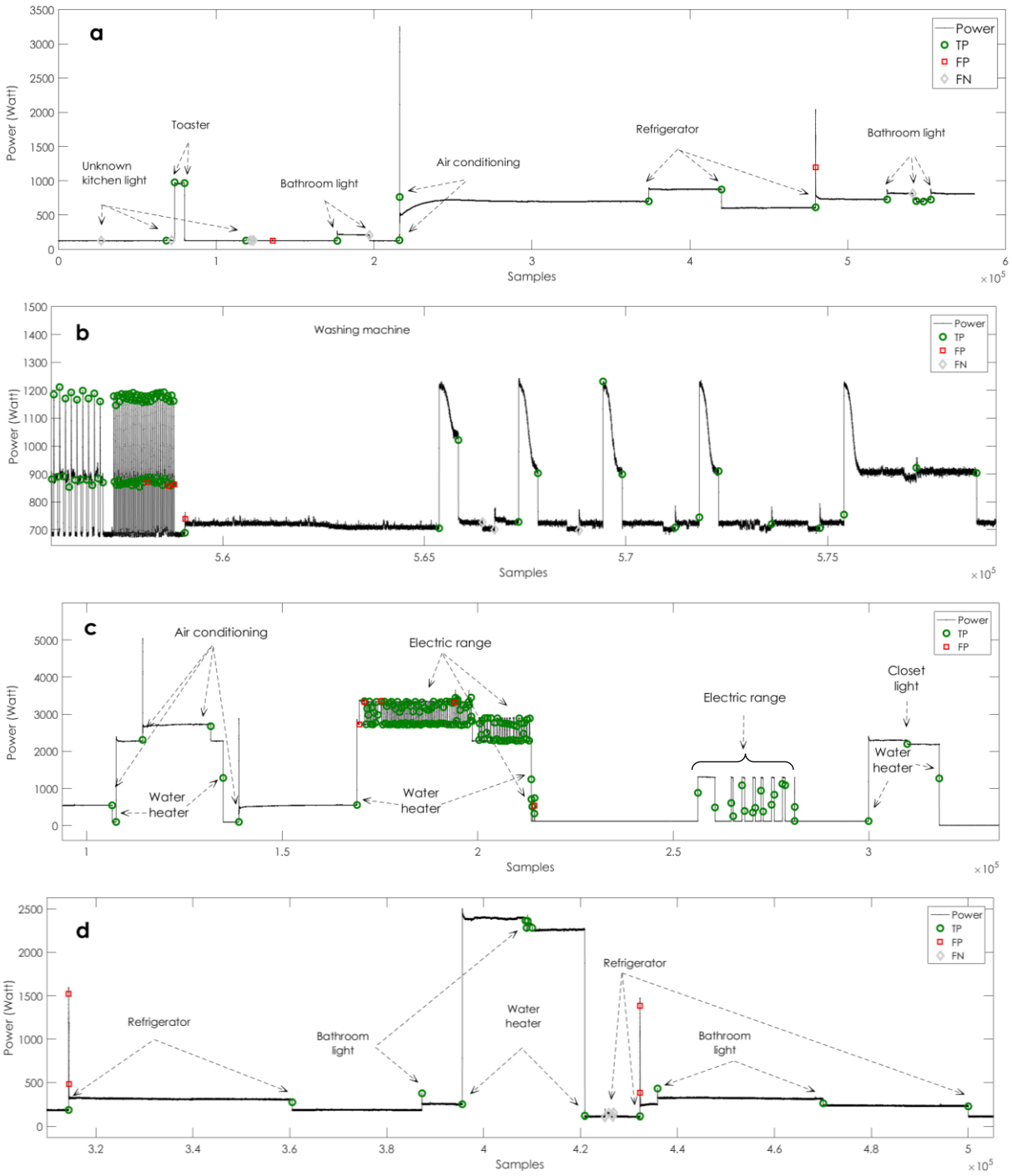


Figure 5-9. Comparison of performance evaluation metrics for motif-based versus GLR event detectors

To provide a better context for the performance of the proposed approach, in Figure 5-10, we have presented visualizations of the detected events on samples of the power time series for a variety of appliances. Short sections of power time-series (less than 3 hours) have been provided for each dataset for visual interpretation. The corresponding motif labels were employed for event classification as well. TP, FP, and FN are depicted based on the ground truth data.


a
Power (Watt)
Samples
×10^5
Power
TP
FP
FN
Unknown kitchen light
Toaster
Bathroom light
Air conditioning
Refrigerator
Bathroom light
b
Washing machine
c
Air conditioning
Electric range
Water heater
Water heater
Electric range
Closet light
Water heater
d
Refrigerator
Bathroom light
Water heater
Refrigerator
Bathroom light

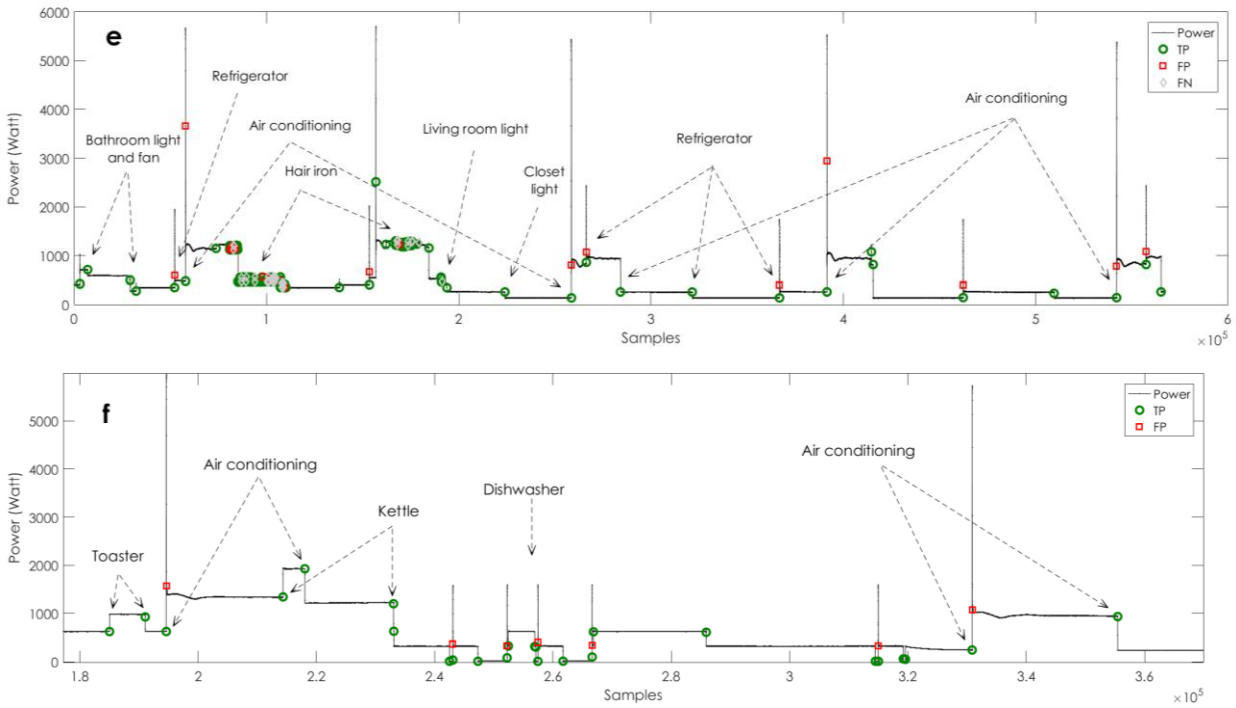


Figure 5-10. Detected events using motif-based approach on samples of power time-series from: (a) Apt 1, phase A, (b) Apt 1, phase B, (c) Apt 2, phase A, (d) Apt 2, phase B, (e) Apt 3, phase A, (f) Apt 3, phase B.

As these graphs and quantitative evaluations show, the motif-based approach has also resulted in a number of false positives and negatives. From this limited observation, it appears that the congested areas (with multiple repeated cycles) could result in higher false negatives. Moreover, it could be seen that the missing events (i.e., false negatives) appear to have lower power draws. Depending on the application domain, the importance of detecting the operational events for appliances might vary. Therefore, we have also evaluated the performance of the motif-based event detection according to the transient power draw values for appliances. In doing so, we have illustrated the trade-off between true detection (TP) and missed detection (FN) for different power ranges in Figure 5-11. As this figure shows, missed detections are mainly attributed to events that have a lower power draw, less than 100 Watts, and only in one case (Apt 1, phase A), some of the very large power draw events (more than 1000 Watts) are not detected. The parameters of the conventional event detector for populating the motif library could play a role in such observations. For example, the GLR could be set to ignore low power variations in consecutive samples to avoid false positives due to noise interference. In such scenarios, the motifs from the low power draw

appliances might not be retrieved and used in motif-based event detection. In this study, we have set the minimum change in power to be equal to 25W.

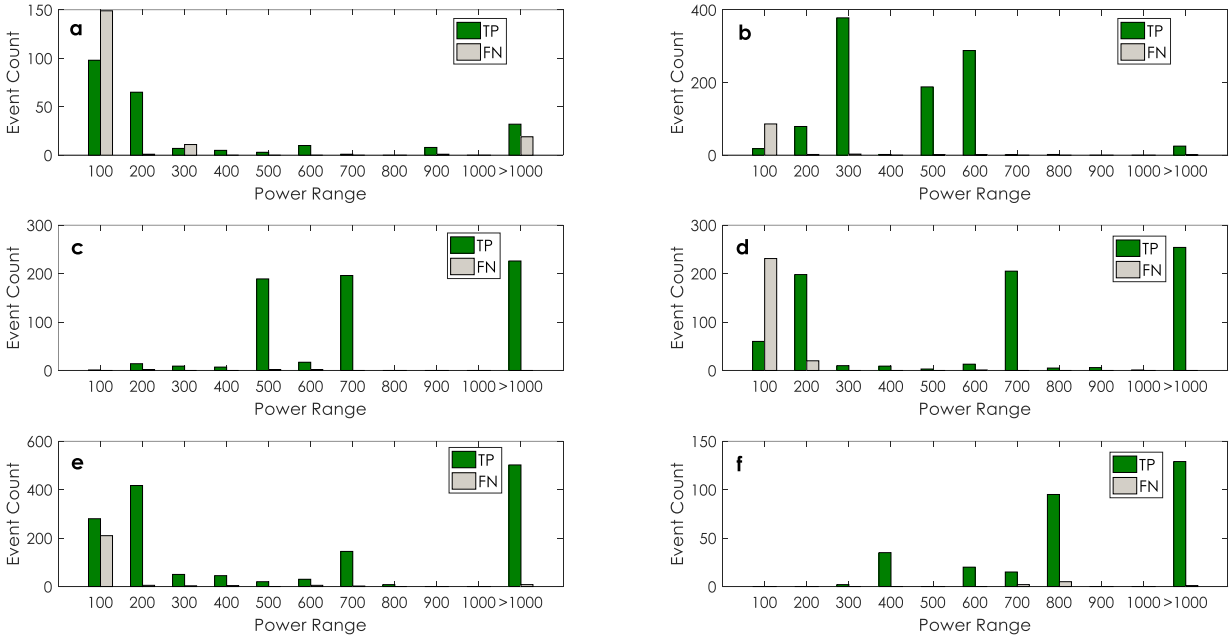


Figure 5-11. Histogram of true positives versus false negatives for the motif-based event detector for different power ranges: (a) Apt 1, phase A, (b) Apt 1, phase B, (c) Apt 2, phase A, (d) Apt 2, phase B, (e) Apt 3, phase A, (f) Apt 3, phase B. The values on the horizontal axis indicate the upper bound for the power range. For example, a value of 100 indicates a power range of 0-100 Watt.

5.4.3.a. Appliance-level evaluation

The evaluation of the performance from the appliance type perspective is important as this knowledge could affect the performance of the end-use applications. Appliances with major power draws have a more significant role in quantifying the energy consumption, and therefore their associated events are of greater importance in that context. For example, air conditioning and heating system events mainly have higher power draws with a longer period of operations, which make them important for energy consumption assessment ([252, 253]). On the other hand, events for appliances with lower power draw play a less important role in energy consumption assessment but they are important in inferring human activities and human-appliance interactions. If low-power events happen frequently and sustain long intervals of operations, their cumulative impact could be also considerable for energy consumption. For example, identifying events for some

appliances such as electric range can be tangibly tied to occupants' activity (e.g., cooking). Specifically, identifying occupants' activity can provide context-aware applications in the buildings. Similarly, fine-grained monitoring of events for several multi-state appliances like the refrigerator or dishwasher can provide an insight for fault detection or demand response opportunities for the residential sector [254].

To provide an insight on the performance for different categories of appliances, we measured the number of true (TP) and missed detection (FN) for each appliance type and presented the results in Table 5-3. To this end, for each label type in the ground truth data, we compared the timestamps between ground truth and predicted events. In this table, the 3-digit labels represent the appliance type according to the EMBED dataset. Some appliances have only turn-on and turn-off events (e.g., lights), while others have different transition states (e.g., Air conditioning (AC) systems).

Table 5-3. Detection rate from appliances' perspective for the test part of the dataset

Dataset	Label	Appliance	# of events	True Positive	False Negative
Apt 1 (Phase A)	111	Refrigerator	65	49	16
	129	TV	96	25	71
	140	Unknown kitchen light	77	1	76
	141	Kitchen light 1	5	5	0
	142	Kitchen fan light	8	8	0
	143	Kitchen light 2	2	2	0
	144	Bathroom light 1	12	12	0
	145	Bathroom light 2	61	60	1
	162	Kettle	9	7	2
	163	Toaster	9	8	1
	180	Air conditioning	38	36	2
	300	Unknown	28	16	12
Apt 1 (Phase B)	120	Laptop	15	2	13
	122	LCD monitor	6	0	6
	180	Air conditioning	27	26	1
	181	Hair dryer	9	8	1
	182	Iron	14	14	0
	183	Washing machine	948	931	17
	300	Unknown	56	0	56
Apt 2 (Phase A)	100	Electric range	217	217	0
	140	Bathroom light	6	6	0
	144	Closet light	21	20	1
	164	Water heater	27	27	0
	180	Air conditioning	394	389	5
Apt 2 (Phase B)	100	Electric range	174	174	0
	111	Refrigerator	455	329	126
	129	TV	63	2	61
	140	Unknown light	46	0	46
	145	Bathroom light	131	122	9
	164	Water heater	34	34	0
	200	Grill	75	75	0
	300	Unknown	38	28	10
Apt 3 (Phase A)	111	Refrigerator	869	810	59
	120	Laptop	12	0	12
	129	TV	16	7	9
	140	Unknown Light	6	2	4
	142	Closet light	10	6	4
	143	Kitchen light	44	42	2
	144	Living room light	19	19	0
	145	Bathroom light and fan	107	98	9
	146	Bedroom lamp	20	15	5
	147	Living room lamp	6	4	2
	166	Hair iron	334	214	120
	180	Air conditioning	277	269	8
	181	Hair dryer	11	11	0
	300	Unknown	3	0	3
Apt 3 (Phase B)	162	Kettle	24	24	0
	163	Toaster	36	36	0
	167	Dishwasher	33	33	0
	180	Air conditioning	211	203	8

Figure 5-12 illustrates the ratio of the correctly detected events for each appliance type (refer to Table 5-3 for label interpretation). As the results show, for apartment 1, appliances like AC, washing machine, bathroom light (which operates a fan as well), iron, and toaster have high detection accuracy, while TV, laptop, and kitchen light, all with low power draw, are hard to detect. For apartment 2, except for TV and the kitchen light, the event detector achieves high accuracy for a variety of appliances like AC, range, water heater, and grill. For apartment 3, except for the laptop that was failed to be recognized, closet light, unknown light, TV, and hair iron showed an average performance. However, the event detector shows a nearly perfect detection rate for the kettle, toasters, dishwasher, AC, dryer, and some of the lights. As can be seen from Figure 5-12, the event detector has a promising performance for a variety of different appliances. However, for the ones with considerably low power draws (less than 100 Watt), the inevitable impact of noise and artifact could be reflected into the shape of appliance signatures, which results in the reduced performance in the clustering procedure (section 5.3.2) or the proximity-based event detector (section 5.3.3).

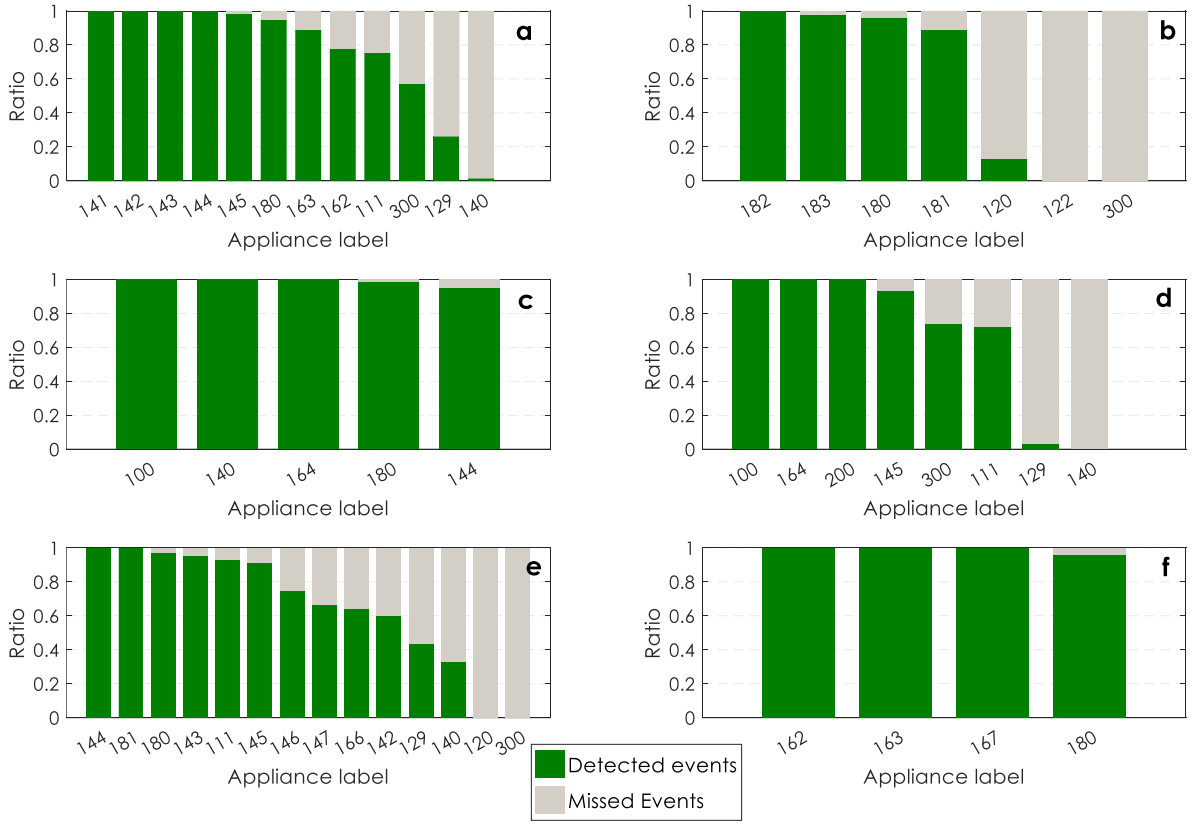


Figure 5-12. The distribution between TP and FN rates for detected events for each appliance type: (a) Apt 1, phase A, (b) Apt 1, phase B, (c) Apt 2, phase A, (d) Apt 2, phase B, (e) Apt 3, phase A, (f) Apt 3, phase B.

