# Big Data Analytics for Demand Response in Smart Grids

James Timilehin Oyedokun

Submitted in fulfilment of the requirements for the

Degree of Doctor of Philosophy

School of Engineering

College of Science and Engineering

University of Glasgow

November 2021

# Abstract

The transition to an intelligent, reliable and efficient smart grid with a high penetration of renewable energy drives the need to maximise the utilisation of customers' demand response (DR) potential. More so, the increasing popularity of smart meters deployed at customers' sites provides a vital resource where data driven strategies can be adopted in enhancing the performance of DR programs. This thesis focuses on the development of new methods for enhancing DR in smart grids using big data analtyics techniques on customers smart meter data. One of the main challenges to the effective and efficient roll out of DR programs particularly for peak load reduction is identifying customers with DR potential. This question is answered in this thesis through the proposal of a shape based clustering algorithm along with novel features to target customers. In addition to targeting customers for DR programs, estimating customer demand baseline is one of the key challenges to DR especially for incentive-based DR. Customer baseline estimation is important in that it ensures a fair knowledge of a customers DR contribution and hence enable a fair allocation of benefits between the utility and customers. A Long Short-Term Memory Recurrent Neural Network machine learning technique is proposed for baseline estimation with results showing improved accuracy compared to traditional estimation methods. Given the effect of demand rebound during a DR event day, a novel method is further proposed for baseline estimation that takes into consideration the demand rebound effect. Results show in addition to customers baseline accurately estimated, the functionality of estimating the amount of demand clipped compared to shifted demand is added.

# Contents

# List of Tables

# List of Figures

# List of Acronyms

**ADP** Adaptive Dynamic Programming

**ARM** Association Rule Mining

**CBL** Customer Baseline Load

**CER** Commission for Energy Regulation

**CFSFDP** Clustering by Fast Search and Find of Density Peaks

**CLARA** Clustering Large Application

**COP** Context-independent, Optimality and Partiality

**CPP** Critical Peak Pricing

**CVM** Core Vector Machine

**DBSCAN** Density-Based Spatial Clustering of Applications with Noise

**DENCLUE** Density Clustering

**DLC** Direct Load Control

**DR** Demand Response

**DT** Decision Tree

**DTW** Dynamic Time Warping

**EEA** European Environmental Agency

**EV** Electric Vehicle

**FCM** Fuzzy C-means

**FERC** Federal Energy Regulatory Commission

**FMM** Finite Mixture Modelling

**FNN** Feed-forward Neural Network

**FP** Frequent Pattern

**GHSOM** Growing Hierarchical Self-Organising Map

**GP** Gaussian Process

**HVAC** Heat, Ventilation, and Air Conditioning

**IBR** Inclining Block Rate

**I/C** Interruptible/Curtailable

**ICA** Independent Component Analysis

**IEA** International Energy Agency

**KS** Kolmogorov-Smirnov

**K-SVD** K-Singular Value Decomposition

**LB** Lower Bound

**LP** Linear Program

**LSTM** Long Short-Term Memory

**MAPE** Mean Absolute Percentage Error

**MDP** Markov Decision Process

**MLib** Machine Learning Library

**MOA** Massive Online Analysis

**MPE** Mean Percentage Error

**NBC** Naive Bayesian Classification

**NC** Nearest to Centroid

**NIST** National Institute of Standards and Technology

**NP** Non-Deterministic Polynomial-Time

**NPR** Non-Parametric Regression

**OCPA** Optimal Charging Point Assignment

**OCSD** Optimal Charging Station Deployment

**OCSP** Optimal Charging Station Placement

**OPTICS** Ordering Points to Identify the Clustering Structure

**PAM** Partitioning Around Medoids

**PCA** Principal Component Analysis

**PMU** Phasor Measurement Unit

**PR** Parametric Regression

**PSO** Particle Swarm Optimisation

**PV** Photovoltaic

**RL** Reinforcement Learning

**RMSE** Root Mean Square Error

**RNN** Recurrent Neural Network

**RTP** Real Time Pricing

**Sarsa** State-action-reward-state-action

**SOC** State of Charge

**SPR** Semi-Parametric Regression

**SMS** Short Message Service

**SVM** Support Vector Machine

**TCL** Thermostatically Controlled Loads

**TD** Temporal Difference

**ToU** Time of Use

**TSA** Transient Stability Assessment

**TVEM** Time-Varying Effects Model

**VDBSCAN** Varied Density-Based Spatial Clustering of Applications with Noise

**WEKA** Waikato Environment for Knowledge Analysis

**XML** Extensible Markup Language

# Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Shengrong Bu for her guidance and support all through my research for which this would not have been possible.

I am very thankful to the Petroleum Trust Development Fund for financing my research through the Presidential Special Scholarship Scheme for Innovation and Development (PRESSID).

I am very grateful to my wife Monique and weans Hamish and Adaira for their support and sacrifices all through my research. Yous kept me going.

Deepest gratitude to my parents Adetoyese and Funmilola for laying a solid academic foundation for me and my siblings and supporting us through the years, prioritising our education above all even in the most difficult of times.

# Declaration

I hereby declare that the contents of this thesis are original and my own work. References made to others work are duly cited and contained in the bibliography. This thesis has not been submitted for the award of any degree either in parts or as a whole.

James Oyedokun

March 2021

# Chapter 1

# Introduction

The increasing urgency and drive towards a carbon neutral global society has led to an increased volume of research across various disciplines targeted at reducing carbon footprint. According to the European Environmental Agency (EEA), energy supply accounted for 29.3% of the global greenhouse gas emissions as at 2014 [1]. The high proportion of emissions from energy supply makes it a particularly important sector of focus for carbon reduction. One of the major drive to achieve emissions reduction in the energy sector is the move to adopt increased utilisation of renewable energy sources in place of energy supply associated with high carbon footprint like coal. In order to keep global emissions down year on year, aside from increasing the renewable source portion of the energy mix, it is important that this increase compensate for rises in demand. According to the International Energy Agency (IEA), despite the increased renewable portion of energy supplied in 2018 compared to 2017, CO2 emissions rose by 1.7% [2]. This rise was due to the increase in demand which mainly resulted from extreme temperatures and economic growth.

Aside increasing the renewable energy mix, the transition of the electricity grid to a smart grid is very important to support the changing energy mix. The transition to smart grids is importatant to enable smart decision making especially to mitigate against the non-dispatchable nature that characterises renewable energy sources like wind and solar. Smart grids refer to the next generation of electricity networks that integrate the functionality of smart sensors, two-way communication technologies and computational intelligence to enhance the reliability, sustainability and efficiency of power generation, transmission, distribution and consumption [3]. In conceptualising smart grids network, the National Institute of Standards and Technology (NIST) highlights seven major domains which are customers, markets, service providers, operations, bulk generation, transmission and distribution. The role of customers in smart grids comes from their flexiility in demand and this is enabled through demand response (DR). The Federal Energy Regulatory Commission (FERC) defines DR as “changes in electric usage by demand-side resources from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized” [4]. DR involves reduction of energy consumption at peak periods through load curtailment strategies (peak clipping), demand shift from peak period to off-peak period (load shifting) and aiding off-peak consumption through storage devices such as rechargeable batteries and electric vehicles (EVs) (valley filling) [5]. Smart grids provide an efficient platform where customers not only consume electricity but also play an important role in achieving an optimal operation of the overall smart grids through their participation in DR programs, and customer site distributed generation and storage.