In the presented analyses, it was assumed that the event detector was evaluated at a time period when no new appliance was added to the environment. If new appliances (with different signatures) are added to a house, their corresponding characteristics need to be added to the library. To tackle this problem, motif processing could be performed regularly in case new appliances are added to the environment to capture their characteristics. In addition, the detection of the appliance type that causes an event is an important step following the event detection for all applications of human-appliance interaction. This process could be carried out by using a classification algorithm that uses a training dataset for inferring the appliance type. Our proposed approach leverages motifs of appliance signatures from similar sets of observations associated with the generated clusters. Therefore, clusters could be used as the training data set for the aforementioned classifier considering that the clusters are labeled by the actual identity of the appliances. In practice, this could be potentially performed through user interfaces (an example of interactive user interface in similar field of disaggregation can be found in [233]).

We have also evaluated the impact of f on the performance of the algorithm in different environments as illustrated in Figure *5-13*. As shown, a value of $10<f<100$ leads to a relatively consistent performance, and the value of f in the selected range does not have an impact on the performance in different contexts.

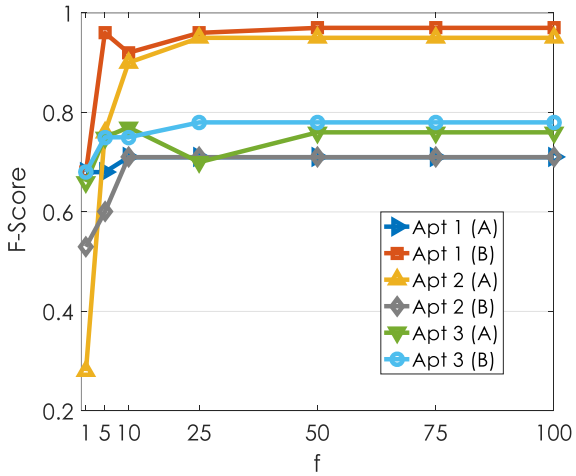


Figure 5-13. Analysis of the impact of f value across different datasets

5.4.4. Limitations

Although the proposed approach has shown an overall promising performance, there are a number of limitations to be addressed: First, the detection rate for some of the low power-draw devices including laptop, TV, and lights was observed to be lower. Figure 5-12 of illustrates these observations. In general, the detection of the operations for low power-draw appliances in the presence of appliances with higher power-draw and in the context of high-resolution electricity signal could be challenging. In fact, developing techniques for detection of miscellaneous electrical loads (MEL) [255], is a topic of research that the Department of Energy has prioritized in the past few years. Second, identifying the status of the appliances is dependent on the detection of on/off events. Similar to other related studies in this domain, missing an on or off event can lead to the wrong estimation of the operational status for different applications. Given that the initial dataset (i.e., buffered data set) is populated through a conventional event-detection algorithm, the outcome of the clustering and the cluster quality analysis process could affect the detection of the events through motif-based event detection. Nonetheless, from the energy perspective, Figure 5-11 of the manuscript shows that for events with higher transient power ranges (>200W), the correct detection rate is high. Third, in our study, it is assumed that after identifying the recurring motifs from the buffered data during the self-training stage, the appliances inside the house remain the same. However, in case of adding a new appliance to the house, it is required to allow time for the motif-based approach to account for the occupant interaction with the new device, in order to learn the contextual information of the newly added load and to update the motif library, accordingly. This also holds true for appliances that have not been used during the data buffering stage. Fourth, in case of having simultaneous events (operational state transitions of different appliances with very close time difference), the shape of the feature vectors will be impacted by the simultaneous events, which can also affect the clustering and the motif-based event detection step. However, considering the fact that the length of the feature vectors was set to a limited duration that preserves the transient shape (less than 2 seconds), the chance of observing a large number of simultaneous events in practice will be low.

5.5. Conclusion

Fine-grained monitoring of operations of the household appliances through power time-series requires the knowledge on the timing of events. However, event detectors typically rely on models

that are configured to a specific environment, in which they are deployed. In this paper, we present the shift from event detection based on abrupt changes in time-series of representative power metrics to a motif-based approach. Motifs are represented by clustered signatures, which represent the transient power draws due to change in operational states of appliances in an environment. Motif-based event detection facilitates the self-configuration of algorithms in a new environment. The realization of the proposed approach calls for data-driven techniques that automatically populate a motif library in an unsupervised manner and use a similarity threshold for comparison between clustered motifs and the new observations. The framework in this study uses a combination of a proposed automated spectral clustering, 1NN classifier, and an outlier detection based on the Frechet distance. The outlier detection autonomously learns its parameters from the clustered signatures in an environment that represent the motifs. The evaluation of the framework power time series on EMBED dataset's three apartments over two weeks demonstrated a promising performance. Overall, the proposed algorithm outperforms the GLR algorithm without the need for detection statistics parameter-tuning in new environments.

As described in this study, the motif-based approach leverages a dense set of observations in the form of a cluster to identify motifs and quantify the event detection thresholds for different clusters. Therefore, the new detected events are associated with a set of observations from the environment that either reflects the automatic change in the operational state of an appliance or a change that is caused by occupants. Accordingly, the framework also provides the ground for more efficient communication with the users in an environment for learning the activities. The framework could be used for facilitated training of a machine learning framework that identifies when different appliances are being used without the need to ask for several inputs for one appliance. Therefore, leveraging the proposed framework for facilitated training of machine learning frameworks that infer the disaggregated energy consumption of appliances or associating the energy consumption to activities of occupants comprise the future direction of this research.

Chapter 6: Quantified investigation of demand and supply balancing among prosumers and consumers: A feasibility assessment for energy trading

Abstract

With the increased adoption of distributed energy resources (DER) and renewables such as solar panels at the building level, consumers turn into prosumers and will supply their own energy demand on site. This will provide peer-to-peer (P2P) energy trading opportunity, in which prosumers offer their surplus energy to consumers. Despite recent attention of P2P research on the residential households and investigating different aspects of virtual and physical layers, empirical analysis and quantified estimation of load balancing potential for energy trading is not well established. Therefore, in this paper, we investigated the load balancing potentials of a community with decentralized DER management systems through a data-driven simulation of varying infrastructure configurations by using real-world data. The concept of load complementarity amongst prosumers and consumers is used for establishing the self-sufficiency of the community. Multiple scenarios by accounting for load profiles uncertainty across households, the PV and battery integration level in the community, and users' flexibility in load operation are investigated to understand the load balancing potential for decentralized distribution. A case study on ~250 residential buildings in Austin, TX, is presented. The findings showed that with a high level of PV integrations (more than 75%), energy trading could result in self-dependency for the entire community during peak generation hours (11am-3pm) while there are limited opportunities during later times after 4pm with PV-standalone systems. As alternatives, it was shown that integrating building level storage and users' flexibility for load shifting during 2-h could improve the self-sufficiently of the community up to 18% and 11%, respectively.

6.1. Introduction

Smart grid rely on novel elements such as Information and Communication Technologies (ICT), metering devices, controllable loads, photovoltaics (PV), and batteries for improved operation of the power system. These elements can revolutionize the energy management flow by moving to decentralized energy distribution and reducing our reliance on centralized distribution systems with fossil fuels. As a result, increasing the utilization of clean and green energy resources will

mitigate the global warming concerns and results in billions of dollars saving in electricity generation [256]. Following this vision, the integration of small-scale distributed energy resource at the household-level has received increasing attention over the last years. Consumers who adopt renewables like PV and incorporate batteries become prosumers. Therefore, they can supply their own energy demand and actively participate in the energy market by selling excessive energy to their neighbors. As such, Peer-to-Peer (P2P) energy trading has been considered as the next-generation approach for supply and demand balancing, in which the collective participation of prosumers and consumers at the household level will have substantial impact for decentralized distribution [257].

To promote the utilization of on-site generation and improving the self-sufficiency, energy policies have been enacted by different countries and states. For example, Germany used to grant a special bonus for self-consumed electricity as defined by its Renewable Energy Act [258]. Similarly, Italy and China have introduced a self-consumption subsidy [259]. Furthermore, European countries (like France and Italy) are imposing modifications to decrease their feed-in-tariff levels [259], and US states consider changing net metering schemes to reduce the rates [260]. In such electricity markets, sending surplus energy back to the grid would be discouraged, and the match of on-site generation and demand will be inherently promoted [261]. In line with this vision, in community-level local energy trading, the goal is to cover the demand of local consumers by the excessive energy of prosumers and small-scale DERs. However, reaching this goal is challenging due to the inherent mismatch between solar generation and household demands, in addition to uncertain and myriad energy consumption patterns observed by households [262].

During recent years, the research on community-level energy trading have studied optimization techniques and simulation [263], integrating the network constraint [264], and investigating financial gain [265]. However, a feasibility investigation on load balancing potential (i.e., matching the available surplus versus deficit energy) at the neighborhood scale is needed to study the role of realistic users' energy behavior and prosumers' assets. Specifically, it is essential to understand the self-sufficiency potential of a community under realistic uncertainties in load profiles that is driven by households' lifestyles, in addition to simulating the impact of PV integration-level, battery adoption, and load flexibility. Accordingly, this could assist energy policymakers to assess reaching the sustainability target goals of communities and decarbonization assessment under different integration levels of small-scale distributed resources.

Inspired by this, in this paper, we carry out a data-driven feasibility assessment on the load balancing potential among prosumers and consumers at the neighborhood scale. As investigation on real consumption and generation data for a community of ~300 households in Austin, TX, from the Pecan Street Dataport project is considered for the case study. The primary questions we sought to answer include:

- Given the realistic household load profiles, what is the load balancing potential for load balancing in communities with equal prosumers/consumers?
- What is the impact of different PV penetration level for load balancing potential?
- How does the inclusion of battery storage systems improve the load balancing potential?
- How does reshaping energy profiles through load flexibility/user practice impact the load balancing potential?

Using data-driven inference on real energy and consumption data, we present quantified results to measure the community self-sufficiency for the abovementioned scenarios. Therefore, the findings of this work can shed light for policy-makers and energy planners for feasibility assessment of energy trading for decentralized distribution and meeting community sustainability goals, under the realistic assumption of using real demand and generation data across a neighborhood of households.

The rest of the paper is structured as follows: Section 6.2 describes the case study, characteristics of the data, and the applied methods. Section 6.3 presents the results, discussion, and implications for different scenarios, and Section 6.4 presents the concluding remarks.

6.2. Methods and materials

The analytical experiments in this paper are carried out on real demand and PV generation data. For each experiment, and to create a community, groups of prosumers and consumers were selected using bootstrapping technique to account for the uncertainty in energy daily profiles across households. Each experiment has been run 100 times through random sampling of prosumers/consumers. For the cases that battery storage was considered, the charging/discharging was simulated for individual households. For the cases that involved load flexibility by users, the impact of load deferral or partial load shedding was simulated for individual households, and daily load profiles under the users' flexible behavior were reconstructed with simulation. Therefore, upon considering the impact of each attribute/energy behavioral response for individual

households, the aggregate impact of energy exchange simulation over the community could have been measured.

6.2.1. Basic definitions

We define the basic definitions that are covered throughout the paper as follows.

Prosumers (H^p): Electricity consumers who install a type of renewable resource are considered as prosumers, therefore, they can generate energy while consuming it. In this paper, we considered PV as the renewable resource, and prosumers are the ones who have installed solar photovoltaic panels. A household in the prosumer group is shown as $n \in H^p$.

Consumers (H^c): Consumers only consume electricity, therefore, they rely on a central market for electricity purchase or their peers who have PV and could share their excessive production. A household in the prosumer group is shown as $n \in H^c$.

Surplus energy: Surplus energy is available when the amount of PV production at each time instance exceeds the instantaneous demand for prosumers. Surplus energy is offered by prosumers, and it could be transferred in a P2P market, saved in battery storage and used at a later time, or being fed back to the grid with net metering/feed-in-tariff options. Considering the surplus energy of a prosumer as $s_n, n \in H^p$, the total surplus energy of the community is $S^c = \sum s_n$.

Deficit energy: Deficit energy is the amount of energy that need to be supplied either from a central market or acquired from peers who have surplus energy. Deficit energy is primarily requested by consumers. However, when prosumers' demand exceed their generation or available savings in the battery, they have deficit energy, too. Considering the deficit energy of a consumer/prosumer as $d_n, n \in \{H^c, H^p\}$, the total deficit energy of the community is $D^c = \sum d_n$.

Net demand: To account for the surplus and deficit energy, the measured power drawn from the electrical grid is the primary factor.

The net demand, $L(t)$, at each time t is measured as:

$$L(t) = P(t) - G(t) - B(t) \quad (1)$$

In which $P(t)$ is the power demand, $G(t)$ is the PV generation and $B(t)$ is the battery power. Here, $P(t)>0$, $G(t)>0$, and $B(t)<0$ while charging and $B(t)>0$ while discharging. Also, $L(t)<0$ contributes the accumulation of surplus energy, while $L(t)>0$ contributes to the accumulation of deficit energy.

Complementarity factor: The ideal objective of a P2P market is to cancel out the aggregate deficit energy in a community through integrating aggregate surplus energy. Here, we define the complementarity factor (CF) percentage as the ratio of deficit energy from consumers/prosumers that can be covered by prosumers. Therefore:

$$CF(\%) = 100 * \left|\frac{S^c}{D^c}\right| \quad (2)$$

In which S^c and D^c are the surplus and deficit energy of the community, respectively. S^c is provided by prosumers whose net energy is negative, while D^c is requested by consumers and/or prosumers whose net energy is positive. A value of CF =100% is ideal since it indicates the complete independence from the central market and meeting the load balance merely through energy trading. A value of CF>100% indicates the presence of extra surplus energy in the community, which can be saved in the battery or be fed back to the grid.

Self-sufficiency: We extend the definition of self-sufficiency [266] for individual households into a community, as an indicator of PV utilization in the presence of energy trading. Self-sufficiency is defined as:

$$\varphi_{ss} = \frac{\int_{t_1}^{t_2} \sum_{n \in H^p} M(t)dt}{\int_{t_1}^{t_2} \sum_{n \in \{H^c, H^p\}} P(t)dt} \quad (3)$$

in which $M(t)$ is the power generation utilized on-site as follows:

$$M(t) = \min\{P(t), G(t) + B(t)\} \quad (4)$$

$[t_1, t_2]$ is the considered timeframe for measuring the self-sufficiency. In this work, assuming that energy trading is carried out during PV output, $[t_1, t_2]$ is measured across all the PV generation hours. The numerator in Eq. 3, as the on-site generation by PV or storage saving is measured for the prosumers' group (H^p), and the denominator, as the energy demand of the community, is measured for both the prosumers' (H^p) and consumers' group (H^C). Simply, self- sufficiency for the community is the share of PV generation, directly consumed by the community during generation hours.

6.2.2. Household selection

Electricity daily load profiles for residential buildings are known to have high variation across households and across days [13, 114]. Although certain pre-defined (with limited number) [267] or synthetic load profiles [268] can be employed to evaluate the impact of PV adoption in different capacities, it could limit the generalization of findings by overlooking the stochastics nature of

household load profiles. Therefore, to present a reasonable estimation of community load balancing potential, it is essential to account for such uncertainties and include myriad possibilities in load profiles, driven by occupants' energy behavior. Here, for simulating the energy trading scenarios under different PV/battery penetration and load flexibility, we use the real demand and generation daily profiles across two months for each households. The bootstrapping sampling technique is used to form communities by selecting different combinations of households. Therefore, to account for the variation of load profiles, we sample m households (ranging from 20 to 100) to form a community as an individual experiment and repeated the experiment 100 times to account for stochastic nature of load profiles for reporting the hypothetical energy exchange.

To calculate the PV generation and net energy for each households and to measure the surplus and deficit values, the numerical integration on daily profiles was carried out. Considering a daily profile as $P(t)$ (or any other forms like $L(t)$ or $G(t)$), the energy is equal to:

$$E = \int_{t1}^{t2} P(t)dt \tag{5}$$

In which $[t_1, t_2]$ is the timeframe of interest for measuring the energy.

6.2.3. Battery modeling

We considered the availability of battery for prosumers. To schedule the prosumers' battery (assuming they have one), we considered charging of the battery at the presence of surplus energy, and discharging later when the deficit is present. The specifications of the battery was assumed similar to the commercial Tesla Powerwall [269], with a capacity of 13.5kWh, and 7kW peak power. Therefore, physical specifications of the battery is accounted in the battery modeling. Figure *6-1* shows the battery scheduling for the simulation. During the times that the generation is higher than demand, the battery is charged, given the physical constraints of the battery. Discharge of the battery happens when the demand gets higher than generation, if there is available energy in battery accumulated from previous times. When the surplus energy of the household exceeds the battery capacity, then the difference can offered for energy trading with consumers. Here, we opted for a simplified and intuitive model for battery scheduling to show the impact of storage for benefiting the load balancing. More sophisticated battery modeling under the constraint of dynamic pricing, cycle life, and losses could be integrated as well [270].