The increasing popularity of smart meters at customer sites provides a huge resource for insight discovery as customers demand patterns can be studied and analysed for the purpose of enhancing their role in demand and supply balance. The importance of electricity demand and supply balance is not limited to available resources but also includes managing supply equipment constraint, enhancing the integration of renewable sources, lowering emissions and enabling an efficient system all through the stages of power flow. The advancement in computing technologies makes the need for the development and implementation of data analytics techniques paramount as value can be extracted from the data generated and collected through the electricity network. The key focus of this thesis is the development and implementation of data analytics techniques to enhance the performance of DR in smart grids.

The large size and heterogeneous properties of data sets especially at utility scale create the need for a robust data management system and novel data analytics solutions for knowledge extraction. The framework of big data technologies for utility applications is illustrated in Fig. 1.1. Data measuring devices including smart meters and network sensors make up Layer 1. The produced data is communicated to relevant node(s) in the network by using state-of-the-art two-way communication technologies in Layer 2. A robust data management system that manages and integrates the collected data is represented in Layer 3. Knowledge extraction which involves the application of big data analytics techniques is implemented in Layer 4. Layer 5 represents the utility applications, which refer to DR in this thesis.

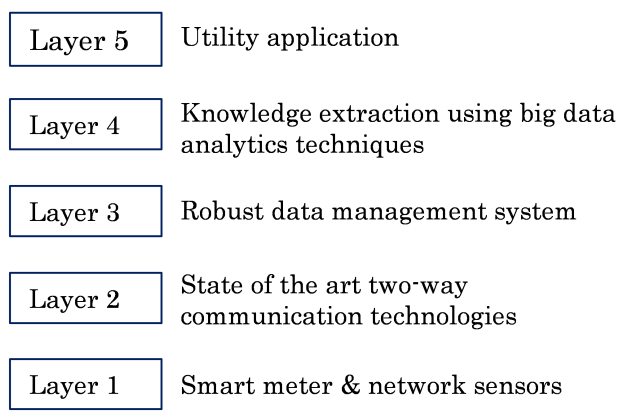


Figure 1.1: The framework of big data technologies for utility applications in smart grids.

## 1.1 Problem Statement

The increasing usage of smart meters at customer sites provides a vital opportunity for network operators to efficiently manage and target customers for DR programs. Prior to the widely use of smart meters, customers are usually recruited for DR programs through cost information on their bills, geographical area and monthly power consumption. These information cannot reflect a consumer's demand characteristics at a fine grain scale (e.g. second, minute, hourly interval), making the targeting of consumers for peak load reduction inefficient. Smart meters can provide the consumers' energy consumption data in real time giving opportunity for DR specific insight discovery. The knowledge derived through the analysis of customers smart meter data can give network operators opportunities to intelligently deploy targeted DR programs based on customers suitability and flexibility. A high participation of highly flexible consumers in DR programs is expected to enhance the grid transformation especially given their ability to contribute to mitigating the intermittent nature of renewable energy sources. In targeting customers for DR programs aimed at peak load reduction, there is a need to group customers based on their demand profile characteristics and also provide a ranking feature on each customer to meet demand curtailment target.

Aside targeting customers for peak load reduction, one of the main opportunities for enhancing DR performance in smart grids is in baseline estimation. Customer baseline refers to the demand a customer would have had if they did not participate in a DR event. Customer baseline estimation is important as it helps in estimating how much demand is reduced by a customer during a DR event. This estimation ensures the benefits of demand reduction is fairly shared between the customers and the electricity supplier. During a DR event, demand reductions usually come as both clipping and shifting. Demand clipping is when demand is directly reduced during DR period. A shift in demand is when the amount of demand reduced during the DR period is moved to other periods. Demand shift can give rise to what is known as the rebound effect and taking this into account can help improve the accuracy of baseline estimation.

This thesis focuses on answering the problem of how customers should be targeted for peak reduction; methodology for customer baseline estimation with a very important feature of taking the demand rebound into consideration.

## 1.2 Research Contributions and Publications

Based on the research work done in this thesis, the following contributions and research publications are presented as follows:

- A comprehensive survey of big data analytics and its application in literature is presented in chapter 3. Big data analytics techniques are classified into three main classes which are Supervised, Unsupervised and Reinforcement learning. The applications of the various techniques in demand response and electric vehicle load integration are discussed.

- A novel methodology is proposed for targeting customers for peak load reduction as well as local peaks. *k*-medoid with dynamic time warping distance measure is proposed as the clustering technique to cluster customers profile. The representational nature of *k*-medoid and the shape based characteristics of the dynamic time warping distance measure makes the methodology of specific application to the problem of targeting customers for peak load reduction. Results shows an improved targeting of customers compared to existing methodology. The contents of this chapter is partly published in the conference paper below:

  - J. Oyedokun, S. Bu, Y. Xiao, and Z. Han, "Smart meter data characterization and clustering for peak demand targeting in smart grids," in Proc. of IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe), Sarajevo, Bosnia and Herzegovina, Oct. 2018.

- A customer baseline estimation methodology is proposed using the Long Short-Term Memory Recurrent Neural Network technique. Historic demand data within the DR event time span is used to train the LSTM model and proposed estimation takes in data of a defined number of previous like-days for estimating customers baseline for the DR event period. This proposed methodology takes a targeted approach of estimating for the DR period rather than the traditional approach of estimating the customer demand profile for the whole day. The contents of this chapter is published in the conference paper below:

- J. Oyedokun, S. Bu, Z. Han, and X. Liu "Customer baseline load estimation for incentive-based demand response using long-short term memory recurrent neural network," in Proc. of IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe), Bucharest, Romania, Sep. 2019.

- A control based residential customer baseline estimation methodology that takes into consideration the DR rebound effect is proposed. Customers are clustered based on the similarity of their demand profile and the demand of non participating DR customers are utilised in estimating the demand of DR participating customers. The proposed methodology not only estimates the amount of demand reduction during DR but also estimates the demand clipped and demand shifted.

## 1.3 Thesis Layout

The rest of the thesis is organised as follows:

- Chapter 2 presents a literature review on DR. The review focuses on programs for DR; DR models and DR benefits.

- Chapter 3 presents a review of big data analytics and its application to DR and EV load integration to smart grids. The features as well as the various classification of big data analytics are also presented.

- Chapter 4 presents the proposed methodology for targeting customers for peak demand reduction as well as local peak DR.

- Chapter 5 presents the proposed estimation methodology for customer baseline using the Long Short-Term Memory Recurrent Neural Network technique.

- Chapter 6 presents the proposed control based estimation of customer baseline estimation methodology that takes into consideration the demand rebound effect.
- Chapter 7 presents the conclusion of the thesis as well as areas for future work.

# Chapter 2

# Demand Response in Smart Grids

Demand response refers to a change in the electricity usage of a consumer in response to price changes in electricity or incentive payment for the purpose of achieving a match to supply and meet network constraint requirements. Demand response generally involve load reduction and shifting from periods of DR events to non-DR events.

In actualising the smart grids objectives, the role of DR cannot be underestimated. The high level of renewable energy in the power supply mix makes flexible assets a key element in achieving supply/demand balance as well as ensure a stable and reliable electricity network. The flexibility of DR will play a significant role in meeting the flexibility needs accompanied by increasing renewable energy.

One of the key focus of this research is identifying and estimating DR resources for meeting DR needs in smart grids.

Demand response (DR) in smart grids is discussed under three main topic headings below which are DR programs; DR models and DR benefits.

## 2.1 Demand Response Programs

DR programs refer to motivating measures used by power utility to incentivize users to reschedule their demand from time to time [6]. The aim of the programs is to shape the customers' demand characteristics to improve reliability of the grids, increase renewable energy integration, and enhance operational efficiency of the grids. DR programs can be categorized into price-based DR programs and incentive-based DR programs based on the motivation methods offered to consumers for their efforts to reduce or shift their energy demand [7].

### 2.1.1 Price-Based DR Programs

These DR programs aim to change electricity demand characteristics of the customers from normal usage pattern using time-varying electricity prices. Customers and utility companies can benefit from DR programs in terms of bills savings and reduced cost of operation, respectively. The extent to which customers respond to varying electricity prices mainly depends on their energy consumption habits. Recent literature on price-based implementation of DR in smart grids involves minimizing utility cost and maximizing customer satisfaction. Recent work in [8] presented the design of a price-based DR program to minimize the utility cost by using the state-of-the-art prediction method of customer demand based on historic consumption data. A price-based DR strategy for multi-zone office buildings was proposed in [9] to co-optimize the energy cost of heating, ventilation, and air conditioning (HVACs) and the thermal preference of each occupant. An optimization framework was proposed in [10] for energy cost minimization and customer convenience maximization based on price-based DR programs.

Price-based DR programs can be classified into time-of-use pricing (ToU) programs, critical peak pricing (CPP) programs, real-time pricing (RTP) programs, and inclining block rate (IBR) programs [6].

#### ToU

This refers to two or more fixed energy price rates, which vary across periods of the day and different seasons of a year [6]. Prices are usually released ahead of time and keep unchanged for a long period of time. The time blocks can be represented differently at the different places. For example, the time blocks are represented as on-peak, mid-peak and off-peak periods in Ontario, Canada [11], whereas ToU features in on-peak (daytime) and off-peak (night time) as described in the Economy 7 tariff plan in the UK [12]. Prices are higher during on-peak time intervals and lower during off-peak time intervals. Minimizing electricity bills can serve as a motivating factor for consumers to shift deferrable loads to off-peak or mid-peak periods. The price rate for each time interval can be changed from season to season depending on the seasonal cost of electricity generation and supply. A radical type of ToU pricing is zero pricing or negative pricing, which means that it is free for customers to consume electricity or customers will be paid to consume electricity during this time period. An increasing number of renewable sources feeding into the grid have made negative prices increasingly common. In [13] the case for zero pricing is presented in comparison to the traditional ToU pricing program with its effectiveness in motivating users to participate in DR programs argued.

#### CPP

This kind of pricing schemes are very similar to ToU except that a much higher rate can be applied for certain days when reliability of the grids is likely to be jeopardized [14]. In general, the much higher peak price takes place only for a limited number of days within a year, when the demand is extremely high. An example of CPP is the Tempo tariff in France [15], where six rates of electricity pricing are used based upon the particular days and hours of use. Days over the year are divided into three groups, which are colour coded into blue, white and red, respectively [15]. Each day also has normal and off-peak periods. Within a year, there are 300 blue days with a low electricity price, 43 white days with a medium electricity price, and 22 red days (critical days) with a high electricity price [15]. The exact colour for a day is determined by the electricity provider based on the forecast of electricity demand for that day and load situation of the electricity network, and then informed to the customers each previous night. A pilot study in [16] shows a 45% and 15% reduction in daily consumption on a red day and white day, respectively, compared to that on a blue day. This result gives an indication of how consumers are likely to adjust their consumption pattern giving high peak prices during critical days, hence a high chance of guaranteeing system reliability.

### RTP

RTP is a pricing scheme where each time period (usually every 15 minutes or hourly) has a different electricity price determined by the supplier usually depending on the wholesale market prices [5]. Customers receive price information ahead of time (usually 1 hour ahead or day-ahead hourly pricing), and then adjust their demands with respect to the received price signals and their energy consumption habits. The need for continuous price information from the utility and consumption level feedback from the consumers makes a two-way communication utmost important in RTP implementation. Smart meters provide two-way communication capability, and hence make their deployment key to the wide spread implementation of RTP.

### IBR

This kind of programs are designed with a multi-level rate structure corresponding to how much electricity a user consumes [5]. Customers are rewarded with a lower electricity rate for consuming electricity below a defined threshold. IBR promotes conservation as consumers' objectives of cost minimisation and comfort are met. IBR motivates consumers to distribute their demand across different times of a day to avoid higher rates thereby, reducing the peak-to-average ratio in the grid.

### 2.1.2 Incentive-Based DR Programs

These refer to programs that offer customers incentives to reduce their electricity demand during a period of supply shortfall and system stress. These incentives can be in addition to or separate from electricity prices, and the incentive amount can be fixed or time-varying [7]. Incentive-based DR programs can be divided into classical programs and market-based programs [7]. Classical programs involve schemes in which customers are awarded bill credits or discounted rates for participating the programs [7]. Classical programs include direct load control (DLC) programs and interruptible/curtailable (I/C) load programs. In market-based programs, participating customers are rewarded financially depending on how much load they curtail during the critical network conditions [7]. Market-based programs include demand bidding programs, and emergency demand reduction programs [7].

#### DLC Programs

These refer to the programs where customers' electrical appliances (e.g. water heater, HVACs) are remotely controlled, e.g., turned off whenever needed by the utility in return for incentive payments. This kind of schemes require direct access to customer equipment by the utility company, which makes DLC highly dispatchable in meeting network and supply constraints compared to other DR schemes.

#### I/C Load Programs

Utilities request a pre-determined level of curtailment from enrolled customers. Incentives takes the form of rate discount or upfront payments. Customers might receive penalties according to the terms of conditions if they do not meet their load reduction commitment.