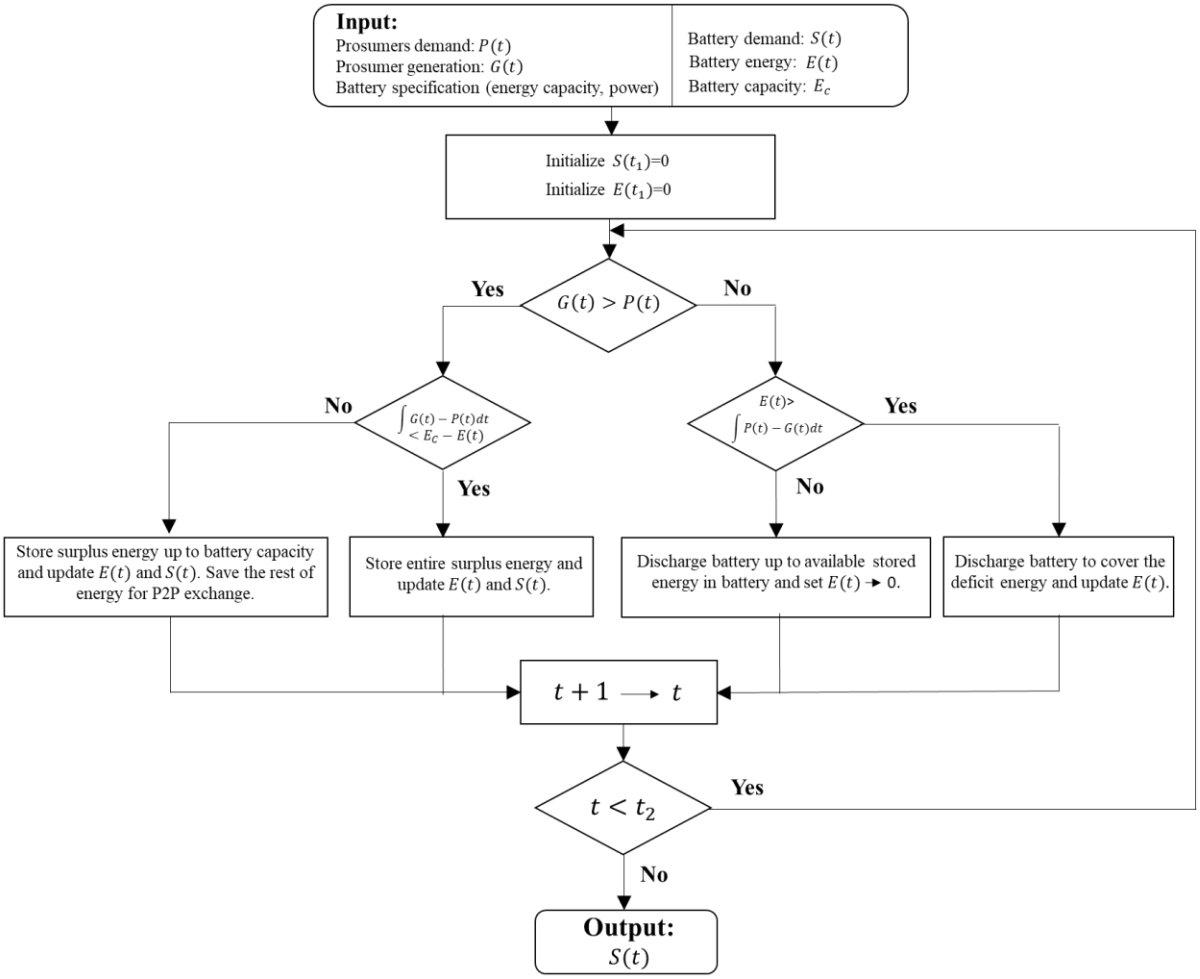
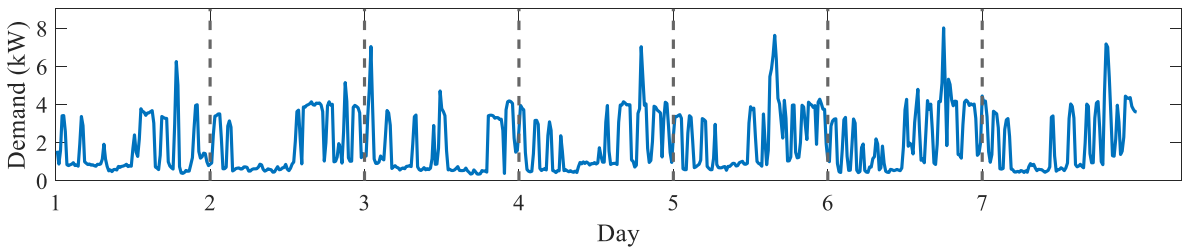


Figure 6-1. Battery scheduling flowchart.

Using the scheduling approach in Figure *6-1*, battery simulation charging/discharging is modeled individually for each prosumers. Figure *6-2* shows an example of charging/discharging pattern of the battery, in addition to demand and generation patterns of one prosumer during one week.



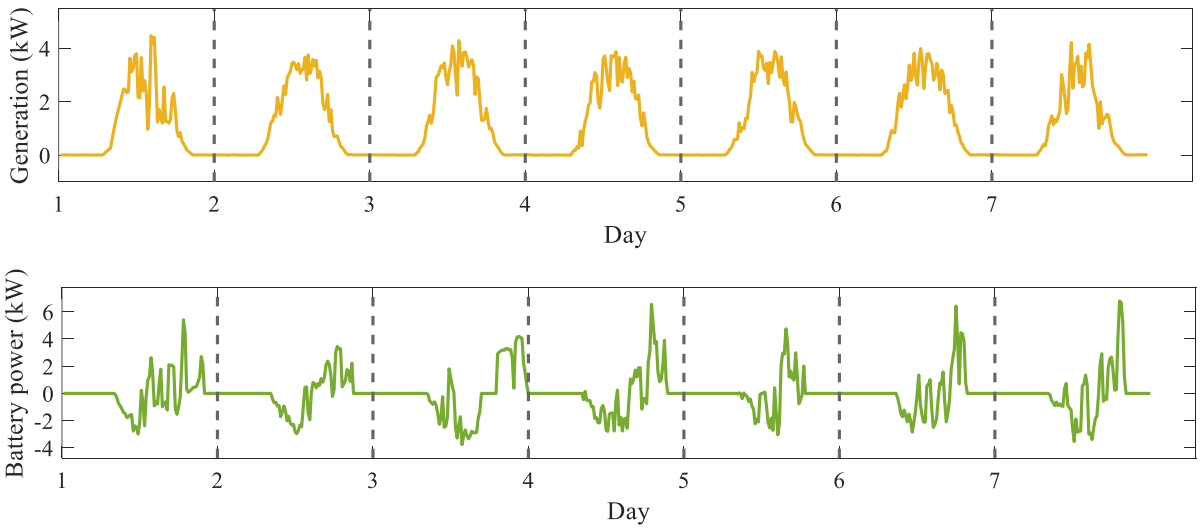


Figure 6-2. Examples of (a) demand, (b) generation, and (c) battery charge/discharge pattern for one household during one week.

6.2.4. Users' flexibility

In the smart grid context, loads flexibility can be achieved by user's practice through postponing the load operation or through adoption of smart loads (i.e., smart appliances) with automated scheduling capabilities. Therefore, flexibility through reshaping the demand profiles provide more room for efficient utilization of PV. Specific appliances have shown to have high flexibility in operation [37, 147, 148], which include EV, AC, and wet appliances (washing machine, dryer, dishwasher). AC's flexibility can be achieved by changing the temperature setpoint to reduce the demand, while EV and wet appliances flexibility is offered through delaying the operation/charging time. Here, we have used the concept of flexibility for investigating the improvement in community load balancing. Through using individual appliance-level data associated with each daily profile in the dataset, the users' flexibility was modeled. Specifically, the EV charging energy (if any) and wet appliances energy (if any) associated with each daily profile of prosumers and consumers was measured on an hourly basis. Similarly, for AC in the residential buildings, through using the suggestions in [271, 272], the energy saving associated with $1°F$, $2°F$, and $2°F$ (with pre-cooling) increase of temperature setpoint was subtracted from the actual AC profiles. 25%, 31%, and 68% of AC power reduction with $1°F$, $2°F$ (with pre-cooling), and $2°F$ temperature setpoint increased was modeled [271]. Therefore, by adjusting a subset of prosumer/consumers who allow for flexible operation of the abovementioned appliances

in our experiments, we could measure the aggregate change of network demand and integrate them with metrics described in section 0.

6.2.5. Case study community

We used the historical energy dataset of 244 residential households located in Austin, TX for the case study. The dataset is available through Pecan Street Project [68]. The community included 119 prosumers with PV and 125 consumers. Since our case study focused on the PV rooftop solar panel as the renewable source, we used the data for July and August, as the representative months in summer, which has warmer climate and sunny days. The dataset was collected during 2015. The demand data and generation data with 15-minute resolution, similar to what smart meters record, was used. For the generation data, less than 0.001% of profiles (8 out of 7152) had missing information, in which were eliminated from the dataset. In total, the prosumers and consumers dataset included 7144 and 7492 daily profiles for the considered timeframe. Each daily profile included 96 data points, which was annotated with 'household ID' and 'day of the year' index for the sake of information retrieval.

6.3. Results and discussion

6.3.1. Exploratory data analysis

We began the analysis by presenting the characteristics of the dataset for prosumers and consumers. This includes the statistics for prosumers and consumer energy usage, PV generation, and the variability analysis of households load shapes.

6.3.1.a. Demand, generation, and household energy patterns

As a key assumption, the complementarity of surplus and deficit energy is required to enable energy exchange. *Figure 6-3* shows the daily load shapes of a consumer (*Figure 6-3*(a)) and a prosumer (*Figure 6-3*(b)) plotted over 20 subsequent days. As can be seen, the complementarity patterns of load shape between the prosumers and consumers in the shaded area (time of generation) can be leveraged for P2P energy trading and load balancing.

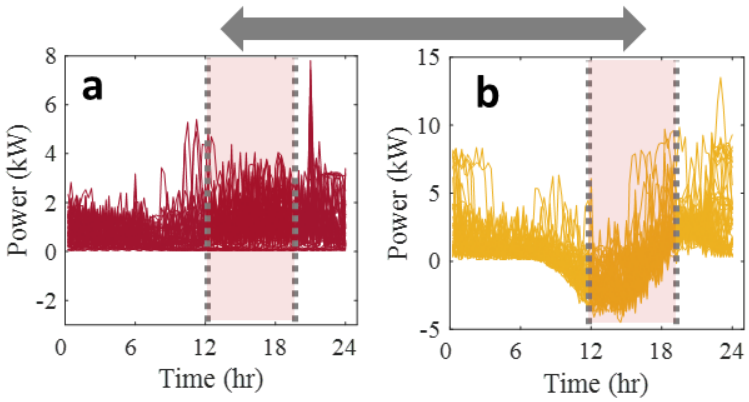


Figure 6-3. Load complementarity for energy trading: (a) a consumer's daily profile with deficit energy (+ net values), (b) a prosumer's daily profiles with surplus energy (- net values).

Figure 6-4(a) shows all the PV generation profiles (7144 profiles) in the dataset for 119 prosumers. The median PV peak power was 3.95kW (5^{th} and 95^{th} quantile of 1.8kW, 6.5kW), while there were outlier profiles with highest PV peak of 11.0kW. *Figure 6-4*(b) shows the distribution of PV generation from 9 am to 7 pm, as the times when solar was available. The highest generation potential is between 2 pm to 3 pm. Furthermore, in the earlier timeframe in the morning spanning later to the maximum generation time, the generation has the potential to entirely supply the demand (and also stored in the battery, if available), while in the later timeframe stretching to the evening, the generation can partially supply the demand (and discharged from the battery, if available, to cover the rest). Since most of PV generation is between 9 am to 7 pm, we considered this timeframe in the rest of the paper for presenting the results.

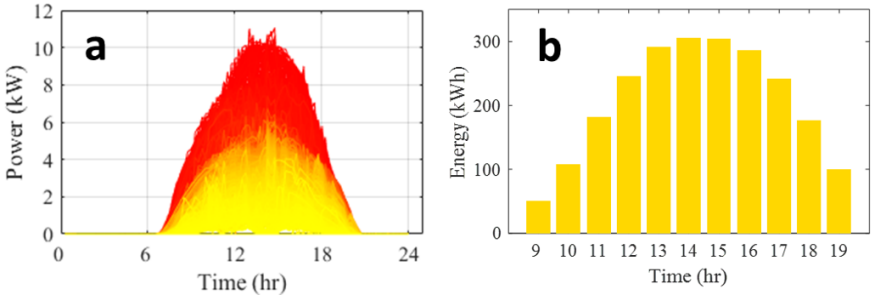
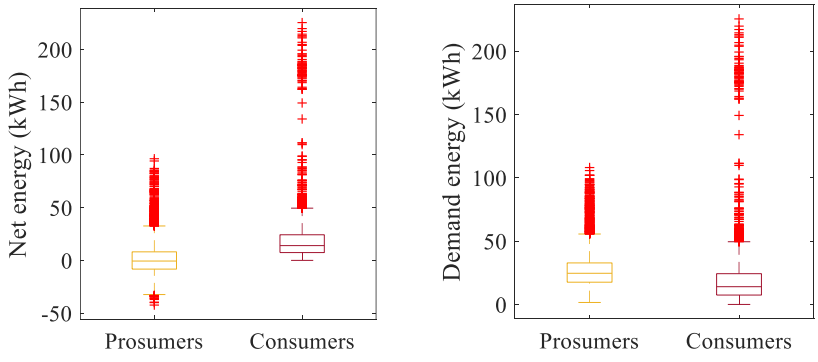


Figure 6-4. Solar generation patterns in the community: (a) PV power profiles, (b) Daily averaged PV generated energy.

As part of the sanity check on the data representation for the community load balancing, the consumers and prosumers should be almost homogeneous in their absolute energy demand. In other words, if the community of prosumers turn out to have significantly higher demand compared

to the consumers pool, they consume their on-site PV generation themselves (i.e., positive net value) and would not be able offer energy trading. Figure 6-5 compares the distribution of the daily net energy (drawn from the grid) and demand energy for prosumers (N=119, 7144 profiles) and consumers (N=125, 7492 profiles). PV generation hours during 9am-7pm was considered. In Figure 6-5(a), the median net energy for prosumers is -0.6kWh (5^{th}, 25^{th}, 75^{th}, and 95^{th} percentile of -18 kWh, -8 kWh, 8 kWh, and 27kWh). The observation on prosumers net energy shows that, on average, the prosumer group can highly supply its demand need, while adding consumers to the community will reduce the self-dependency. In Figure 6-5(b), it is show that the demand energy of prosumers (without considering PV generation), is higher on average compared to the consumers. For the prosumers, the median demand energy for is 25kWh (5^{th}, 25^{th}, 75^{th}, and 95^{th} percentile of 8 kWh, 18 kWh, 33 kWh, and 52 kWh) while for the consumers, the median demand energy for is 14kWh (5^{th}, 25^{th}, 75^{th}, and 95^{th} percentile of 3 kWh, 7 kWh, 24 kWh, and 42 kWh). This observation inclines that in the community, PV owners have heavier consumption, which could be associated with external factors such as larger household size, or the presence of modern appliances with higher usage and/or owning a plug-in EV. However, as the distributions of energy demand between consumers and prosumers are relatively comparable, we considered the dataset suitable for the presentation of load balancing potential results.

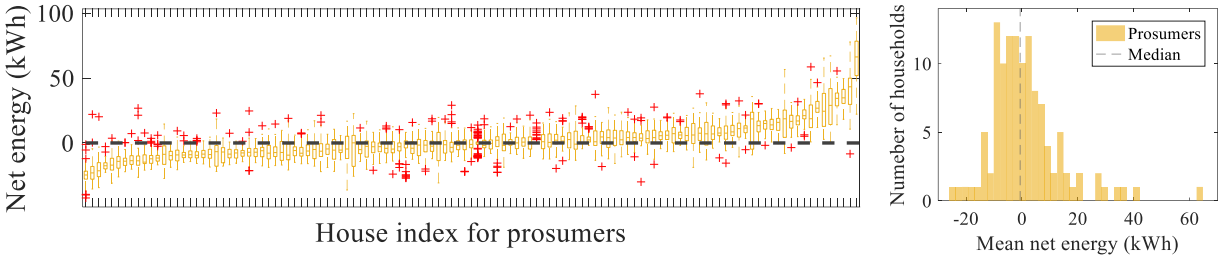


(a) (b)

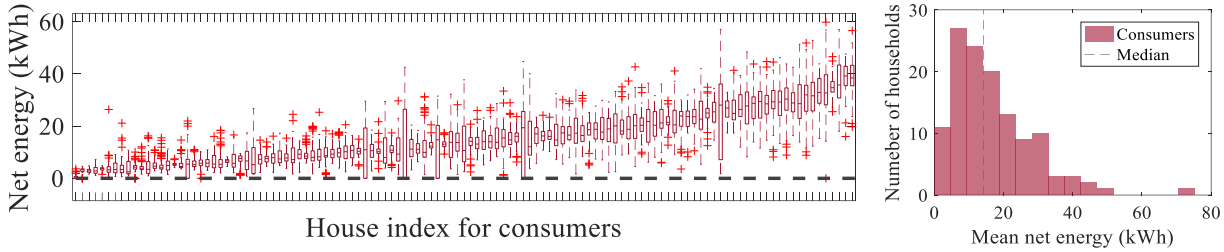
Figure 6-5. Distribution of (a) demand for the entire community and (b) drawn from the grid energy and from 9 am-7 pm.

As noted, a high variation in daily energy usage exists across both households and subsequent days. Figure 6-6 presents the variation of total daily energy consumption across households, each with two months of daily profiles. In Figure 6-6(a), 53% of prosumers (63, $N = 119$), on average

have a daily negative net energy (i.e., surplus), which they could offer to neighbors, while maintaining their self-dependency (dashed baseline of 0 in Figure 6-6(a)). For consumers (Figure 6-6(b)), all have positive net energy intuitively, while both the average daily energy and inter-day difference is considerable across households.



(a)



(b)

Figure 6-6. Distribution of net energy for (a) prosumers and (b) consumers from 9 am-7pm for all households.

6.3.1.b. Entropy analysis for household load profiles

Households exhibit variously different load profiles across multiple days. Accordingly, the variability of energy profiles shows the extent to which a prosumer/consumer is stable or unpredictable. To demonstrate the distribution of household variability in energy patterns, the concept of entropy was used. Using a clustering technique, and assigning the entire library of daily profiles into different cluster, the entropy of a household, E_n, is defined as:

$$E_n = -\sum_{i=1}^{K} P(C_i)\log\big(P(C_i)\big) \tag{6}$$

In which $P(C_i)$ is the probability of observing cluster i, and K is the total number of cluster. A low value of E_n indicates higher stability of households' load profiles across different days, while a higher E_n denotes low predictability. The lowest value for E_n is zero, when all daily load profiles of a household belongs to one cluster.

The K-means clustering technique was used on daily load profiles of prosumers (7144 profiles) and consumers (7492 profiles). Upon empirical observation, using 5 clusters revealed distinct load profiles with different patterns. *Figure 6-7* shows the clusters of load profiles (net energy value) in addition to the entropy of individual households. For the prosumers community (*Figure 6-7*(a)), except cluster 2 with 20% frequency, which did not offer surplus energy at any time of the day, other clusters all had some surplus energy with different peak levels. Therefore, cluster 2 have no potential for energy trading, while cluster 3 has the highest potential due to its sharp surplus peak during PV generation in addition to its moderately low peak during evening when PV generation diminishes. For the prosumer community (*Figure 6-7*(b)), cluster 2 is a suitable candidate energy trading due to sharp deficit peak at PV generation time, while cluster 3 (with constant consumption), and cluster 1 and 5 (with peak demand around 6:00 pm) also offer some potential for P2P exchange. Cluster 4 has low consumption until around 11 pm, in which a peak arises, therefore, not justified for energy trading (unless their prosumers neighbors have extra energy in storage). *Figure 6-7*(c) and *Figure 6-7*(d) shows the distribution of entropy with respect to average daily net energy for prosumers and consumers, and each data point is one household. The horizontal/vertical dashed line in those plots reflect the 25th and 75th percentile of values for entropy/net energy. As the distribution shows, for each of the nine areas, the community includes a variety of households with low to high predictability and energy demand, which further reflects the varied range of households with different energy behavior.