**Demand Bidding Programs**

In this kind of programs, customers bid to curtail electricity use during peak demand and system stress [7]. The bid would normally include the demand reduction capacity of the customers and their requested prices. Participating customers are usually large consumers who offer bids in the electricity wholesale market [17]. Customers with low electricity consumption can also participate through a third party that aggregates their curtailment and represents them to bid [5].

**Emergency Demand Reduction Programs**

Incentive payments are provided to users in response to their curtailment during system emergency situation and when the grid is out of reserve. Demand reduction by larger electricity users can provide auxiliary services to the utility by acting as virtual spinning reserves [5].

## 2.2 Demand Response Models

In DR applications, utility functions, cost functions [5], or a combination of both have been widely used in literature to model the behaviours of the users and power utility. The need to reduce carbon emissions of power systems in view of meeting overall carbon emission targets presents the need for a carbon emission function in DR modelling. Therefore, in the following, carbon emission functions for DR in smart grids is introduced and formulated.

### 2.2.1 Utility Functions

A utility function describes the level of satisfaction users obtain as a function of their energy usage, which is non-decreasing and concave [5]. Different utility functions are used to represent the behaviours of different users. Quadratic utility functions are commonly used to model consumers' utility as it is non-decreasing and its marginal benefit to the consumer is linearly decreasing [18] [19] [20]. Other utility functions can also be used for DR if they meet the following two properties [21].

- Utility functions are non-decreasing: This means users prefer to consume more electricity until the desired level of electricity consumption is met [21].
- The marginal benefit to the customer is a non-increasing function: This implies that the utility function is concave and the comfort obtained by the user will saturate as soon as the desired energy consumption level is reached. The participation of various customers' load components depends on their load flexibility and how they impact customers' utility level during DR events. Residential loads can be classified into thermostatically controlled loads (TCLs), urgent non-TCLs, non-urgent non-TCLs and battery-based loads, and work in [22] proposed models to describe the level of flexibility of each load type.

### 2.2.2 Cost Functions

This kind of functions represent the cost of electricity generation and delivery by the power utility, which is increasing and strictly convex [21]. Price-based DR can be modelled with cost functions describing the characteristics of energy cost in response to level of demand. Two properties of cost functions are outlined for modelling DR as follows [5]:

- Firstly, the energy cost continually increases with the increase of demand.
- The marginal cost of the power utility is increasing.

Examples of cost functions that satisfy these two properties are piece-wise linear functions and quadratic functions.

### 2.2.3 Carbon Emission Functions

Carbon emission functions are introduced here for DR to maximize the use of renewable energy. Depending on the level of renewable energy penetration in the grid per unit time, carbon footprint per unit energy consumption varies. On-site generation (e.g photovoltaic (PV), wind) can reduce carbon footprint. DR and energy storage can also enhance the deferment of consumption from time periods of higher carbon emissions with low renewable energy penetration to those of lower carbon emissions with high renewable energy penetration. The two properties of the carbon emission functions that can be used for DR are defined as follows:

- The amount of carbon emissions is non-decreasing as demand increases.
- The marginal carbon emissions of the generating units can be either concave, convex or zero depending on the emissions characteristics of the power generator.

## 2.3 Benefits of Demand Response Programs

The benefits of DR is discussed in a view to emphasise the need for solutions aimed at maximising its potential in smart grids. 4 main benefits are discussed in the following subsection

### 2.3.1 Bill Savings

Participating customers receive savings on their electricity consumption either through reduced electricity prices in the case of price-based DR or incentives received from the supplier in the case of incentive-based DR. The amount of savings depend on how much demand is shifted to periods of either low price or non DR event period. According to work done in [23], bill savings can amount to about 38% of electricity cost for participating customers.

### 2.3.2 Reduced Cost of Electricity

Aside participating customers, other customers can benefit from reduced cost of electricity. During peak periods, peaking generators are usually employed to meet demand spikes. These generators are characterised with quick start up and shut down times and are usually very expensive. Demand reduction from participating customers can translates to reduced cost of meeting aggregate demand due to reduced need for peaking generators. How much cost reduction can be achieved will depend on the extent and proportion of customers participating in DR.

### 2.3.3 Enhancement in System Security

DR can provide flexibility to the system operator especially in mitigating the uncertainty that characterises renewable energy sources. During periods of high demand and limited supply, DR can help avert brownouts and potential blackouts through customers clipping and shifting their demand.

### 2.3.4 Investment Deferral

Increasing demand can bring about the need to invest in larger distribution and transmission capacity. Given that peak periods take a short period of time compared to the aggregate demand profile of customers, the decision to increase capacity to meet growing demand is usually an expensive and cost ineffective option. DR can help achieve reduced aggregate peak even with increasing total demand thereby offering a non expensive solution to increasing demand and deferring investment.

### 2.3.5 Emission Reduction

The reduced need for peaking generators at peak period given a wide uptake of DR programs by customers can have a huge effect on reduced carbon emissions. The flexibility of customers can also enable increased integration of renewable energy to the grid which are characteristically non dispatchable. Curtailment activity of renewable energy can be averted by customers shifting demand to periods when resources are plentiful.

# Chapter 3

# Big Data Analytics: A Review

## 3.1 Features of Big Data in Demand Response

The digitization of electricity grids brings about the generation of a huge amount of real-time data across the nodes in the network. Smart meters, phasor measurement units (PMUs) and other network sensors measure and record the continuous data describing the power flow through the nodes of the overall network. Other data resources, e.g., weather data and renewable sources data, are needed to deliver solutions to the grids. The variety of these data provides a wide dimension of information for which decisions can be made for the smooth and efficient running of the smart grids, and also leads to big data challenges.

Big data is usually characterised by "3V", which are volume, velocity and variety [24]. Some characterization includes value and/or veracity giving a 4V or 5V description [25]. Value describes the benefits derived from analysing large datasets, which can be fundamentally characterized by high volume, velocity and wide variety. Veracity refers to reliability of the data. To identify the veracity of the large datasets, some preliminary processing and analysis need to be done on the data, and the results will be compared with those of baseline methods or benchmark trends. In the following, 3V will be discussed, since we consider that 3V fundamentally characterize big data in smart grids, and value and veracity are more of an aftermath of processing big data.

### 3.1.1 Data Volume in Demand Response

Volume refers to the quantity of data generated and stored. In smart grids, smart meters and network sensors generate a huge amount of data in real time. According to the MaRS Market Insights report, a smart meter reporting data at 15-minute intervals generates about 400MB of data per year [26]. As a result, about 9.44 million smart meters installed till date in the UK [27] will generate 3.776 petabytes of data per year. In DR applications, combining with data from other consumption influencing factors such as weather and social events will further increase the volume and complexity of the data. In the case of EV users, the users' charging data can be used to learn the driving characteristics of each individual user, and to estimate their potential participation in DR.

### 3.1.2 Data Velocity in Demand Response

Velocity refers to the speed at which data is generated, transmitted and processed. Smart meters and network sensors generate data in resolutions of seconds, minutes or hours depending on the requirements of the applications. An application with quicker response requires a shorter time period between the moment the data is generated, communicated and processed for knowledge extraction. At the utility scale, a robust data analytics framework needs to be designed to meet up with both kinds of applications with short and long response time.

### 3.1.3 Data Variety in Demand Response

Variety refers to various types of data. Data utilised for DR and smart grid applications can be classified as structured, semi-structured or unstructured data [28]. Structured data (e.g., energy consumption data) have a high level of organization, and can be easily queried from a relational database using structured query language [28] [29]. Semi-structured data is a form of structured data that is not organized like the structure associated with relational databases, but does contain elements that can separate the data into various hierarchies [30]. For example, data exchanged between energy management platform and third party aggregators using extensible markup language (XML) or Web services are semi-structured data [31]. Unstructured data are not organized in a pre-defined format. Examples are notification of energy use through emails or short message service (SMS). A big data analytics framework needs to take cognisance of these features of data, process, and integrate them for grid applications.

## 3.2 Techniques for Big Data Analytics

Processing and analyzing these big data pose challenges for traditional data analytics platforms, and therefore, big data analytics are needed in DR applications for smart grids. Big data analytics techniques refer to means and procedures utilised in extracting useful knowledge and insights from big datasets, which can enhance the reliability and efficiency of the smart grids.

Based on the depth of analysis, big data analytics can be divided into three levels which are descriptive, predictive and prescriptive analytics [32] [33]. Descriptive analytics exploits historical data in order to present knowledge on past events. Predictive analytics focuses on forecasting and deriving possible future outcomes using historical data in combination with rules and sometimes external data. Prescriptive analytics which is the highest level, extends the scope of big data analytics to decision making. The question of what, when and why an event will occur based on historical and real time data is answered under the scope of prescriptive analytics. In extracting knowledge and gathering intelligence from big data, various data mining and machine learning techniques can be used for descriptive, predictive or prescriptive analytics.

Machine learning and data mining techniques used for big data analytics can be categorized as supervised-learning techniques, unsupervised-learning techniques and reinforcement learning (RL) techniques, illustrated in Fig. 3.1.

### 3.2.1 Supervised-Learning Techniques

Supervised-learning techniques make use of labelled training datasets to infer functions or make predictions for unseen points [34]. The training datasets include input vectors and desired outputs examples, where the algorithms learn a general rule that maps inputs to outputs. The goal is to approximate the mapping function so well that the output of a new input data can be predicted. Supervised-learning techniques can be divided into two categories based on the characteristics of the outputs: classification methods if the outputs are discrete, and regression methods if the outputs are continuous [35].

#### Classification Methods

Classification methods assign data inputs into classes based on the defined labels [36]. This kind of methods identify which group a new observation belongs to based on the training set of data with known groups. Classifiers are extracted from the training dataset to map new data to known groups. Classification techniques include Bayesian classification, Naïve Bayesian classification (NBC), support vector machine (SVM), decision trees (DTs) [37]. In the following, each of these techniques is discussed as well as their applications in big data sets.

Bayesian classifiers are statistical classifiers, and the class membership is determined based on the probability that a given tuple belongs to a particular class [37]. Bayesian classification technique is based on the Bayes theorem which stipulates that for a data tuple $X$ belonging to a specified class, given hypothesis $H$, $P(H|X) = P(X|H) * P(H)/P(X)$. $P(H|X)$ is the probability that for the observed data tuple $X$, the hypothesis holds, i.e., posterior probability of $H$ given $X$. $P(X|H)$ is the posterior probability of $X$ given $H$.

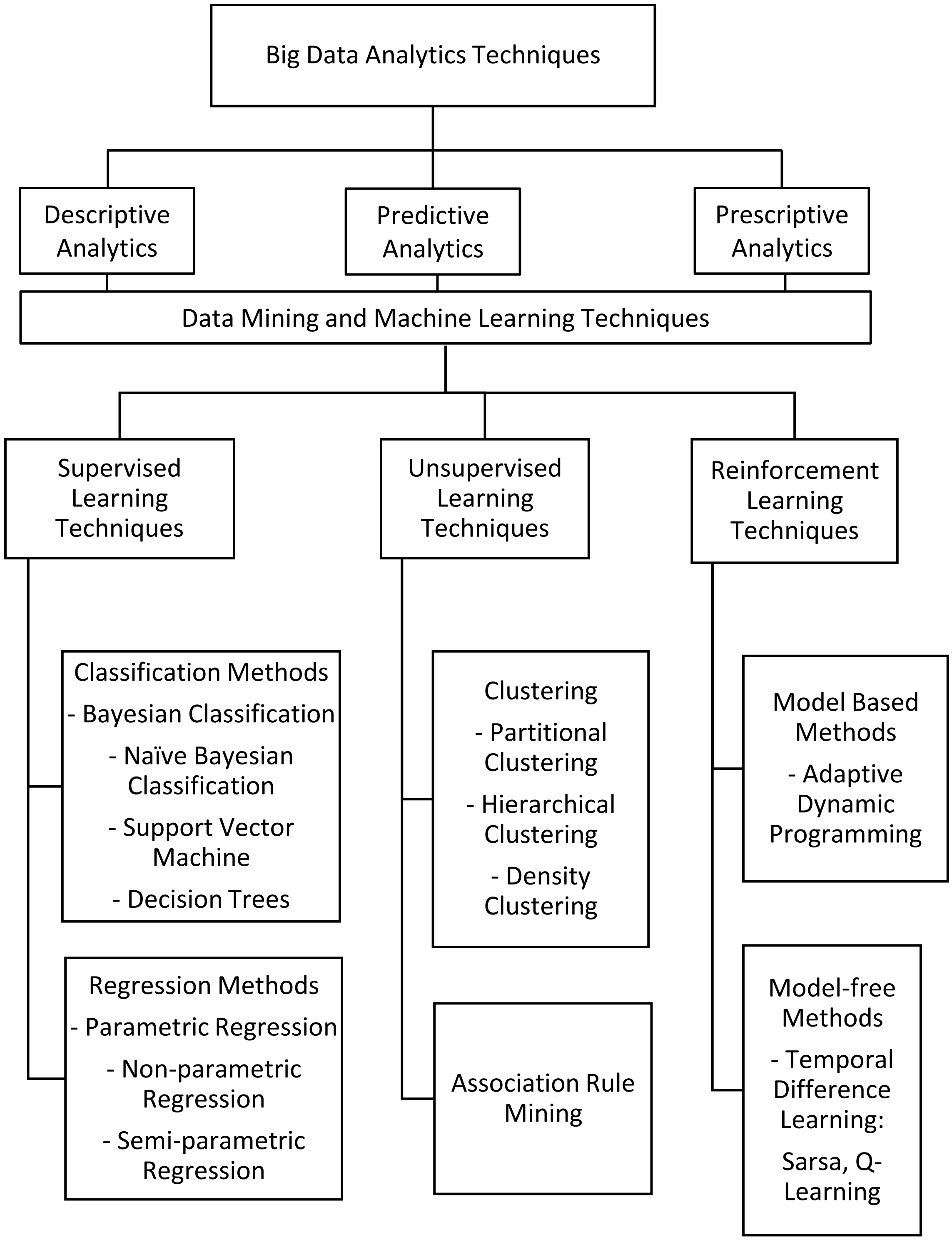


Figure 3.1: Techniques for big data analytics.