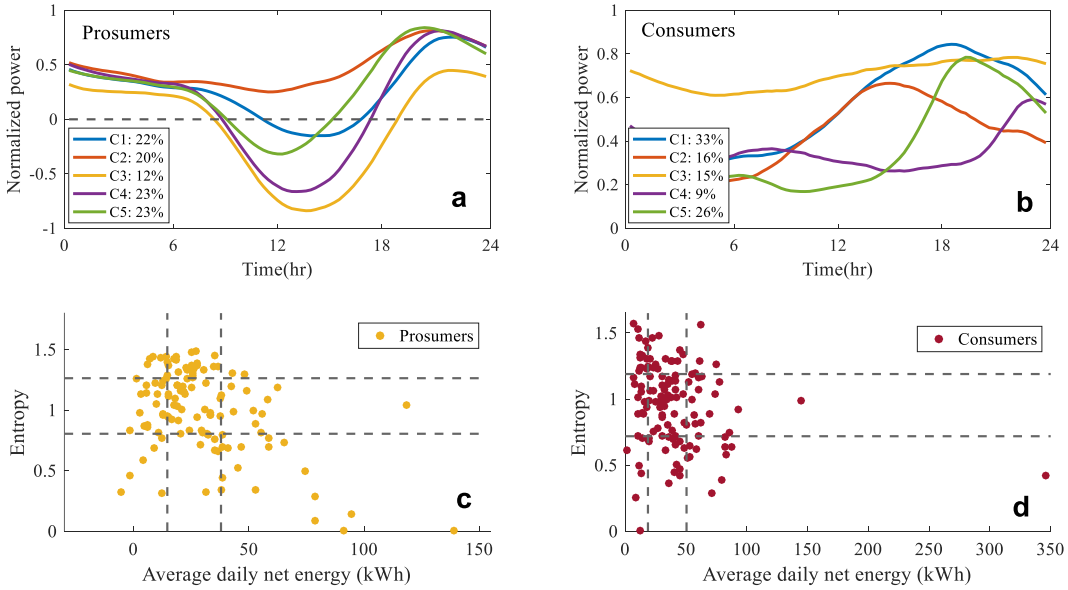


Figure 6-7. Clustered load profiles of prosumers and consumers and their entropy distribution: (a) clusters of prosumers daily profiles, (b) clusters of consumers daily profiles, (c) entropy of prosumers households, and (d) entropy of consumer households. Values in the legend for subplot (a) and (b) for shows the frequency of clusters in the community.

6.3.2. Load complementarity for energy balancing

This section presents the results of load balancing potential under the impact of PV integration, battery adoption, and users' load flexibility.

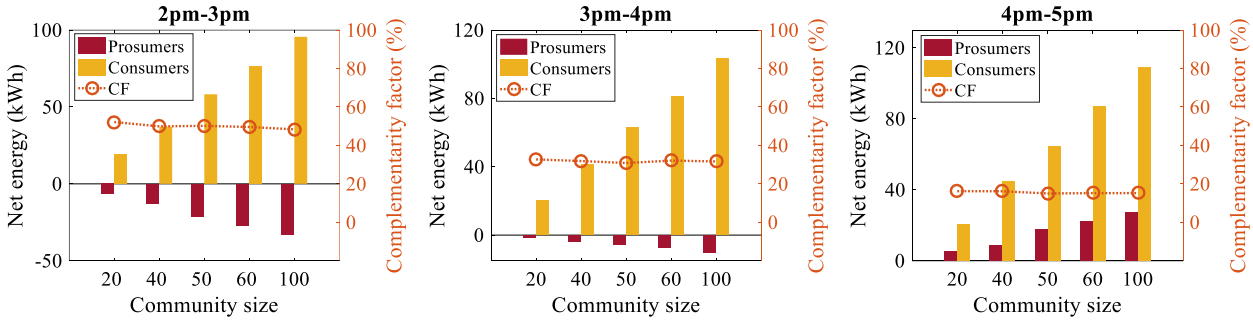
6.3.2.a. Equal distribution of prosumers and consumer

In this scenario, we considered ***"****what is the load balancing potential for communities with the same number of prosumers and consumers with no change in behavior?".* Here, it was assumed that the community contains the same number of prosumers and consumers, and no flexibility in behavior or storage is available. Therefore, it reflects the potential of energy trading just based on PV generation.

Using subnetworks of n households with surplus and deficit energy, we performed a simulation and measured the extent to which energy balancing that can be achieved. A range of $n = \{20,40,60,80,100\}$ as the household sample size in the community was considered for measuring the complementarity factor. Therefore, the number of prosumers and consumers was assumed to be $n/2$ in this case. For each community size and each hour of PV generation, 100 experiments, indicating 100 communities, was considered. Figure 6-8 presents the variation of net energy by prosumers, deficit energy by consumers, and complementarity factor (%) for three subsequent hours (2pm-5pm). In all subplots, it is shown that increasing the community size results in linear change for surplus energy and deficit energy ($R^2 > 0.98$ for all cases). As a result, the complementarity factor (CF) for various community size remains almost constant ($\sigma < 1.5$ for all three subplots). During 2-3 pm, which is the highest PV generation time, CF reaches 50.0% on average for various community size, while it declines to 31.8% and 15.5% during 3-4 pm and 4-5 pm, respectively. The reduction of CF is associated both with the decline of PV generation, in addition to the considerable increased demand of prosumers and consumers at the hours stretching to the evening. Specifically, during 4-5pm, the community of prosumers' net energy is positive (generation do not cover the demand). Therefore, unlike 2-4pm, the aggregate surplus energy offered by a subset of prosumers is not able to cancel out the deficit of another subset of prosumers.

A complete statistics for different hours of PV generation (similar to what presented in *Figure 6-8*), is presented in *Table 6-1*. As shown, the highest CF happens at 12-1 pm, in which prosumers surplus energy covers could cover up to 72.5% of the entire community's demand through energy trading. However, during 16-19pm and 9-10am, prosumers do not offer any surplus energy within themselves (i.e., positive net energy values), which reduces the potential for energy trading. In the extreme cases, at and 5-6pm and 6-7pm, the CF is less than 5% and 1%, indicating almost no potential for energy trading. Apart from these hours, the CF is positive at different hours, varying between 15.5 (3-4 pm) to 62.9 (11-12 pm).

To summarize, *in communities with equal distribution of prosumers and consumers and without any storage capacity or user flexibility, the CF (%) is highly dependent on PV generation, therefore, exhibiting high variation at different times of generation (standard deviation of* 24% *in complementarity). During hours of PV generation from 9am-7pm, 60% of time, prosumers group on aggregate have surplus energy and could offer to consumers. In the most productive hour, CF reaches to 72.5%, while on average it is 35.6 %. Besides, increasing the community size leads to linear increase in surplus and deficit energy and therefore the CF remains almost constant*.



(a) (b) (c)

Figure 6-8. Comparison of surplus energy, deficit energy, and complementarity factor for various community size with equal number of prosumers and consumers at (a) 2-3pm, (b) 3-4pm, and (c) 4-5pm. The left axis represents the net energy for bar charts, and the right axis represents the complementarity factor for the line.

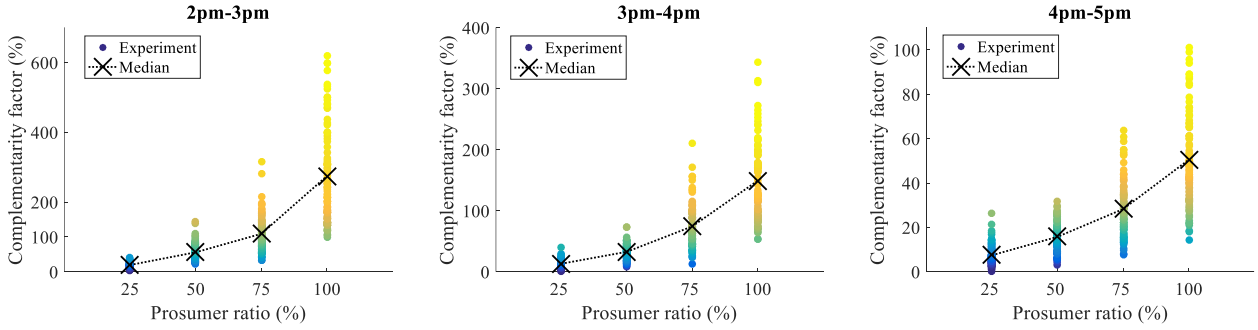
Table 6-1. Surplus energy, deficit energy, and complementarity factor for various community size with equal number of prosumers and consumers from 9am-7pm.

Time of day	Community size	Prosumers net energy (kWh)*	Consumers net energy (kWh)*	CF (%)*
9 am-10 am	20	3.9	10.1	17.2
	40	8.6	22.1	14.8
	60	12.0	33.0	15.3
	80	16.4	42.6	15.6
	100	20.0	54.3	15.1
10 am-11 am	20	-2.6	13.0	45.7
	40	-3.8	23.5	43.1
	60	-6.1	36.0	41.5
	80	-9.4	48.4	42.5
	100	-10.1	60.3	41.2
11 am-12 pm	20	-6.0	13.7	62.2
	40	-13.6	27.5	63.9
	60	-21.5	41.5	64.6
	80	-28.2	54.8	63.7
	100	-34.1	71.6	59.8
12 pm-1 pm	20	-10.2	15.3	79.3
	40	-19.6	33.2	69.6
	60	-29.4	47.1	71.7
	80	-39.8	62.6	72.1
	100	-48.9	78.9	69.9
1 pm-2 pm	20	-8.5	18.8	61.4
	40	-17.3	35.2	61.8
	60	-25.0	55.2	57.8
	80	-32.6	72.1	57.3
	100	-42.3	89.4	58.4
2 pm-3 pm	20	-6.6	18.9	52.2
	40	-12.6	37.1	49.9
	60	-21.1	58.2	50.2
	80	-27.0	76.4	49.6
	100	-32.7	95.6	48.3
3 pm-4 pm	20	-1.5	20.2	32.7
	40	-3.8	41.5	31.9
	60	-5.3	63.2	30.8
	80	-7.6	81.2	32.1
	100	-10.0	103.2	31.6
4 pm-5 pm	20	5.6	20.7	16.2
	40	9.0	44.8	16.1
	60	17.5	64.8	15.0
	80	22.4	86.9	15.2
	100	27.4	109.1	15.2
5 pm-6 pm	20	14.3	23.3	4.9
	40	29.4	48.1	4.5
	60	42.9	70.7	4.6
	80	57.8	94.9	4.6
	100	71.1	117.2	4.8
6 pm-7 pm	20	23.8	25.2	0.8
	40	48.0	51.4	0.8
	60	70.0	75.6	0.9
	80	94.1	101.0	0.9
	100	117.9	125.3	0.9
Average				35.6

* Values for each cell in these columns reflects the average of 100 experiment.

6.3.2.b. Varying prosumers distribution

In this scenario, we considered *"what is the load balancing potential for communities with different ratio of prosumers?".* Therefore, we considered varying level of PV integration in the community, without considering any storage or users' flexibility. A community size of $N = 20$, and 100 repetition of experiment with bootstrapping was considered. *Figure 6-9* presents the complementarity factor (%) for three subsequent hours (2pm-5pm). Prosumers ratio (i.e., PV integration) of {25%, 50%, 75%, 100%} was considered for the community. Therefore, in each experiment, 5, 10, 15, and 20 (out of 20) prosumers exist in the community. As can be seen, during 2-3 pm, with 75% and 100% prosumer ratio, the CF exceeds 100% (109% and 274%, respectively), thereby covering the entire demand of both prosumers and consumers and offering excess surplus, which could go back to the grid or stored in battery. For 25% and 50% of prosumers ratio in that case, the median of CF is 19% and 56%, respectively. During 3-4pm, only with 100% prosumers ratio, the CF exceeds 100% (149%) while for 25%, 50%, and 75% prosumer ratio, 13%, 33%, and 74% of the network demand can be supplied by prosumers. However, during 4-5pm, with 100% prosumer ratio, the CF only reaches up to 50%, while lowering PV integration further limits the complementarity in network (CF of only 7% with 25% prosumer ratio).



(a) (b) (c)

Figure 6-9. Comparison of complementarity factor for varying level of prosumers in the community ($N = 20$) at (a) 2-3pm, (b) 3-4pm, and (c) 4-5pm.

Due to the highly varied energy demand and load profiles, the CF values in *Figure 6-9* shows considerable variation in different experiments. *Figure 6-10* shows the distribution of complementarity factor at 2-3pm for a prosumer ratio of 0.25 (first case in Figure *6-9*(a)). Due to the similarity to the normal distribution, a two-sample Kolmogorov-Smirnov (KS) test on the histogram and normal distribution was conducted. The test result was not significant (*p*-value=0.89), indicating the normality of the distribution. The KS test for other scenarios (different

prosumers ratio) in addition to different hours in *Figure 6-9* showed *p*-values of higher than 0.05 except in one case.

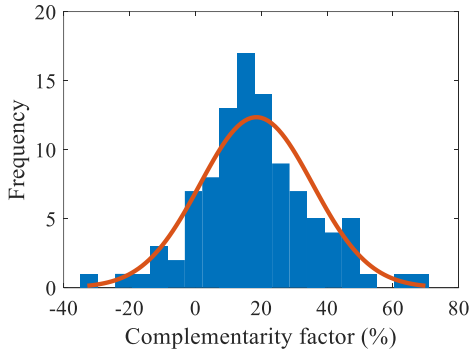


Figure 6-10. Histogram of complementarity factor at 14pm-15pm for prosumer ratio of 0.25.

Figure 6-11 compares the average complementarity factor and its 95% confidence interval for different prosumer ratio during hours of PV generation. Due to the high variation across different prosumer ratio and hour of the day, the y-axis is shown in logarithmic scale. For all the scenarios, it was observed that during the timeframe of 9am-4pm, the aggregate net energy of prosumers was negative (i.e., surplus energy), with high potential of P2P exchange. For 25% and 50% prosumer ratio, the highest CF, during 12-13pm, offers a value 27% and 75%. However, for 75% and 100% prosumer ratio, the CF exceeds 100% from around 10am-3pm, with the highest value of 165% and 420% during 12-13pm. Additionally, during 11am-3pm, with high integration of 75% and 100% PV, not only prosumers can supply their own demand and consumer peers, but on average they still have a surplus energy equal to 30% and 217% of the network demand. However, without any type of storage, the surplus energy has to go back to the grid with the option of net metering/feed-in-tariff. This observation shows that, while increasing PV generation improves load balancing potential, its benefit is limited to the hours of highest PV generation, but not during later times in evening when the demand gets higher. Specifically, regardless of the PV integration level, after 4pm, the aggregate prosumers net energy was positive, which considerably limit the complementarity factor. Even with 100% prosumer ratio in community, only 12% and 2% of the network deficit (from both prosumers and consumers) could be supplied by PV generation.

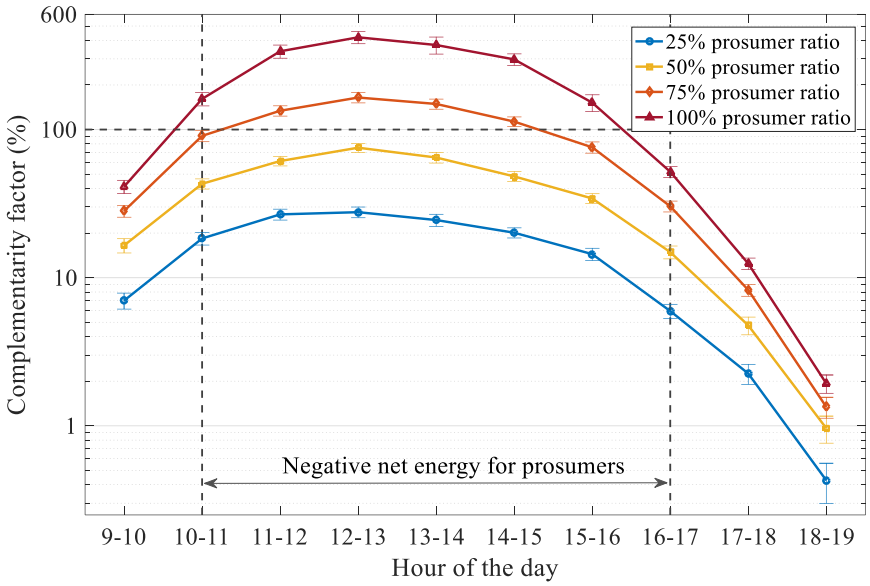


Figure 6-11. Impact of varying prosumer ratio on complementarity factor of the community during PV generation hours for 25%, 50%, 75%, and 100% prosumer ratio.

To summarize, *in communities with high ratio of prosumers with PV standalone systems (more than 75%), energy trading can cover all the demand (100%) of community during hours of highest PV generation (11am-3pm). Additionally, extra surplus energy is available during those hours with the potential of storage or feeding back to the grid. However, in communities with PV standalone systems, significant shortage in supply of the community can be induced during later times of PV generation (4-7pm), in which prosumers are not able to supply their own demand. Therefore, to efficiently utilize the PV generation for energy trading leading to later times (4-7pm), alternate measures, including deploying battery system or users' flexibility is needed.*

6.3.2.c. Integration of battery storage

In this scenario, we considered *"what is the load balancing potential for communities with different ratio of battery storage?"*. Therefore, in addition to setting varying ratios of PV integration in the previous section, we also considered different levels of battery storage adoption in the community. The adoption of battery storage is only considered for the prosumers. Therefore, for each ratio of prosumers, we considered the different battery integration levels for prosumers. PV ratios of {0.25, 0.5, 0.75, and 1} and battery ratios of {0.25, 0.5, 0.75, and 1} was used for forming communities with N=20 through repeating 100 experiments. As an example for PV ratio of 0.5, and battery ratio of 0.5, 10 prosumers and 10 consumers includes the community, out of which 5 prosumers have battery storage.

Initially, the presence of a community storage level without physical constraint for charging was considered to show the most optimistic load balancing potential with unlimited storage. *Figure 6-12* shows the impact of the community storage battery for different prosumers ratios. During 10 to 15, which were the hours that CF exceeds 100% with PV-standalone system (untapped potential for PV utilization), the battery stores the extra energy and CF is set 100. From 16 to 19, which were the hours with low chance of energy trading (*Figure 6-11*), a 100% ratio of prosumers could result $CF = 100$ during all evening hours, given unlimited storage. With 75% prosumer ratio, the CF at 4-5pm, 5-6pm and 6-7pm, would be 85%, 49%, and 10%, which shows an improvement of 180%, 75% and 80% compared to the similar case with the PV-standalone system.