$P(X)$ is the prior probability of $X$, and $P(H)$ is the prior probability of $H$ [37]. For big data applications, a parallel and incremental classification approach was proposed in [38] to learn Bayesian networks using MapReduce[1]. Bayesian networks were also used within the MapReduce framework to discover user similarities from large social media datasets [39].

NBC is the extension version of Bayesian classification. The effect of an attribute on a given class is assumed to be independent of the other attributes, which is referred as class-conditional independence and aims to simplify the computation involved in data tuple classification [37]. The big data application of NBC was implemented in [40] using the Hadoop framework to cluster movie review data. In order to reduce data classification time and take account of the contribution weight of different data characteristics, a classification framework was proposed employing principal component analysis for dimension reduction of multidimensional datasets [41]. To enhance diagnosis accuracy and reduce diagnosis time, NBC was used to classify patients illness based on historical patients symptoms data [42].

SVM is a machine learning technique that constructs an optimal hyperplane for classification in a feature space or higher dimensional kernel space using training datasets [43] [44]. Optimal hyperplane maximises the distance margin between classes, thereby minimising the generalization error of the classifier [43]. SVM can be used to classify both linear and non-linear data. Non-linear data classification can be achieved by mapping the data tuples into a higher dimension until a linear optimal separating hyperplane is found [37]. For big datasets classification, work in [45] and [46] implemented a divide and conquer approach by splitting the data into manageable sized partitions before obtaining local support vectors from each partition. However, this splitting procedure can result in local support vectors not aligning with global support vectors as the partitioned datasets might not essentially represent the distribution of the whole dataset. To address the shortfall of this approach, a distribution preserving kernel SVM that retains the mean and variance of the whole dataset in each partition was proposed in [47]. A density-dependent vector quantization method was proposed in [48] for least square SVM applications on

[1] A programming model for processing large datasets.

big datasets. The quantization method maps outputs to input data density using a single shrinkage threshold thereby considerably reducing the sample size of the whole data. In power systems application, an online transient stability assessment (TSA) framework was proposed in [49] using core vector machine (CVM) evolved from SVM [50] classification technique on large PMUs datasets.

DTs technique utilises flowchart-like tree structures to classify input datasets into respective groups. Each internal node represents test on an attribute while each branch and terminal node represents test outcome and class label, respectively [37] [51]. To classify a data tuple, the attribute of the given data tuple is tested through the flowchart of the decision tree. The path trace from the root node to the leaf node (i.e., terminal node) shows the technique procedure of class assignment. DT classifiers are generally accurate, simple and easy to assimilate. DTs can also be classed as a regression method as it can deal with outputs with continuous features. A regression-based DT was proposed in [52] for fault distance estimation in transmission lines. For big data applications, fuzzy DTs has been shown to effectively deal with problems associated with uncertain, incomplete data and multiple solutions [53]. A distributed fuzzy decision tree learning scheme using MapReduce programming was developed in [54] for generating both binary and multi-way fuzzy decision trees for big data applications. A distributed fuzzy discretizer was used to generate strong fuzzy partitions on attributes based on fuzzy information entropy [54]. The accuracy and speed of large traffic data classification were tested in [55] using decision trees on the Hadoop infrastructure with various classifier tools including WEKA classifier, MOA classifier and Spark MLib. A distributed environment to implement KS-Tree algorithm was proposed for large data set classification [56].

### Regression Methods

Regression methods are used to estimate the relationships between the dependent output variable and one or more independent input variables, when the output variable is continuous [57]. Mathematical relationships between input and output examples are usually derived with functions, which are then used to predict the outcome of future input variables. For DR applications, various input variables that influence the characteristics of customers' demand can be modelled to predict energy usage of the customers. Input variables include weather, time of the day, type of day (workday or holiday, special events), and so on. The magnitude of influence each input has on the outputs is reflected in the mathematical function as a gain or weight factor. Regression techniques can be classified as parametric regression (PR), non-parametric regression (NPR), and semi-parametric regression (SPR).

PR uses mathematical functions with parameter vector representation to express the relationship between the dependent and the independent variables. Examples of PR are linear, polynomial, logistic, mixture models, $k$-means, hidden Makov models and factor analysis regression [58]. PR can be used when the test data is normally distributed, homogeneous and autocorrelated [58]. A mode-based parametric regression method was proposed in [59] to analyse and derive relationships between variables in a large dataset. The concept of parametric Gaussian processes and its suitability on large datasets are introduced and discussed in place of stochastic variational inference and the need of scalable algorithms for posterior distribution approximation [60].

NPR can be used when the conditions for parametric regression are not met. For this kind of models, the assumption is that data distribution cannot be defined by a finite set of parameters. Examples of NPR are Gaussian processes (GPs), kernel, Dirichlet process, infinite hidden Markov models and infinite latent factor models [58]. To scale non-parametric and semi-parametric regression applications to big datasets, a parallel distributed framework was proposed using Spark data structures on MapReduce [61]. Sparse GP and latent variable model was implemented on large datasets using the distributed MapReduce architecture [62]. An online anomaly detection algorithm was proposed based on NPR to detect abnormal access requests in a cloud environment using large communication datasets [63].

SPR uses a combination of parametric and nonparametric regression models [64], and therefore combines the advantages of both kinds of models. SPR can serve as a great technique in situations where a parametric method is desired but error distribution is unknown. Some examples of semi-parametric regression are generalized additive models, generalized partial linear models, index models and varying coefficient models [65]. A time-varying effects model (TVEM) was proposed to improve revenue generation by optimally allocating market resource big data [66]. A SPR model was proposed in [67] to explore and analyse large and complex genomic datasets to identify transcription factors for bridging functional genomics with disease risk loci. A framework was proposed for the real time application of SPR analysis on distributed large scale data [68]. An SPR approach based on generalized additive models was implemented for short-term and middle-term electricity load forecasting, and the relationship between demand, temperature and other variables was derived using weather temperature and load data from 2,200 substations [69].

### 3.2.2 Unsupervised Learning Techniques

This kind of techniques use algorithms to discover hidden patterns and structures from their unlabelled input datasets [70]. Unsupervised learning techniques are mainly categorized as clustering and association rule mining (ARM).

Clustering, the most commonly used unsupervised-learning technique, is an exploratory procedure of finding interesting subgroups in given data. The main classes of clustering techniques are partitional, hierarchical and density clustering, which are discussed below [71].

### Partitional Clustering

This kind of clustering techniques construct $k$ partitions from a set of $n$ objects [37], where $k \leq n$. Each partition represents a cluster, and each object belongs to only one cluster. The objects in a cluster are close to each other and are as far as possible to those in other clusters. Examples of partitional clustering are $k$-means and $k$-medoids clustering [37].

$k$-means clustering is a centroid-based technique, where initially $k$ points (centroids) are randomly located in space. The following two steps are performed iteratively: (1) each object is assigned to a cluster with the nearest centroid, and (2) within each cluster, the centroid is moved to the mean of the objects assigned to it [37]. The algorithm continues until no instance changes cluster membership. For nominal data, a variant of $k$-means method called $k$-modes can be used [37]. Here, the means of clusters are replaced with modes while new dissimilarity measures and frequency-based methods deal with nominal objects and cluster modes updates, respectively. For big data applications, a distributed framework based on Spark was proposed to implement $k$-means clustering [72]. The Hadoop and MapReduce frameworks were also used to implement $k$-means clustering based on Particle Swarm Optimisation (PSO) for large datasets [73]. PSO has the advantage of selecting optimal centroids from the datasets thereby, improving the clustering speed and accuracy. A MapReduce based $k$-means clustering was implemented for large datasets in a cloud computing environment [74].

$k$-medoids is a representative object-based technique, where actual objects are used to represent each cluster instead of their mean values [37]. Clusters are partitioned by minimising the sum of dissimilarities between each object in a cluster and its representative object, and the dissimilarity is measured by the absolute-error criterion [37]. For small dataset applications, $k$-medoids can be realized by the partitioning around medoids (PAM) algorithm, which solves the clustering problem in an iterative, greedy way like the $k$-means in which representative objects are replaced until the clustering quality stops improving by any replacement. For large dataset applications, a sampling-based method called clustering large applications (CLARAs) is used [37]. CLARA applies PAM algorithm on a random sample of the large dataset to determine the best medoids from the sample. How effective this method works depends on how closely the sample data represents the large dataset. The MapReduce framework has also been used to implement $k$-medoid clustering for big data application [75] [76].

#### Hierarchical Clustering

In this kind of clustering techniques, objects of a dataset are grouped into a hierarchy or tree of clusters based on the distance measure. Hierarchical clustering includes agglomerative and divisive clustering [37]. Agglomerative hierarchical clustering uses a bottom-up approach, where objects of a dataset is taken as clusters and iteratively merged to form larger clusters. Divisive clustering or called top-down clustering, takes the whole dataset as a cluster and then iteratively splits it into smaller clusters till a terminating condition is met. There are two main classes of distance measure in clustering which are algorithmic and probabilistic distance methods. Algorithmic methods use deterministic distances between objects for clustering while probabilistic methods use probabilistic models to measure the distance between objects [37].

Algorithmic methods utilise deterministic distance measures for both agglomerative and divisive hierarchical clustering. There are four deterministic distance measures approach generally used to separate clusters: minimum, maximum, mean and average distance [37]. Algorithms that use minimum distance to measure distance between clusters are referred to as nearest-neighbour clustering algorithms [37]. Farthest-neighbour clustering algorithms [37] use the maximum distance to measure distance between clusters. Mean distance represents the difference between the cluster mean of two clusters. Average distance represents the average of the difference between objects in one cluster and objects in another cluster. Multiple iterations are run to derive the hierarchical clustering until the distance measure objective is achieved for each cluster thereby requiring a quadratic time and space complexity [77]. A new linkage method called 'nearest to centroid (NC-link)' was proposed to measure cluster distance for large dataset applications [77]. The NC-link's distance measure is the distance between two data points, where each of the data point is the closest to the centroid of each cluster. The iterative distance update is not required for NC-link after merging data points, thus reducing space usage and response time [77]. In [78], a distributed hierarchical clustering approach was proposed for applications in large datasets. An hierarchical $k$-means algorithm was proposed in [79] for large dataset clustering. The clustering algorithm was used to cluster a large advanced metering infrastructure dataset consisting of 18,622 daily patterns [79].

Probabilistic methods measures the distance between clusters using probabilistic models [80]. Probabilistic methods have the advantage of been applicable in a situation where some attribute values are missing [81]. Algorithmic methods in this situation is difficult to implement as distance measure cannot be computed. Also, the challenge of deciding what distance measure to use does not arise when using probabilistic methods. Probabilistic clustering is carried out by accurately estimating a generative model from the datasets to be clustered [37]. A generative model describes probabilistically how a dataset may be formed [82]. A probabilistic hierarchical clustering model using marginal likelihood was proposed to assign datasets to respective clusters [81]. Dirichlet process mixture model was used as the generative model in [81] with Bayesian hypothesis testing to determine which cluster merger is beneficial. A model was proposed in [83] to perform joint clustering and feature selection using a statistical generative model based on the hierarchical Pitman-Yor process and the generalized Dirichlet distributions. An optimal probabilistic estimation approach was developed in [84] for hierarchical clustering using the survival of the fittest principle.

**Density-based Clustering**

This technique models clusters in data space as dense regions partitioned by sparse regions [37]. An advantage of density clustering techniques is its ability to find clusters of arbitrary shape whereas, partitional and hierarchical clustering methods are designed to discover spherical-shaped clusters [37]. Three main density-based clustering techniques and their applications in large datasets are discussed as follows.

Density-based spatial clustering of applications with noise (DBSCAN): DBSCAN performs clustering on datasets by locating core objects with dense neighbourhoods [85]. A core object is identified when it fulfils the following two conditions: every object is within a defined radius distance $\varepsilon$ to another object and the object neighbourhood contains a minimum number of objects $MinPts$ [85]. These two parameters give DBSCAN the ability to cluster datasets in arbitrary shapes. The DBSCAN algorithm has been widely used in literature, and one of its recent applications is to perform real-time image superpixel segmentation [86]. However, when using DBSCAN, setting the appropriate values for these two parameters can be challenging, which can also affect the overall clustering quality [87]. A varied DBSCAN (VDBSCAN) was proposed to find clusters in datasets of varying densities [88]. The proposed approach was based on ranking each object in the dataset by the distance to their kth nearest neighbour ($k$-dist). The $k$-dist of each object is sorted in ascending order and plotted to display the datasets density levels. DBSCAN is then applied to each density group of datasets to derive clusters. VDBSCAN addresses the challenge of varying densities that can be associated with some datasets. An algorithm improving the performance of VDBSCAN was proposed in [89] with the use of parallelism techniques in the graphics processing unit.