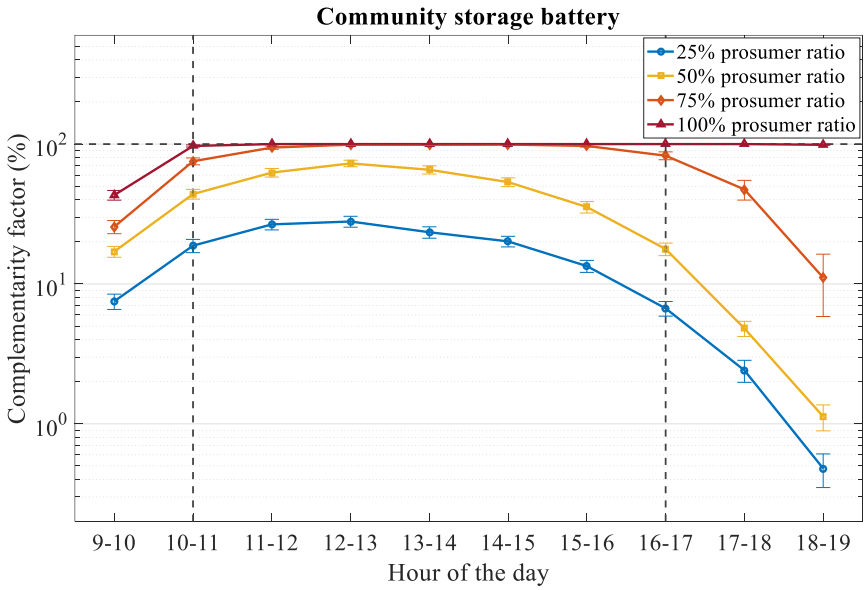


Figure 6-12. Impact of community storage battery without physical constraints during PV generation hours for 25%, 50%, 75%, and 100% prosumer ratio.

Figure 6-13 presents the self-sufficiency of the community during the PV generation hours under different PV and battery adoption level. As can be seen, the increase of battery adoption results in improvement in self-sufficiency of the community. Particularly, a maximum improvement of 4.8%, 11.3%, 13.0, and 17.0% in self-sufficiency is observed with full integration of batteries, under 25%, 50%, 75%, and 100% of PV ratio respectively. Specifically, with 100% PV-battery adoption, the self-sufficiency of the community increases to 83%.

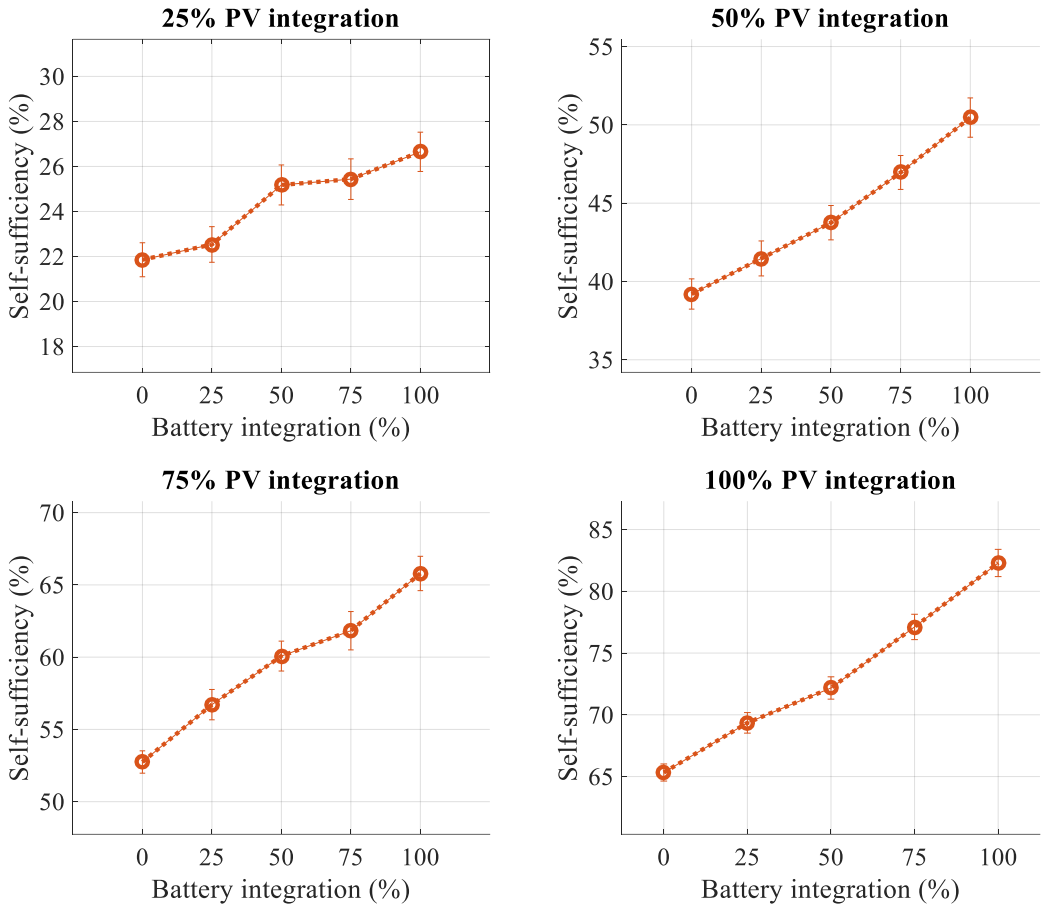


Figure 6-13. Self-sufficiency of the community with different levels of PV and battery.

6.3.2.d. Impact of users' flexibility

In this scenario, we considered **"***what is the impact of users' flexibility in improving the load balancing potential?".* We specifically considered different levels of contribution as {20%, 40%, 60%, 80%, 100%} from users under different combinations of flexible loads {AC 1°, AC 1° + deferrable loads, AC 2°(with pre-cooling), AC 2°(pre-cooling) + deferrable loads, AC 3°, AC 3° + deferrable loads}. Given the low potential of P2P exchange during hours leading to evening time (*Figure 6-11*), the flexibility was considered during a relatively short needed time, which is the timeframe of 5pm-7pm.

Figure 6-14 present the self-sufficiency of the community under different PV integration (each subplot) and different flexibility contributions from prosumers/consumers. Each line represent the type of load/temperature set point change in which the flexibility was considered. Each data point shows the mean value from 100 experiments. As can be seen, increasing the flexibility will result

in increased self-sufficiency. Particularly, for 25% PV, a maximum increase of 4.4% (22.1% to 26.5%), for 50% PV, an increase of 5.4% (39.3% to 44.7%), for 75% PV, an increase of 9.0% (53.2% to 62.2%), and for 100% PV, an increase of 10.4% (65.1% to 75.5%) was observed. Regarding the level of flexibility contribution, on average (regardless of PV integration), an improvement of 2.2%, 4.1%, 5.0%, and 7.5% with 25%, 50%, 75%, and 100% flexibility participation over the baseline is observed.

To summarize, *flexibility in users' behavior, with small temperature setpoint change (2°) and deferring the flexible load operation, during hours of low generation (5pm-7pm) can improve the self-sufficiency up to more than 10%. Specifically, in communities with 100% PV ratio, the self-sufficiency at PV generation time could reach to ~75%.*

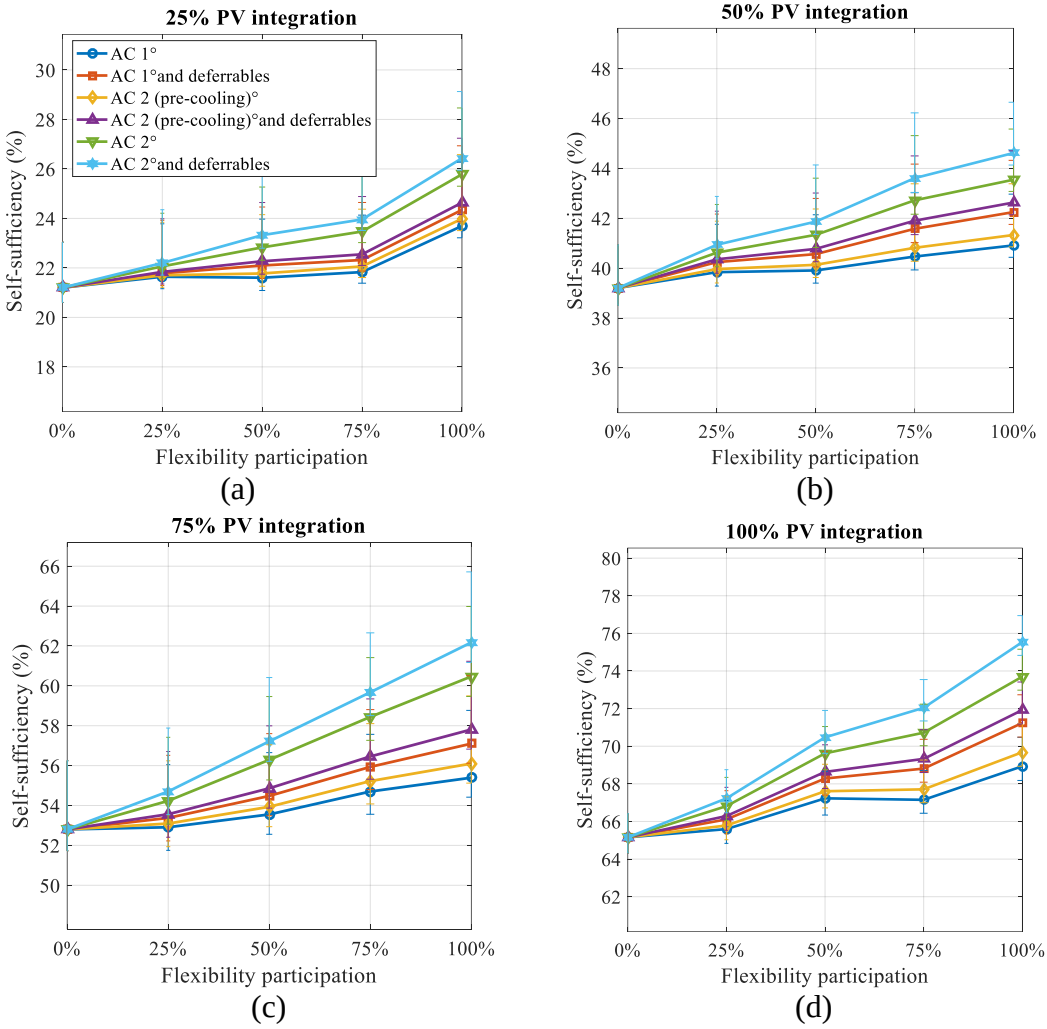


Figure 6-14. Impact of flexible behavior on self-sufficiency for (a) 25% PV integration, (b) 50% PV integration, (c) 75% PV integration and (d) 100% PV integration.

6.3.2.e. Comparison of users' flexibility versus battery storage

The results in Sections 6.3.2.c and 6.3.2.d showed the potential of two attributes (battery storage and user's flexibility) for improving the load balancing potential during PV generation hours. To compare the impact of these two attributes, *Figure 6-15* shows the improvement of self-sufficiency through users' flexibility and battery storage compared to the baseline (i.e., just PV integration). As can be seen, with increasing the PV ratio in the community, the improvement by including users' flexibility and battery storage is more highlighted. Furthermore, with low PV ratio (25%) the impact of user's flexibility is comparable with the battery storage while with high PV ratio, battery storage shows higher improvement compared to the user's flexibility. Nonetheless, with high PV ratios of 75% and 100% in the community, the users' flexibility practice improvements showed to be 69% and 61% of that of the battery storage. Therefore, in the case of low/none batteries adoption in the community, users' flexibility can be deemed as a viable/alternate approach for utilizing PV generation for energy trading. This could be achieved by using advanced technologies like smart appliances and smart thermostat, which allow for automation under user's allowance, or users' manual practice in operation of flexible loads.

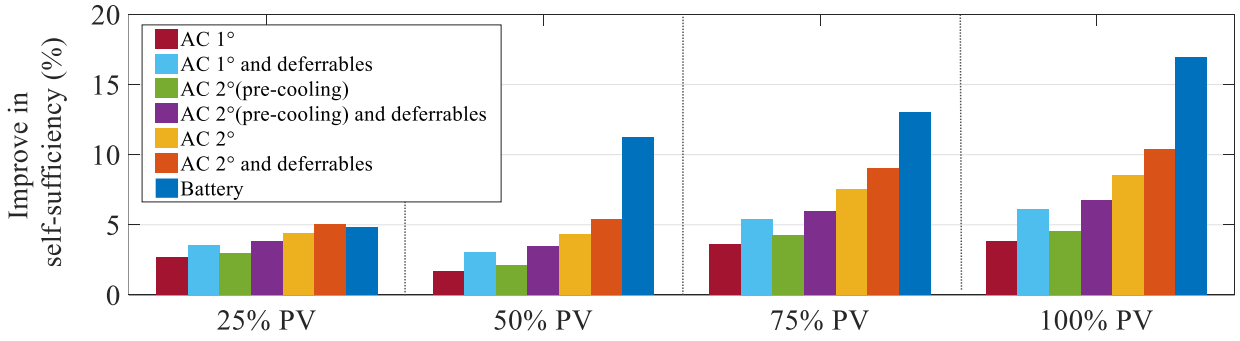


Figure 6-15. Comparison of self-sufficiency improvement with different users' flexibility and battery storage.

6.3.3. Limitation

There are a number of limitations associated with this work: (1) Like any data-driven studies, the findings presented here were extracted from the dataset. Therefore, the generalizability of findings is associated with similarity in energy profiles of prosumers/consumers. Nonetheless, we used the data from Pecan Street Project [68], which is currently one of the largest campaign for energy initiative, and selected ~250 households from the ERCOT grid. Nonetheless, the majority of analysis presented here included sub-hourly resolution demand and generation data. Demand data

is extracted from smart metering, which is a ubiquitous and available metering infrastructure in most regions. Generation data, in the absence of real PV data, can also be estimated from temperature and geographical location information with open-source solutions (e.g., PV Watts[1]). Therefore, similar type of analysis can be carried out easily on other datasets. (2) The findings for load balancing potential for the community was obtained through the aggregation of surplus and deficit energy of individual consumers, assuming the energy trading without including transmission loss or network constraints. Studying these factors could further shed light on the findings of this work. (3) For the battery scenario, we considered flat pricing rate. Therefore, battery was discharged at the earliest time that demand exceeds generation. In the presence of dynamic pricing, sophisticated optimization techniques can be used for battery modeling and ideal times for energy trading. (4) We relied on the presence of PV systems that are installed in prosumers' households, with the same capacity as specified in the dataset. Therefore, the impact of different PV capacities or battery sizing was not modeled. (5) We considered discrete values for PV, battery, and flexibility rates. Nonetheless, the model is dynamic and can map the results for any arbitrary values. Such criteria can be set by utilities or decision-makers. (6) We presented the results of case-study for the summer. Therefore, integrating the results over the yearly period could show the impact of seasonality.

6.4. Conclusion

This paper presents data-driven quantification of load balancing potential amongst prosumers and consumers, which could be achieved through energy trading. A case-study of ~250 buildings was done to quantify the surplus/deficit energy of prosumers and consumers under the presence of PV solar generation for communities. The impact of hour of the day during PV generation, PV integration rate, battery integration rate, and users' flexibility on load operation was modeled to measure complementarity and self-sufficiency.

The following key observations were made: Standalone PV in households would considerably help in balancing the surplus and deficit energy within a community during times of high solar generation. Specifically, with equal distribution of prosumers and consumers, ~75% of deficit energy in the community can be canceled out prosumers at the highest PV peak hour. Furthermore, with 100% PV integration, the community can become entirely self-dependent during at high PV

[1] https://pvwatts.nrel.gov/

generation times (11am-3pm) through covering 100% of deficit energy by energy trading. However, during the times leading to evening (4pm afterwards to 7pm when PV generation diminishes), the prosumers no longer could supply their own demand. As a result, the impact of alternate measures such as battery storage systems or user's flexibility was studied. It was observed that with 100% battery integration or 100% users' flexibility in adjusting flexible loads during critical times, the self-sufficiency of the community could improve around 85% and 75% respectively. Furthermore, at the absence of storage systems, the user's flexibility showed the improvement in self-sufficiency by ~60% of what could be offered by commercial conventional storage batteries.

Future directions of this research comprise of studying the seasonality impact, evaluation on other datasets at different geographical locations, the impact of different battery and PV sizing, and including the network constraints to model P2P energy trading.

Chapter 7: Conclusion

7.1. Summary of studies

In this dissertation, data-driven techniques are proposed for analyzing the temporal energy data of households with application to distributed energy management. As the core objective, the solutions addressed in this work aim at improving the match between energy demand and energy supply of the communities through mining the high-resolution energy data of households. Specifically, the problems associated with this dissertation included (1) segmentation of households according to their peak demand shaving potential for DR events, (3) inference of time-of-use events of appliances for DR, and (4) impact of integrating small-scale distributed elements (e.g., solar panels, storage, and smart loads) for balancing the demand-supply.

For the first problem, a segmentation approach through time-series analysis on energy data and statistical modeling was introduced. The complex interactional behavior of human-appliances were modeled by extracting the frequency, consistency, and peak time usage on energy data. Using the approach, the community of households were ranked based on their peak shaving potential for DR events. The results showed the applicability of the predictive approach for segregating different households according to their peak shaving potential. Through quantitative analysis, the potential of different appropriate loads for DR applications (e.g., EV, AC, and dryer) were ranked. Furthermore, through simulation, the proposed segmentation approach showed the applicability of avoiding the rebounded effect (i.e., creating undesired peak demand right after DR event) through justified identification of a small set of households in the community.

For the second problem, ML solutions were introduced to infer the timing of appliance events from the whole-house energy data. Given the primary limitation of prior efforts that require considerable effort for model parameter selection, the objective was to provide a generalizable training dataset to avoid in-situ training. The problem was investigated for both low-resolution (e.g., 15 minute) and high-resolution data (60Hz). For the low-resolution data, we introduced a framework that first identify the similar households in the community based on energy behavior, and then used classifiers for appliance time-of-use inference. Neural network, SVM, random forest, and decision tree were tested to classify the appliance ‘On’, ‘Off’ events from the sub-intervals of energy load shapes. Results show average F-score of 83%, and 71% for EV and dryer across ten test households, solely by training the model on a set of neighbors in the community with known labels.

For the high-resolution data, we first introduced a self-tuning spectral clustering approach that first cluster the appliance signature without a priori information. The extracted clusters were used to create a processed library of appliance signature in the same environment. Thereafter, outlier detection was used on the whole-house energy samples to identify similar signatures to those in the library and thereby identifying new events. The evaluation on ~15000 events showed a high accuracy for event detection. Specifically, for major appliances for DR applications, AC, dishwasher, and washing machine had an F-score of higher than 90%.

For the third study, under the realistic uncertainty of load profiles, the impact of small-scale distributed elements (solar panels, storage, and smart loads) for improving the demand-supply problem was studied. Communities including more than 100 prosumer and 100 consumers in Austin, TX was used as a case-study. It was observed that communities with equal distribution of prosumers and consumers, ~75% of deficit energy in the community can be canceled out by prosumers at the highest PV peak hour. Furthermore, with 100% PV integration, the community can become entirely self-dependent during at high PV generation times (11am-3pm). The integration of other elements such as battery storage and smart appliances showed the potential of increasing the self-sufficiency of the community to around 85% and 75%, respectively. Furthermore, the user's flexibility showed to be a viable solution to improve the self-sufficiency by offering ~60% of what could be achieved by commercial battery storage.

7.2. Summary of contribution

In summary, this dissertation has contributed to adaptive operation of distributed energy management applications by leveraging the interactional energy consumption behavior of households through data analytics. The contributions of this work, which addressed the research questions, are outlined as follows:

(1) This study introduced a segmentation approach for ranking the households based on peak reduction potential. Furthermore, through community-level analysis, the flexibility of different major appliances, at different hours of the day were quantified. This actionable information could assist electric utilities for resources planning and program design for DR. (2) This study introduced machine learning-based event detection frameworks for identifying the major appliances time-of-use. To envision an scalable approach, the proposed solutions do not call for model parameter search, and the basis for model training/parameter search was merely based on automated

techniques or learning from similar households with known information. (3) This study investigated the role of small-scale distributed resources for improving the demand/supply match under the impact of uncertainty in energy behavior at the community-scale. The data-driven analysis could assist utilities and energy planners for resource allocation to improve the self-sufficiency of communities.

7.3. Future research directions

Future works of this research can address several questions, as elaborated here:

(1) Integrating the users' perception and how different subset of households respond to DR events could shed light on the potential of stratified engagement of households for DR application and reshaping the network demand. For our first study, we considered the user compliance factor by relying on previous community-level empirical studies. However, accounting for different levels of willingness in responding to the DR requests for buildings needs more comprehensive investigation. For instance, users with frequent and consistent energy patterns might be less willing to shift their loads, even with proper forms of incentives. Therefore, further investigations based on behavioral theories and pilot studies are important to associate the actual user compliance and historically observed consumption behavior.

(2) ML-based methods and optimizations techniques have high potential to improve the flexibility of the decentralized energy systems. For example, advanced forecasting techniques can be used for predicting the demand/generation of consumers/prosumers, and optimization techniques could be used to model the P2P exchange under non-flat pricing (i.e., dynamic pricing) to maximize the energy saving and minimize generation cost. Furthermore, the rise of secure and decentralized solutions like blockchain provide opportunities for modeling P2P exchange transactions. Through integrating these elements, future research could focus on developing P2P exchange solutions by minimizing the forecasting errors and maximizing the utilization of renewables for decentralized applications.

(3) Extreme events like power outage are undesired situations that impact grid reliability. Although this work mainly focused on community-level data for improved energy balancing of the network to potentially avoid those issues, looking at individual elements, i.e., buildings, can shed light on the smart operation of loads in case of facing these undesired events. Specifically, through the adoption of battery storage coupled with Home Energy Management (HEM) systems, modules

could be developed for adjusting the priority levels of individual loads operation. Accordingly, developing optimization models for HEM based on user priority is regarded as an active area of research.

(4) To further improve the model training for ToU inference, using larger samples for training could be helpful. Currently, commercial off-the-shelf smart plugs (e.g., Amazon smart plugs) are available as easily deployable solutions to record the information of individual loads. Therefore, through a larger adoption of these devices, the larger samples for training can be obtained, which in turn could improve the model performance. However, one open question is the private information that could be revealed by obtaining this information from households. Although the model training and analytics are carried out on a cloud server, it could still be subject to privacy leaks. Recent attempts (e.g., [273, 274]) have proposed privacy-preserving techniques to obfuscate energy data while still allowing for sophisticated machine learning solutions, and this is an ongoing area of research.

(5) Like any data-driven methods, our findings were presented based on evaluation of the considered datasets in this work. Nonetheless, we primarily used the samples from the Pecan Street Projects, as the most largest energy dataset publicly available to scholars. However, with the increased adoption of novel metering devices, more data will be available for further evaluation. Therefore, to increase the generalizability of the findings, future work could look at larger spatiotemporal scope, through investigation of seasonality impact and geographical diversity.

Chapter 8: References

[1] U. S. Environmental Protection Agency. *Electricity consumers*. Available: https://www.epa.gov/energy/electricity-customers. Last retrieved on June 22, 2020.

[2] U. S. Energy Information Administration. *U.S. Energy-Related Carbon Dioxide Emissions*. Available: https://www.eia.gov/environment/emissions/carbon/. Last retrieved on June 22, 2020.

[3] K. J. Lomas, "Carbon reduction in existing buildings: a transdisciplinary approach," ed: Taylor & Francis, 2010.

[4] U. S. D. o. Energy, "Energy Efficiency Trends in Residential and Commercial Buildings," 2008.

[5] L. Gelazanskas and K. A. Gamage, "Demand side management in smart grid: A review and proposals for future direction," *Sustainable Cities and Society*, vol. 11, pp. 22-30, 2014.

[6] F. Mwasilu, J. J. Justo, E.-K. Kim, T. D. Do, and J.-W. Jung, "Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration," *Renewable and sustainable energy reviews*, vol. 34, pp. 501-516, 2014.

[7] U. S. Energy Information Administration. (2016). *Today in energy*. Available: https://www.eia.gov/todayinenergy/. Last retrieved on June 22, 2020.

[8] U. S. Energy Information Administration. *California - State Energy Profile Overview* Available: https://www.eia.gov/state/?sid=CA. Last retrieved on June 22, 2020.

[9] U. S. Energy Information Administration, "Annual Energy Outlook 2019 with projections to 2050," 2019, Available: https://www.eia.gov/outlooks/aeo/pdf/aeo2019.pdf. Last retrieved on June 22, 2020.

[10] U. S. Energy Information Administration. (2018). *U.S. coal consumption in 2018 expected to be the lowest in 39 years*. Available: https://www.eia.gov/todayinenergy/detail.php?id=37692. Last retrieved on June 22, 2020.

[11] R. Tonkoski, D. Turcotte, and T. H. El-Fouly, "Impact of high PV penetration on voltage profiles in residential neighborhoods," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 3, pp. 518-527, 2012.

[12] S. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings, "Putting the" smarts" into the smart grid: A grand challenge for artificial intelligence," *Communications of the ACM*, vol. 55, no. 4, pp. 86-97, 2012.

[13] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420-430, 2014.

[14] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, 2018.

[15] U. S. Energy Information Administration. (2017). *Nearly half of all U.S. electricity customers have smart meters*. Available: https://www.eia.gov/todayinenergy/detail.php?id=34012. Last retrieved on June 22, 2020.

[16] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Applied energy*, vol. 141, pp. 190-199, 2015.

[17] G. J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, "Two-stage pattern recognition of load curves for classification of electricity customers," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1120-1128, 2007.

[18] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, no. 1, pp. 68-80, 2012.

[19] G. Tsekouras, P. Kotoulas, C. Tsirekis, E. Dialynas, and N. Hatziargyriou, "A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers," *Electric Power Systems Research*, vol. 78, no. 9, pp. 1494-1510, 2008.

[20] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *IEEE Transactions on power systems*, vol. 20, no. 2, pp. 596-602, 2005.

[21] D. Hsu, "Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data," *Applied Energy*, vol. 160, pp. 153-163, 2015.
[22] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 933-940, 2006.
[23] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, "Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data," *Applied Energy*, vol. 87, no. 11, pp. 3538-3545, 2010.
[24] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Load pattern-based classification of electricity customers," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1232-1239, 2004.
[25] I. P. Panapakidis, T. A. Papadopoulos, G. C. Christoforidis, and G. K. Papagiannis, "Pattern recognition algorithms for electricity load curve analysis of buildings," *Energy and Buildings*, vol. 73, pp. 137-145, 2014.
[26] B. P. Bhattarai, S. Paudyal, Y. Luo, M. Mohanpurkar, K. Cheung, R. Tonkoski, R. Hovsapian, K. S. Myers, R. Zhang, and P. Zhao, "Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions," *IET Smart Grid*, 2019.
[27] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, vol. 52, pp. 213-234, 2013.
[28] D. Parker, D. Hoak, A. Meier, and R. Brown, "How much energy are we using? Potential of residential energy demand feedback devices," *Proceedings of the 2006 Summer Study on Energy Efficiency in Buildings, American Council for an Energy Efficient Economy, Asilomar, CA*, 2006.
[29] B. Neenan, J. Robinson, and R. Boisvert, "Residential electricity use feedback: A research synthesis and economic framework," *Electric Power Research Institute*, vol. 3, 2009.
[30] K. Herter and S. Wayland, "Behavioral Experimentation with Residential Energy Feedback Through Simulation Gaming," *Technical ReportsCEC-500-2009-XXX, California Energy Commission, PIER Buildings End-Use Energy Efficiency Program*, 2009.
[31] R. G. Pratt, P. J. Balducci, C. Gerkensmeyer, S. Katipamula, M. C. Kintner-Meyer, T. F. Sanquist, K. P. Schneider, and T. J. Secrest, "The smart grid: an estimation of the energy and CO2 benefits," Pacific Northwest National Laboratory (PNNL), Richland, WA (US)2010.
[32] S. Darby, "The effectiveness of feedback on energy consumption," *A Review for DEFRA of the Literature on Metering, Billing and direct Displays*, vol. 486, no. 2006, 2006.
[33] C. Fischer, "Feedback on household electricity consumption: a tool for saving energy?," *Energy efficiency*, vol. 1, no. 1, pp. 79-104, 2008.
[34] B. Neupane, T. B. Pedersen, and B. Thiesson, "Towards flexibility detection in device-level energy consumption," in *International Workshop on Data Analytics for Renewable Energy Integration*, 2014, pp. 1-16: Springer.
[35] E. Erell, B. A. Portnov, and M. Assif, "Modifying behaviour to save energy at home is harder than we think…," *Energy and Buildings*, vol. 179, pp. 384-398, 2018.
[36] H. Rashid, P. M. Mammen, S. Singh, K. Ramamritham, P. Singh, and P. Shenoy, "Want to reduce energy consumption?: don't depend on the consumers!," in *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*, 2017, p. 16: ACM.
[37] M. Afzalan and F. Jazizadeh, "Residential loads flexibility potential for demand response using energy consumption patterns and user segments," *Applied Energy*, vol. 254, p. 113693, 2019.
[38] M. Afzalan and F. Jazizadeh, "Data-driven identification of consumers with deferrable loads for demand response programs," *IEEE Embedded Systems Letters*, 2019.
[39] M. Afzalan and F. Jazizadeh, "A Machine Learning Framework to Infer Time-of-Use of Flexible Loads: Resident Behavior Learning for Demand Response," *IEEE Access*, vol. 8, 2020.
[40] M. Afzalan and F. Jazizadeh, "An automated spectral clustering for multi-scale data," *Neurocomputing*, vol. 347, pp. 94-108, 2019.
[41] M. Afzalan, F. Jazizadeh, and J. Wang, "Self-configuring event detection in electricity monitoring for human-building interaction," *Energy and Buildings*, vol. 187, pp. 95-109, 2019.

[42] P. Basak, S. Chowdhury, S. H. nee Dey, and S. Chowdhury, "A literature review on integration of distributed energy resources in the perspective of control, protection and stability of microgrid," *Renewable and Sustainable Energy Reviews,* vol. 16, no. 8, pp. 5545-5556, 2012.
[43] P. Alstone, D. Gershenson, and D. M. Kammen, "Decentralized energy systems for clean electricity access," *Nature Climate Change,* vol. 5, no. 4, p. 305, 2015.
[44] R. D'hulst, W. Labeeuw, B. Beusen, S. Claessens, G. Deconinck, and K. Vanthournout, "Demand response flexibility and flexibility potential of residential smart appliances: Experiences from large pilot test in Belgium," *Applied Energy,* vol. 155, pp. 79-90, 2015.
[45] B. Sparn, X. Jin, and L. Earle, "Laboratory Testing of Demand-Response Enabled Household Appliances," National Renewable Energy Lab.(NREL), Golden, CO (United States)2013.
[46] A. Zipperer, P. A. Aloise-Young, S. Suryanarayanan, R. Roche, L. Earle, D. Christensen, P. Bauleo, and D. Zimmerle, "Electric energy management in the smart home: Perspectives on enabling technologies and consumer behavior," *Proceedings of the IEEE,* vol. 101, no. 11, pp. 2397-2408, 2013.
[47] M. Pipattanasomporn, M. Kuzlu, and S. Rahman, "An algorithm for intelligent home energy management and demand response analysis," *IEEE Transactions on Smart Grid,* vol. 3, no. 4, pp. 2166-2173, 2012.
[48] K. Vanthournout, B. Dupont, W. Foubert, C. Stuckens, and S. Claessens, "An automated residential demand response pilot experiment, based on day-ahead dynamic pricing," *Applied Energy,* vol. 155, pp. 195-203, 2015.
[49] H. Shareef, M. S. Ahmed, A. Mohamed, and E. Al Hassan, "Review on home energy management system considering demand responses, smart technologies, and intelligent controllers," *IEEE Access,* vol. 6, pp. 24498-24509, 2018.
[50] O. Erdinc, N. G. Paterakis, T. D. Mendes, A. G. Bakirtzis, and J. P. Catalão, "Smart household operation considering bi-directional EV and ESS utilization by real-time pricing-based DR," *IEEE Transactions on Smart Grid,* vol. 6, no. 3, pp. 1281-1291, 2014.
[51] M. H. Albadi and E. F. El-Saadany, "A summary of demand response in electricity markets," *Electric power systems research,* vol. 78, no. 11, pp. 1989-1996, 2008.
[52] H. T. Haider, O. H. See, and W. Elmenreich, "A review of residential demand response of smart grid," *Renewable and Sustainable Energy Reviews,* vol. 59, pp. 166-178, 2016.
[53] D. J. Hammerstrom, R. Ambrosio, T. A. Carlon, J. G. DeSteese, G. R. Horst, R. Kajfasz, L. L. Kiesling, P. Michie, R. G. Pratt, and M. Yao, "Pacific Northwest GridWise™ Testbed Demonstration Projects; Part I. Olympic Peninsula Project," Pacific Northwest National Lab.(PNNL), Richland, WA (United States)2008.
[54] C. B. Kobus, E. A. Klaassen, R. Mugge, and J. P. Schoormans, "A real-life assessment on the effect of smart appliances for shifting households' electricity demand," *Applied Energy,* vol. 147, pp. 335-343, 2015.
[55] S. Nistor, J. Wu, M. Sooriyabandara, and J. Ekanayake, "Capability of smart appliances to provide reserve services," *Applied Energy,* vol. 138, pp. 590-597, 2015.
[56] J. Widén, "Improved photovoltaic self-consumption with appliance scheduling in 200 single-family buildings," *Applied Energy,* vol. 126, pp. 199-212, 2014.
[57] R. Yin, E. C. Kara, Y. Li, N. DeForest, K. Wang, T. Yong, and M. Stadler, "Quantifying flexibility of commercial and residential loads for demand response using setpoint changes," *Applied Energy,* vol. 177, pp. 149-164, 2016.
[58] T. Teeraratkul, D. O'Neill, and S. Lall, "Shape-based approach to household electric load curve clustering and prediction," *IEEE Transactions on Smart Grid,* vol. 9, no. 5, pp. 5196-5206, 2018.
[59] P. Bradley, M. Leach, and J. Torriti, "A review of the costs and benefits of demand response for electricity in the UK," *Energy Policy,* vol. 52, pp. 312-327, 2013.
[60] P. Guo, V. O. Li, and J. C. Lam, "Smart demand response in China: Challenges and drivers," *Energy Policy,* vol. 107, pp. 1-10, 2017.

[61] P. Guo, J. C. Lam, and V. O. Li, "Drivers of domestic electricity users' price responsiveness: A novel machine learning approach," *Applied energy,* vol. 235, pp. 900-913, 2019.
[62] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE transactions on industrial informatics,* vol. 7, no. 3, pp. 381-388, 2011.
[63] P. H. Nardelli and F. Kühnlenz, "Why Smart Appliances May Result in a Stupid Energy Grid?," *arXiv preprint arXiv:1703.01195,* 2017.
[64] L. Stankovic, V. Stankovic, J. Liao, and C. Wilson, "Measuring the energy intensity of domestic activities from smart meter data," *Applied energy,* vol. 183, pp. 1565-1580, 2016.
[65] D. Frazzetto, B. Neupane, T. B. Pedersen, and T. D. Nielsen, "Adaptive User-Oriented Direct Load-Control of Residential Flexible Devices," *arXiv preprint arXiv:1805.05470,* 2018.
[66] J. Kwac, J. I. Kim, and R. Rajagopal, "Efficient customer selection process for various DR objectives," *IEEE Transactions on Smart Grid,* 2017.
[67] P. M. Mammen, H. Kumar, K. Ramamritham, and H. Rashid, "Want to Reduce Energy Consumption, Whom should we call?," in *Proceedings of the Ninth International Conference on Future Energy Systems*, 2018, pp. 12-20: ACM.
[68] Source: Pecan Street Dataport. 2017. [Online]. Available: https://dataport.pecanstreet.org/.
[69] F. Elghitani and W. Zhuang, "Aggregating a large number of residential appliances for demand response applications," *IEEE Transactions on Smart Grid,* vol. 9, no. 5, pp. 5092-5100, 2018.
[70] E. Klaassen, C. Kobus, J. Frunt, and J. Slootweg, "Responsiveness of residential electricity demand to dynamic tariffs: Experiences from a large field test in the Netherlands," *Applied Energy,* vol. 183, pp. 1065-1074, 2016.
[71] A. Malik, N. Haghdadi, I. MacGill, and J. Ravishankar, "Appliance level data analysis of summer demand reduction potential from residential air conditioner control," *Applied energy,* vol. 235, pp. 776-785, 2019.
[72] E. C. Kara, M. D. Tabone, J. S. MacDonald, D. S. Callaway, and S. Kiliccote, "Quantifying flexibility of residential thermostatically controlled loads for demand response: a data-driven approach," in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, 2014, pp. 140-147: ACM.
[73] P. Srikantha, C. Rosenberg, and S. Keshav, "An analysis of peak demand reductions due to elasticity of domestic appliances," in *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet*, 2012, p. 28: ACM.
[74] F. Fernandes, H. Morais, Z. Vale, and C. Ramos, "Dynamic load management in a smart home to participate in demand response events," *Energy and Buildings,* vol. 82, pp. 592-606, 2014.
[75] A. Di Giorgio and L. Pimpinella, "An event driven smart home controller enabling consumer economic saving and automated demand side management," *Applied Energy,* vol. 96, pp. 92-103, 2012.
[76] P. Du and N. Lu, "Appliance commitment for household load scheduling," *IEEE transactions on Smart Grid,* vol. 2, no. 2, pp. 411-419, 2011.
[77] Z. Yu, L. Jia, M. C. Murphy-Hoye, A. Pratt, and L. Tong, "Modeling and stochastic control for home energy management," *IEEE Transactions on Smart Grid,* vol. 4, no. 4, pp. 2244-2255, 2013.
[78] A. Saha, M. Kuzlu, M. Pipattanasomporn, S. Rahman, O. Elma, U. S. Selamogullari, M. Uzunoglu, and B. Yagcitekin, "A robust building energy management algorithm validated in a smart house environment," *Intelligent Industrial Systems,* vol. 1, no. 2, pp. 163-174, 2015.
[79] S. Althaher, P. Mancarella, and J. Mutale, "Automated demand response from home energy management system under dynamic pricing and power and comfort constraints," *IEEE Transactions on Smart Grid,* vol. 6, no. 4, pp. 1874-1883, 2015.
[80] X. Jin, K. Baker, D. Christensen, and S. Isley, "Foresee: A user-centric home energy management system for energy efficiency and demand response," *Applied energy,* vol. 205, pp. 1583-1595, 2017.

[81] Z. Wang and R. Paranjape, "Optimal residential demand response for multiple heterogeneous homes with real-time price prediction in a multiagent framework," *IEEE transactions on smart grid*, vol. 8, no. 3, pp. 1173-1184, 2015.
[82] N. G. Paterakis, O. Erdinc, A. G. Bakirtzis, and J. P. Catalão, "Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 6, pp. 1509-1519, 2015.
[83] P. Yi, X. Dong, A. Iwayemi, C. Zhou, and S. Li, "Real-time opportunistic scheduling for residential demand response," *IEEE Transactions on smart grid*, vol. 4, no. 1, pp. 227-234, 2013.
[84] J. Kwac and R. Rajagopal, "Demand response targeting using big data analytics," in *Big Data, 2013 IEEE International Conference on*, 2013, pp. 683-690: IEEE.
[85] V. Chandan, T. Ganu, T. K. Wijaya, M. Minou, G. Stamoulis, G. Thanos, and D. P. Seetharam, "idr: Consumer and grid friendly demand response system," in *Proceedings of the 5th international conference on Future energy systems*, 2014, pp. 183-194: ACM.
[86] J. Kwac, J. Flora, and R. Rajagopal, "Lifestyle segmentation based on energy consumption data," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2409-2418, 2018.
[87] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of electricity consumption behavior dynamics toward big data applications," *IEEE transactions on smart grid*, vol. 7, no. 5, pp. 2437-2447, 2016.
[88] T. K. Wijaya, T. Ganu, D. Chakraborty, K. Aberer, and D. P. Seetharam, "Consumer segmentation and knowledge extraction from smart meter and survey data," in *Proceedings of the 2014 SIAM International Conference on Data Mining*, 2014, pp. 226-234: SIAM.
[89] J. C. Holyhead, S. D. Ramchurn, and A. Rogers, "Consumer targeting in residential demand response programmes," in *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*, 2015, pp. 7-16: ACM.
[90] Y. Ji and R. Rajagopal, "Demand and flexibility of residential appliances: An empirical analysis," in *Signal and Information Processing (GlobalSIP), 2017 IEEE Global Conference on*, 2017, pp. 1020-1024: IEEE.
[91] S. Lin, F. Li, E. Tian, Y. Fu, and D. Li, "Clustering load profiles for demand response applications," *IEEE Transactions on Smart Grid*, 2017.
[92] B. P. Esther and K. S. Kumar, "A survey on residential demand side management architecture, approaches, optimization models and methods," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 342-351, 2016.
[93] M. Afzalan and F. Jazizadeh, "Efficient integration of smart appliances for demand response programs," in *Proceedings of the 5th Conference on Systems for Built Environments*, 2018, pp. 29-32: ACM.
[94] H. Rashid, P. Singh, and K. Ramamritham, "Revisiting selection of residential consumers for demand response programs," in *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*, 2017, p. 30: ACM.
[95] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. Jennings, "Agent-based control for decentralised demand side management in the smart grid," in *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, 2011, pp. 5-12: International Foundation for Autonomous Agents and Multiagent Systems.
[96] B. Yildiz, J. Bilbao, J. Dore, and A. Sproul, "Recent advances in the analysis of residential electricity consumption and applications of smart meter data," *Applied Energy*, vol. 208, pp. 402-427, 2017.
[97] Michigan New Energy Policy. *Common Demand Response Practices and Program Designs*. Available: https://www.michigan.gov/documents/energy/Common_Practices_Feb22_522983_7.pdf. Last retrieved on June 22, 2020.
[98] N. C. Truong, J. McInerney, L. Tran-Thanh, E. Costanza, and S. D. Ramchurn, "Forecasting multi-appliance usage for smart home energy management," 2013.

[99] Draft Logic. List of the Power Consumption of Typical Household Appliances. Available: https://www.daftlogic.com/information-appliance-power-consumption.htm. Last retrieved on June 22, 2020.
[100] NILM wiki. Available: http://wiki.nilm.eu/appliance.html. Last retrieved on June 22, 2020.
[101] G. Koutitas, "Control of flexible smart devices in the smart grid," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1333-1343, 2012.
[102] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, 1998, pp. 59-66: IEEE.
[103] U. S. Census Bureau. *QuickFacts, Austin city, Texas*. Available: https://www.census.gov/quickfacts/fact/table/austincitytexas/LND110210. Last retrieved on June 22, 2020.
[104] J.-H. Kim and A. Shcherbakova, "Common failures of demand response," *Energy*, vol. 36, no. 2, pp. 873-880, 2011.
[105] A. Munshi and Y. A.-R. Mohamed, "Unsupervised Non-Intrusive Extraction of Electrical Vehicle Charging Load Patterns," *IEEE Transactions on Industrial Informatics*, 2018.
[106] Z. Zhang, J. H. Son, Y. Li, M. Trayer, Z. Pi, D. Y. Hwang, and J. K. Moon, "Training-free non-intrusive load monitoring of electric vehicle charging with low sampling rate," *arXiv preprint arXiv:1404.5020*, 2014.
[107] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. Zhao, and F. Luo, "A hierarchical hidden markov model framework for home appliance modeling," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3079-3090, 2018.
[108] Y. Liu, X. Wang, and W. You, "Non-intrusive Load Monitoring by Voltage-Current Trajectory Enabled Transfer Learning," *IEEE Transactions on Smart Grid*, 2018.
[109] S. Iyengar, S. Lee, D. Irwin, and P. Shenoy, "Analyzing energy usage on a city-scale using utility smart meters," in *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, 2016, pp. 51-60: ACM.
[110] H. Rashid, P. Singh, V. Stankovic, and L. Stankovic, "Can non-intrusive load monitoring be used for identifying an appliance's anomalous behaviour?," *Applied Energy*, vol. 238, pp. 796-805, 2019.
[111] O. Alrumayh and K. Bhattacharya, "Flexibility of Residential Loads for Demand Response Provisions in Smart Grid," *IEEE Transactions on Smart Grid*, 2019.
[112] J. Kwac, J. Flora, and R. Rajagopal, "Lifestyle segmentation based on energy consumption data," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2409-2418, 2016.
[113] M. S. Piscitelli, S. Brandi, and A. Capozzoli, "Recognition and classification of typical load profiles in buildings with non-intrusive learning approach," *Applied Energy*, vol. 255, p. 113727, 2019.
[114] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841-851, 2017.
[115] M. Afzalan and F. Jazizadeh, "Investigating the Appliance Use Patterns on the Households' Electricity Load Shapes from Smart Meters," in *Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience*: American Society of Civil Engineers Reston, VA, 2019, pp. 154-161.
[116] K. Cetin, P. Tabares-Velasco, and A. Novoselac, "Appliance daily energy use in new residential buildings: Use profiles and variation in time-of-use," *Energy and Buildings*, vol. 84, pp. 716-726, 2014.
[117] Y. Ji and R. Rajagopal, "Demand and flexibility of residential appliances: An empirical analysis," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2017, pp. 1020-1024: IEEE.
[118] S. Lin, F. Li, E. Tian, Y. Fu, and D. Li, "Clustering load profiles for demand response applications," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1599-1607, 2017.

[119] K. Basu, L. Hawarah, N. Arghira, H. Joumaa, and S. Ploix, "A prediction system for home appliance usage," *Energy and Buildings,* vol. 67, pp. 668-679, 2013.
[120] A. Barbato, A. Capone, M. Rodolfi, and D. Tagliaferri, "Forecasting the usage of household appliances through power meter sensors for demand management in the smart grid," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 404-409: IEEE.
[121] J. Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, 2015, pp. 55-64: ACM.
[122] K. S. Barsim and B. Yang, "On the feasibility of generic deep disaggregation for single-load extraction," *arXiv preprint arXiv:1802.02139*, 2018.
[123] R. Bonfigli, A. Felicetti, E. Principi, M. Fagiani, S. Squartini, and F. Piazza, "Denoising autoencoders for Non-Intrusive Load Monitoring: Improvements and comparative evaluation," *Energy and Buildings,* vol. 158, pp. 1461-1474, 2018.
[124] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, and D. Deschrijver, "Appliance classification using VI trajectories and convolutional neural networks," *Energy and Buildings*, vol. 158, pp. 32-36, 2018.
[125] F. Jazizadeh, M. Afzalan, B. Becerik-Gerber, and L. Soibelman, "EMBED: A Dataset for Energy Monitoring through Building Electricity Disaggregation," in *Proceedings of the Ninth International Conference on Future Energy Systems*, 2018, pp. 230-235: ACM.
[126] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE transactions on Consumer Electronics,* vol. 57, no. 1, pp. 76-84, 2011.
[127] A. Zoha, A. Gluhak, M. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors,* vol. 12, no. 12, pp. 16838-16866, 2012.
[128] S. M. Tabatabaei, S. Dick, and W. Xu, "Toward non-intrusive load monitoring via multi-label classification," *IEEE Transactions on Smart Grid,* vol. 8, no. 1, pp. 26-40, 2017.
[129] M. Berges, E. Goldman, H. S. Matthews, L. Soibelman, and K. Anderson, "User-centered nonintrusive electricity load monitoring for residential buildings," *Journal of Computing in Civil Engineering,* vol. 25, no. 6, pp. 471-480, 2011.
[130] F. Jazizadeh, B. Becerik-Gerber, M. Berges, and L. Soibelman, "An unsupervised hierarchical clustering based heuristic algorithm for facilitated training of electricity consumption disaggregation systems," *Advanced Engineering Informatics,* vol. 28, no. 4, pp. 311-326, 2014.
[131] F. Jazizadeh, B. Becerik-Gerber, M. Berges, and L. Soibelman, "Unsupervised clustering of residential electricity consumption measurements for facilitated user-centric non-intrusive load monitoring," *Comput. Civ. Build. Eng,* pp. 1869-1876, 2014.
[132] J. Z. Kolter, S. Batra, and A. Y. Ng, "Energy disaggregation via discriminative sparse coding," in *Advances in Neural Information Processing Systems*, 2010, pp. 1153-1161.
[133] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, and D. Deschrijver, "On the Bayesian optimization and robustness of event detection methods in NILM," *Energy and Buildings,* vol. 145, pp. 57-66, 2017.
[134] M. Gaur and A. Majumdar, "Disaggregating Transform Learning for Non-Intrusive Load Monitoring," *IEEE Access,* vol. 6, pp. 46256-46265, 2018.
[135] N. Batra, A. Singh, and K. Whitehouse, "Gemello: Creating a detailed energy breakdown from just the monthly electricity bill," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 431-440: ACM.
[136] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-Markov models," *Journal of Machine Learning Research,* vol. 14, no. Feb, pp. 673-701, 2013.
[137] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," in *Artificial Intelligence and Statistics*, 2012, pp. 1472-1482.

[138] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. Zhao, and F. Luo, "A hierarchical hidden Markov model framework for home appliance modeling," *IEEE Transactions on Smart Grid,* vol. 9, no. 4, pp. 3079-3090, 2016.
[139] N. Batra, H. Wang, A. Singh, and K. Whitehouse, "Matrix factorisation for scalable energy breakdown," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
[140] F. E. R. Commission, "Assessment of demand response and advanced metering," 2008.
[141] Pecan Street, "Dataport: the world's largest energy data resource," *Pecan Street Inc,* 2015.
[142] M. Afzalan and F. Jazizadeh, "Semantic search in household energy consumption segmentation through descriptive characterization," in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 2019, pp. 263-266.
[143] T. Hasanin, T. M. Khoshgoftaar, J. L. Leevy, and N. Seliya, "Examining characteristics of predictive models with imbalanced big data," *Journal of Big Data,* vol. 6, no. 1, p. 69, 2019.
[144] T. Hasanin, T. M. Khoshgoftaar, J. L. Leevy, and R. A. Bauder, "Severely imbalanced Big Data challenges: investigating data sampling approaches," *Journal of Big Data,* vol. 6, no. 1, p. 107, 2019.
[145] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research,* vol. 16, pp. 321-357, 2002.
[146] J. Kwac and R. Rajagopal, "Data-driven targeting of customers for demand response," *IEEE Transactions on Smart Grid,* vol. 7, no. 5, pp. 2199-2207, 2015.
[147] M. Pipattanasomporn, M. Kuzlu, S. Rahman, and Y. Teklu, "Load profiles of selected major household appliances and their demand response opportunities," *IEEE Transactions on Smart Grid,* vol. 5, no. 2, pp. 742-750, 2013.
[148] K. Knezović, M. Marinelli, P. Codani, and Y. Perez, "Distribution grid services and flexibility provision by electric vehicles: A review of options," in *2015 50th International Universities Power Engineering Conference (UPEC)*, 2015, pp. 1-6: IEEE.
[149] J. Y. Park, E. Wilson, A. Parker, and Z. Nagy, "The good, the bad, and the ugly: Data-driven load profile discord identification in a large building portfolio," *Energy and Buildings,* vol. 215, p. 109892, 2020.
[150] A. Capozzoli, M. S. Piscitelli, S. Brandi, D. Grassi, and G. Chicco, "Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings," *Energy,* vol. 157, pp. 336-352, 2018.
[151] B. Gao, T.-Y. Liu, T. Qin, X. Zheng, Q.-S. Cheng, and W.-Y. Ma, "Web image clustering by consistent utilization of visual features and surrounding texts," in *Proceedings of the 13th annual ACM international conference on Multimedia*, 2005, pp. 112-121: ACM.
[152] Y. Yang, D. Xu, F. Nie, S. Yan, and Y. Zhuang, "Image clustering using local discriminant models and global integration," *IEEE Transactions on Image Processing,* vol. 19, no. 10, pp. 2761-2773, 2010.
[153] J. Cao and H. Li, "Energy-efficient structuralized clustering for sensor-based cyber physical systems," in *Ubiquitous, Autonomic and Trusted Computing, 2009. UIC-ATC'09. Symposia and Workshops on*, 2009, pp. 234-239: IEEE.
[154] K. Y. Yeung and W. L. Ruzzo, "Principal component analysis for clustering gene expression data," *Bioinformatics,* vol. 17, no. 9, pp. 763-774, 2001.
[155] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, no. 8, pp. 888-905, 2000.
[156] Y. Weiss, "Segmentation using eigenvectors: a unifying view," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, 1999, vol. 2, pp. 975-982: IEEE.
[157] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Advances in neural information processing systems,* vol. 2, pp. 849-856, 2002.
[158] F. Lauer and C. Schnörr, "Spectral clustering of linear subspaces for motion segmentation," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 678-685: IEEE.

[159] X. Zhang, L. Jiao, F. Liu, L. Bo, and M. Gong, "Spectral clustering ensemble applied to SAR image segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2126-2136, 2008.
[160] F. R. Bach and M. I. Jordan, "Learning spectral clustering, with application to speech separation," *Journal of Machine Learning Research*, vol. 7, no. Oct, pp. 1963-2001, 2006.
[161] A. Paccanaro, C. Chennubhotla, J. Casbon, and M. Saqi, "Spectral clustering of protein sequences," in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, 2003, vol. 4, pp. 3083-3088: IEEE.
[162] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *NIPS*, 2004, vol. 17, no. 1601-1608, p. 16.
[163] T. Xiang and S. Gong, "Spectral clustering with eigenvector selection," *Pattern Recognition*, vol. 41, no. 3, pp. 1012-1029, 2008.
[164] F. Zhao, L. Jiao, H. Liu, X. Gao, and M. Gong, "Spectral clustering with eigenvector selection based on entropy ranking," *Neurocomputing*, vol. 73, no. 10, pp. 1704-1717, 2010.
[165] G. Sanguinetti, J. Laidler, and N. D. Lawrence, "Automatic determination of the number of clusters using spectral algorithms," in *Machine Learning for Signal Processing, 2005 IEEE Workshop on*, 2005, pp. 55-60: IEEE.
[166] M. Steinbach, L. Ertöz, and V. Kumar, "The challenges of clustering high dimensional data," in *New Directions in Statistical Physics*: Springer, 2004, pp. 273-309.
[167] H. Goncalves, A. Ocneanu, M. Berges, and R. Fan, "Unsupervised disaggregation of appliances using aggregated consumption data," in *The 1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, 2011.
[168] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395-416, 2007.
[169] D. Verma and M. Meila, "A comparison of spectral clustering algorithms," *University of Washington Tech Rep UWCSE030501*, vol. 1, pp. 1-18, 2003.
[170] F. Tung, A. Wong, and D. A. Clausi, "Enabling scalable spectral clustering for image segmentation," *Pattern Recognition*, vol. 43, no. 12, pp. 4069-4076, 2010.
[171] R. Mall, R. Langone, and J. A. Suykens, "Self-tuned kernel spectral clustering for large scale networks," in *Big Data, 2013 IEEE International Conference on*, 2013, pp. 385-393: IEEE.
[172] T. Jebara, Y. Song, and K. Thadani, "Spectral clustering and embedding with hidden markov models," in *European Conference on Machine Learning*, 2007, pp. 164-175: Springer.
[173] Z. Li, J. Liu, S. Chen, and X. Tang, "Noise robust spectral clustering," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1-8: IEEE.
[174] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, "Spectral grouping using the Nystrom method," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 2, pp. 214-225, 2004.
[175] K. Taşdemir, "Vector quantization based approximate spectral clustering of large datasets," *Pattern Recognition*, vol. 45, no. 8, pp. 3034-3044, 2012.
[176] X. Chen and D. Cai, "Large Scale Spectral Clustering with Landmark-Based Representation," in *AAAI*, 2011, vol. 5, no. 7, p. 14.
[177] J. Cao, P. Chen, Q. Dai, and W.-K. Ling, "Local information-based fast approximate spectral clustering," *Pattern Recognition Letters*, vol. 38, pp. 63-69, 2014.
[178] J. Yin and Q. Yang, "Integrating hidden Markov models and spectral analysis for sensory time series clustering," in *Data Mining, Fifth IEEE International Conference on*, 2005, p. 8 pp.: IEEE.
[179] X. Li, W. Hu, C. Shen, A. Dick, and Z. Zhang, "Context-aware hypergraph construction for robust spectral clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2588-2597, 2014.
[180] J. Li, Y. Xia, Z. Shan, and Y. Liu, "Scalable constrained spectral clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 2, pp. 589-593, 2015.

[181] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1650-1654, 2002.
[182] M. Meila and J. Shi, "Learning segmentation by random walks," in *NIPS*, 2000, vol. 14.
[183] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 5, pp. 603-619, 2002.
[184] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 17, no. 8, pp. 790-799, 1995.
[185] P. Perona and W. Freeman, "A factorization approach to grouping," *Computer Vision—ECCV'98*, pp. 655-670, 1998.
[186] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433-459, 2010.
[187] M. Hosseini and F. T. Azar, "A new eigenvector selection strategy applied to develop spectral clustering," *Multidimensional Systems and Signal Processing*, vol. 28, no. 4, pp. 1227-1248, 2017.
[188] X.-Y. Li and L.-J. Guo, "Constructing affinity matrix in spectral clustering based on neighbor propagation," *Neurocomputing*, vol. 97, pp. 125-130, 2012.
[189] N. Tremblay, G. Puy, R. Gribonval, and P. Vandergheynst, "Compressive spectral clustering," in *International Conference on Machine Learning*, 2016, pp. 1002-1011.
[190] Y. Li, J. Huang, and W. Liu, "Scalable Sequential Spectral Clustering," in *AAAI*, 2016, pp. 1809-1815.
[191] M. Meila, "Spectral Clustering: a Tutorial for the 2010's," ed: CRC Press, 2016, pp. 1-23.
[192] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," in *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, 2010, pp. 911-916: IEEE.
[193] J. D. Hamilton, *Time series analysis*. Princeton university press Princeton, 1994.
[194] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, M. J. Er, W. Ding, and C.-T. Lin, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664-681, 2017.
[195] T. W. Liao, "Clustering of time series data—a survey," *Pattern recognition*, vol. 38, no. 11, pp. 1857-1874, 2005.
[196] A. Little and A. Byrd, "A Multiscale Spectral Method for Learning Number of Clusters," in *Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on*, 2015, pp. 457-460: IEEE.
[197] W. Ye, S. Goebl, C. Plant, and C. Böhm, "Fuse: Full spectral clustering," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1985-1994: ACM.
[198] A. Damle, V. Minden, and L. Ying, "Robust and efficient multi-way spectral clustering," *arXiv preprint arXiv:1609.08251*, 2016.
[199] Energy Information Administration (2017). *Electricty Explained-Use of Electricty*. Available: https://www.eia.gov/energyexplained/index.php?page=electricity_use.
[200] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, and S. Patel, "Disaggregated end-use energy sensing for the smart grid," *IEEE Pervasive Computing*, vol. 10, no. 1, pp. 28-39, 2011.
[201] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. D. Abowd, "At the flick of a switch: Detecting and classifying unique electrical events on the residential power line," *Lecture Notes in Computer Science*, vol. 4717, pp. 271-288, 2007.
[202] T. Hargreaves, M. Nye, and J. Burgess, "Keeping energy visible? Exploring how householders interact with feedback from smart energy monitors in the longer term," *Energy policy*, vol. 52, pp. 126-134, 2013.
[203] E. Hailemariam, R. Goldstein, R. Attar, and A. Khan, "Real-time occupancy detection using decision trees with multiple sensor types," 2011, pp. 141-148: Society for Computer Simulation International.

[204] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, "Occupancy Detection from Electricity Consumption Data," presented at the Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings, Roma, Italy, 2013.
[205] W. Kleiminger, C. Beckel, and S. Santini, "Household occupancy monitoring using electricity meters," 2015, pp. 975-986: ACM.
[206] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, and P. Shenoy, "Non-Intrusive Occupancy Monitoring using Smart Meters," presented at the Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings, Roma, Italy, 2013.
[207] A. Akbar, M. Nati, F. Carrez, and K. Moessner, "Contextual occupancy detection for smart office by pattern recognition of electricity consumption data," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 561-566.
[208] R. Brown, N. Ghavami, M. Adjrad, M. Ghavami, and S. Dudley, "Occupancy based household energy disaggregation using ultra wideband radar and electrical signature profiles," *Energy and Buildings*, vol. 141, pp. 134-141, 2017.
[209] K. X. Perez, M. Baldea, and T. F. Edgar, "Integrated HVAC management and optimal scheduling of smart appliances for community peak load reduction," *Energy and Buildings*, vol. 123, pp. 34-40, 2016.
[210] M. Zeifman, C. Akers, and K. Roth, "Nonintrusive Appliance Load Monitoring (NIALM) for Energy Control in Residential Buildings: Review and Outlook," in *IEEE transactions on Consumer Electronics*, 2011.
[211] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE transactions on Consumer Electronics*, vol. 57, no. 1, 2011.
[212] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838-16866, 2012.
[213] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli, "A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring," *Applied Energy*, vol. 185, pp. 331-344, 2017.
[214] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, "Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models," *Applied Energy*, vol. 208, pp. 1590-1607, 2017.
[215] M. Bedir, E. Hasselaar, and L. Itard, "Determinants of electricity consumption in Dutch dwellings," *Energy and buildings*, vol. 58, pp. 194-207, 2013.
[216] R. V. Jones and K. J. Lomas, "Determinants of high electrical energy demand in UK homes: Appliance ownership and use," *Energy and Buildings*, vol. 117, pp. 71-82, 2016.
[217] G. Flett and N. Kelly, "A disaggregated, probabilistic, high resolution method for assessment of domestic occupancy and electrical demand," *Energy and Buildings*, vol. 140, pp. 171-187, 2017.
[218] S. De Lauretis, F. Ghersi, and J.-M. Cayla, "Energy consumption and activity patterns: an analysis extended to total time and energy use for French households," *Applied Energy*, vol. 206, pp. 634-648, 2017.
[219] S. Ahmadi-Karvigh, B. Becerik-Gerber, and L. Soibelman, "A framework for allocating personalized appliance-level disaggregated electricity consumption to daily activities," *Energy and Buildings*, vol. 111, pp. 337-350, 2016.
[220] A. Faustine, N. H. Mvungi, S. Kaijage, and K. Michael, "A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem," *arXiv preprint arXiv:1703.00785*, 2017.
[221] D. Luo, L. K. Norford, S. R. Shaw, and S. B. Leeb, "Monitoring HVAC equipment electrical loads from a centralized location--methods and field test results/Discussion," *ASHRAE Transactions*, vol. 108, p. 841, 2002.
[222] Y. Jin, E. Tebekaemi, M. Berges, and L. Soibelman, "Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities," in *ICASSP*, 2011, pp. 4340-4343.

[223] H. Azzini, R. Torquato, and L. C. da Silva, "Event detection methods for nonintrusive load monitoring," in *PES General Meeting| Conference & Exposition, 2014 IEEE*, 2014, pp. 1-5: IEEE.
[224] K. Barsim, R. Streubel, and B. Yang, "Unsupervised Adaptive Event Detection for Building-Level Energy Disaggregation," *Proceedings of power and energy student summt (PESS), Stuttgart, Germany*, 2014.
[225] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. Rizzoli, "A Hybrid Signature-based Iterative Disaggregation algorithm for Non-Intrusive Load Monitoring," *Applied Energy*, vol. 185, pp. 331-344, 2017.
[226] S. Henriet, U. Simsekli, B. Fuentes, and G. Richard, "A Generative Model for Non-Intrusive Load Monitoring in Commercial Buildings," *arXiv preprint arXiv:1803.00515*, 2018.
[227] B. Kalluri, A. Kamilaris, S. Kondepudi, H. W. Kua, and K. W. Tham, "Applicability of using time series subsequences to study office plug load appliances," *Energy and Buildings*, vol. 127, pp. 399-410, 2016.
[228] H. N. Rafsanjani, C. R. Ahn, and J. Chen, "Linking building energy consumption with occupants' energy-consuming behaviors in commercial buildings: Non-intrusive occupant load monitoring (NIOLM)," *Energy and Buildings*, vol. 172, pp. 317-327, 2018.
[229] N. Sadeghianpourhamami, J. Ruyssinck, D. Deschrijver, T. Dhaene, and C. Develder, "Comprehensive feature selection for appliance classification in NILM," *Energy and Buildings*, vol. 151, pp. 98-106, 2017.
[230] A. S. Bouhouras, P. A. Gkaidatzis, E. Panagiotou, N. Poulakis, and G. C. Christoforidis, "A NILM algorithm with enhanced disaggregation scheme under harmonic current vectors," *Energy and Buildings*, vol. 183, pp. 392-407, 2019.
[231] M. Aiad and P. H. Lee, "Non-intrusive load disaggregation with adaptive estimations of devices main power effects and two-way interactions," *Energy and Buildings*, vol. 130, pp. 131-139, 2016.
[232] M. Aiad and P. H. Lee, "Unsupervised approach for load disaggregation with devices interactions," *Energy and Buildings*, vol. 116, pp. 96-103, 2016.
[233] D. García, I. Díaz, D. Pérez, A. A. Cuadrado, M. Domínguez, and A. Morán, "Interactive visualization for NILM in large buildings using non-negative matrix factorization," *Energy and Buildings*, vol. 176, pp. 95-108, 2018.
[234] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870-1891, 1992.
[235] A. Iwayemi and C. Zhou, "SARAA: Semi-supervised learning for automated residential appliance annotation," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 779-786, 2017.
[236] L. Farinaccio and R. Zmeureanu, "Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses," *Energy and Buildings*, vol. 30, no. 3, pp. 245-259, 1999.
[237] Y. Jin, E. Tebekaemi, M. Berges, and L. Soibelman, "Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 4340-4343: IEEE.
[238] H. A. Azzini, R. Torquato, and L. C. da Silva, "Event detection methods for nonintrusive load monitoring," in *PES General Meeting| Conference & Exposition, 2014 IEEE*, 2014, pp. 1-5: IEEE.
[239] K. S. Barsim, R. Streubel, and B. Yang, "Unsupervised adaptive event detection for building-level energy disaggregation," *In proceedings of Power and Energy Student Summt (PESS), Stuttgart, Germany*, 2014.
[240] B. Wild, K. S. Barsim, and B. Yang, "A new unsupervised event detector for non-intrusive load monitoring," in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*, 2015, pp. 73-77: IEEE.
[241] B. Zhao, L. Stankovic, and V. Stankovic, "On a training-less solution for non-intrusive appliance load monitoring using graph signal processing," *IEEE Access*, vol. 4, pp. 1784-1799, 2016.

[242] G. L. Grinblat, L. C. Uzal, and P. M. Granitto, "Abrupt change detection with one-class time-adaptive support vector machines," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7242-7249, 2013.
[243] C. S. Hilas, I. T. Rekanos, and P. A. Mastorocostas, "Change point detection in time series using higher-order statistics: a heuristic approach," *Mathematical Problems in Engineering*, vol. 2013, 2013.
[244] A. N. Milioudis, G. T. Andreou, V. Katsanou, K. I. Sgouras, and D. P. Labridis, "Event detection for load disaggregation in smart metering," in *Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*, 2013, pp. 1-5: IEEE.
[245] L. Ye and E. Keogh, "Time series shapelets: a new primitive for data mining," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 947-956: ACM.
[246] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in neural information processing systems*, 2002, pp. 849-856.
[247] B. Mohar, "Some applications of Laplace eigenvalues of graphs," in *Graph symmetry*: Springer, 1997, pp. 225-275.
[248] F. Jazizadeh, "Building Energy Monitoring Realization: Context-Aware Event Detection Algorithms for Non-Intrusive Electricity Disaggregation," in *Construction Research Congress 2016*, pp. 839-848.
[249] H. Alt and M. Godau, "Computing the Fréchet distance between two polygonal curves," *International Journal of Computational Geometry & Applications*, vol. 5, no. 01n02, pp. 75-91, 1995.
[250] J. Liang, S. K. Ng, G. Kendall, and J. W. Cheng, "Load signature study—Part I: Basic concept, structure, and methodology," *IEEE transactions on power Delivery*, vol. 25, no. 2, pp. 551-560, 2010.
[251] K. D. Anderson, M. E. Bergés, A. Ocneanu, D. Benitez, and J. M. Moura, "Event detection for non intrusive load monitoring," in *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*, 2012, pp. 3312-3317: IEEE.
[252] I. Rahman, M. Kuzlu, and S. Rahman, "Power disaggregation of combined HVAC loads using supervised machine learning algorithms," *Energy and Buildings*, vol. 172, pp. 57-66, 2018.
[253] C. Deb, M. Frei, J. Hofer, and A. Schlueter, "Automated load disaggregation for residences with electrical resistance heating," *Energy and Buildings*, vol. 182, pp. 61-74, 2019.
[254] M. Pipattanasomporn, M. Kuzlu, S. Rahman, and Y. Teklu, "Load profiles of selected major household appliances and their demand response opportunities," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 742-750, 2014.
[255] U. S. Department of Energy. (2016). *Miscellaneous Electric Loads: What Are They and Why Should You Care?* Available: https://www.energy.gov/eere/buildings/articles/miscellaneous-electric-loads-what-are-they-and-why-should-you-care. Last retrieved on June 22, 2020.
[256] A. Chopra and V. Kundra, "A POLICY FRAMEWORK FOR THE 21st CENTURY GRID: Enabling Our Secure Energy Future," 2011.
[257] W. Tushar, T. K. Saha, C. Yuen, D. Smith, and H. V. Poor, "Peer-to-Peer Trading in Electricity Networks: An Overview," *IEEE Transactions on Smart Grid*, 2020.
[258] *German Renewable Energy Sources Act*. Available: https://en.wikipedia.org/wiki/German_Renewable_Energy_Sources_Act. Last retrieved on June 22, 2020.
[259] S. Nowak, "Trends 2013 in photovoltaic applications: Survey report of selected IEA countries between 1992 and 2012," *Int. Energy Agency (IEA). Paris, France. Tech. Rep. IEA-PVPS T1-23*, vol. 2013, 2013.
[260] Inside Climate News. (2019). *As Rooftop Solar Grows, What Should the Future of Net Metering Look Like?* Available: https://insideclimatenews.org/news/11062019/rooftop-solar-net-metering-

rates-renewable-energy-homeowners-utility-state-law-changes-map. Last retrieved on June 22, 2020.
[261] J. Salom, J. Widén, J. Candanedo, and K. B. Lindberg, "Analysis of grid interaction indicators in net zero-energy buildings with sub-hourly collected data," *Advances in Building Energy Research*, vol. 9, no. 1, pp. 89-106, 2015.
[262] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125-3148, 2018.
[263] Z. Zhang, R. Li, and F. Li, "A Novel Peer-to-Peer Local Electricity Market for Joint Trading of Energy and Uncertainty," *IEEE Transactions on Smart Grid*, 2019.
[264] T. Liu, X. Tan, B. Sun, Y. Wu, X. Guan, and D. H. Tsang, "Energy management of cooperative microgrids with p2p energy sharing in distribution networks," in *2015 IEEE international conference on smart grid communications (SmartGridComm)*, 2015, pp. 410-415: IEEE.
[265] S. Cui, Y.-W. Wang, and J.-W. Xiao, "Peer-to-peer energy sharing among smart energy buildings by distributed transaction," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6491-6501, 2019.
[266] R. Luthander, J. Widén, D. Nilsson, and J. Palm, "Photovoltaic self-consumption in buildings: A review," *Applied energy*, vol. 142, pp. 80-94, 2015.
[267] J. Linssen, P. Stenzel, and J. Fleer, "Techno-economic analysis of photovoltaic battery systems and the influence of different consumer load profiles," *Applied Energy*, vol. 185, pp. 2019-2025, 2017.
[268] S. Quoilin, K. Kavvadias, A. Mercier, I. Pappone, and A. Zucker, "Quantifying self-consumption linked to solar home battery systems: Statistical analysis and economic assessment," *Applied Energy*, vol. 182, pp. 58-67, 2016.
[269] *Tesla Powerwall*. Available: https://www.tesla.com/powerwall. Last retrieved on June 22, 2020.
[270] E. Barbour and M. C. González, "Projecting battery adoption in the prosumer era," *Applied energy*, vol. 215, pp. 356-370, 2018.
[271] M. Hu, F. Xiao, and L. Wang, "Investigation of demand response potentials of residential air conditioners in smart grids using grey-box room thermal model," *Applied energy*, vol. 207, pp. 324-335, 2017.
[272] M. Hu and F. Xiao, "Investigation of the demand response potentials of residential air conditioners using grey-box room thermal model," *Energy Procedia*, vol. 105, pp. 2759-2765, 2017.
[273] P. Bovornkeeratiroj, S. Iyengar, S. Lee, D. Irwin, and P. Shenoy, "RepEL: A Utility-Preserving Privacy System for IoT-Based Energy Meters," in *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)*, 2020, pp. 79-91: IEEE.
[274] Z. Guan, G. Si, J. Wu, L. Zhu, Z. Zhang, and Y. Ma, "Utility-privacy tradeoff based on random data obfuscation in internet of energy," *IEEE Access*, vol. 5, pp. 3250-3262, 2017.

Licensing your original work with a **Creative Commons License**

http://www.creativecommons.org

The six Creative Commons licenses permit varying types of uses. The following chart illustrates the permissions, requirements, and restrictions of the CC licenses, from the least restrictive, to the most restrictive.

Link	Icon	Licenses	Author allows users to	Author requires users to	Author restricts users from
Link		**CC BY**	Copy, distribute, display, perform, and remix the work.	Attribute or credit the author as requested.	
Link		**CC BY-SA** (CC BY Share Alike)	Copy, distribute, display, perform, and remix the work.	Attribute or credit the author as requested. Apply the same CC license used by the author to the derivative work.	
Link		**CC BY-NC** (CC BY Non-Commercial)	Copy, distribute, display, perform, and remix the work.	Attribute or credit the author as requested.	Copying, distributing, displaying, performing, or remixing the work for commercial purposes.
Link		**CC BY-ND** (CC BY No Derivative Works)	Copy, distribute, display, and perform verbatim (unchanged) copies of the work.	Attribute or credit the author as requested.	Remixing or creating derivatives of the work.
Link		**CC BY-NC-SA** (CC BY Non-Commercial, Share Alike)	Copy, distribute, display, perform, and remix the work for non-commercial purposes.	Attribute or credit the author as requested. Apply the same CC license used by the author to the derivative work.	Copying, distributing, displaying, performing, and remixing the work for commercial purposes.
Link		**CC BY-NC-ND** (CC BY Non-Commercial, No Derivative Works)	Copy, distribute, display, and perform verbatim (unchanged) copies of the work for non-commercial purposes.	Attribute or credit the author as requested.	Remixing or creating derivatives of the work. Copying, distributing, displaying, performing, and remixing the work for commercial purposes.