Ordering Points to Identify the Clustering Structure (OPTICS): OPTICS addresses one of DBSCAN's major weaknesses: the problem of detecting meaningful clusters in data of varying density. OPTICS outputs a cluster ordering such that spatially close points become neighbours in the ordering [90]. The clustering order is derived using information on the core distance and reachability distance of each object in the datasets. The core distance of Object $p$ is the smallest $\varepsilon$ such that there contains at least $MinPts$ objects in its $\varepsilon$-neighbourhood. The reachability distance of Object $p$ from $q$ refers to the minimum radius value that makes $p$ density reachable from $q$ that is, $max\{core\_distance(q), dist(p,q)\}$ [90]. In a situation where $p$ is directly reachable to more than one core objects, $p$ is assigned to the cluster whose core object has the least reachability distance from $p$ [37]. The advantage of OPTICS over DBSCAN is its ability to cluster datasets of varying densities. In order to cluster large trajectory datasets, a scalable density-based trajectory clustering algorithm was proposed based on OPTICS [91]. A scalable parallel OPTICS data clustering was also proposed in [92] using graph algorithmic techniques.

DENsity CLUstEring (DENCLUE): DENCLUE groups objects into clusters based on a set of density distribution functions [37]. The idea behind this technique is that each data point of a dataset can be modelled using an influence function, which describes its impact within its neighbourhood [93]. The density function of a data space is the sum of the influence functions of all its constituent data point. Clusters are mathematically determined by identifying local maxima (also known as density attractors) in the overall density function using a hill climbing procedure [93]. Advantages of using DENCLUE include: its ability to cluster datasets of high dimension, handle datasets with large noise, and significantly faster compared to DBSCAN [93]. In order to improve the execution time of DENCLUE, simulated annealing and genetic algorithm was proposed in [94] to identify the local maxima of density functions in place of the traditional hill climbing procedure.

**Association Rule Mining**

ARM is a rule-based unsupervised learning method used for discovering interesting relationship and rules between variables in large databases [37]. ARM can be used to find recurring patterns and correlation between items in transactional databases [37]. Association rule is mined in two steps which are frequent item set mining followed by rule generation [37]. The ARM technique has been applied to various big data applications. For DR applications, ARM can be used to study users' consumption pattern and identify customers who are likely to respond and contribute significantly to demand reduction during DR events. ARM was used to improve prediction accuracy of power transformer states by combining the Apriori algorithm and a probabilistic graphical model [95]. To address the challenge of high volume in big data, ARM was implemented on the MapReduce framework in [96]. A Spark-based ARM using frequent pattern (FP) growth algorithm was proposed in [97] to mine network performance data with the aim of discovering key quality indicator anomalies regarding user experience.

### 3.2.3 Reinforcement Learning

RL is an area of learning concerned with how agents should take actions in an environment so as to maximize some notion of cumulative reward [98]. The environment of the agent is generally formulated as a Markov decision process (MDP) [99] [98] [100]. An MDP consists of a set of environment states $\mathscr{S}$, agent actions $\mathscr{A}$, reward function $\mathscr{R} : \mathscr{S} * \mathscr{A} \rightarrow \mathscr{R}$ and the state transition probability function $\mathscr{T} : \mathscr{S} * \mathscr{A} \rightarrow \Pi(\mathscr{S})$ [98]. State $s_t \in \mathscr{S}$ is the environmental state at time $t$. The transition function associated with an action $a \in \mathscr{A}$ by the agent is $P^a_{s \rightarrow s'} = P(s'|s,a)$, where $s'$ denotes the new state and $R^a_{s \rightarrow s'}$ is the reward function. The agent aims to maximise reward by deriving an optimal policy $\pi(s,a) = P(a_t = a|s_t = s)$ [99] [101]. RL is a technique for the agent to learn the MDP, and solve it for the optimal policy at the same time [98].

RL can be categorized as model-based and model-free RL [98] [102] [103]. Model-based RL techniques are usually applied to learn the MDP model of the environment, or an approximation of it. In model-based RL, the agent chooses the action that gives the most reward based on the model. In situations where the model is difficult to represent and/or learn, model-free RL techniques become very useful. Model-free RL relies solely on trial-and-error experience to derive the optimal policy for maximum reward.

### Model-based

Model-based RL methods use experience to build an internal model of the environment's transitions and corresponding rewards [103]. The constructed model is used by the agent to take actions to find the optimal policy. One common model-based approach is adaptive dynamic programming (ADP), which has been applied widely to energy storage problems in smart grids. An iterative ADP algorithm was proposed for optimal battery control in smart home energy systems [104]. A mixed iterative ADP algorithm combining policy and value iterations was proposed in [105] to solve the optimal control and management of energy storage in smart micro-grid systems. In addition to solving energy storage problems in smart grids, ADP has been applied to DR problems. An ADP learning algorithm was proposed for the demand side management of domestic electric water heaters [106]. ADP learning was also employed to solve the residential energy scheduling problem of solar energy in micro grids [107].

### Model-free

In model-free RL methods, the agents' optimal policy is derived through a trial-and-error interaction with the environment without explicitly learning the model [102]. One common model-free RL approach is temporal difference (TD) learning. In TD learning, agents learn by bootstrapping from the current approximation of the value function [108] [109]. Insight from value iteration is used to modify the projected value of a state based on the current reward and the projected value of the next state [98]. There have been application of TD learning for power systems. An on-line adaptive dynamic power management framework was proposed in [110] based on TD learning technique. In [111], an optimal power sytem wide area controller was proposed based on a hybrid RL and TD framework. TD learning can be also applied to control problems using Sarsa and Q-learning TD control [109]. Sarsa is also known as state-action-reward-state-action, which considers the transition from state-action pair to state-action pair while learning the value of state-action pairs [109]. Sarsa is an on-policy TD learning method as the same policy used to generate the current action is used for the next action [109]. Q-learning is an off-policy TD learning method which chooses actions that maximize the estimated utility function.

## 3.3 Enhancement of Demand Response Based on Big Data Analytics

To successfully actualise smart, distributed, efficient and reliable future grids, deriving valuable knowledge from smart grid data resources is very important. The role of DR in actualising the objectives of future smart grids is very important [112]. Developing frameworks and utilising efficient data analytics technique at utility scale is thereby important to extract valuable information necessary for ensuring the efficient and reliable future smart grids. Literature on the enhancement of DR applications using big data analytics is presented.

### 3.3.1 Applications using Smart Meter Data

Traditionally, DR programs are deployed to customers with the aim of actualising a desired aggregate demand. These programs do not usually guarantee meeting the DR objectives as the customer demand characteristics per unit time is unknown. Smart meters provide a huge mass of real-time consumption data of customers, which has potential of enhancing DR programs based on the demand characteristics of each individual consumer.

In recent literature, there has been an increasing interest in knowledge extraction from smart meters data using data analytics techniques. Recent data-driven applications for DR have involved DR targeting of customers and customer DR characteristics clustering. In this section, the literature on the applications of data-driven techniques for enhancing DR

Table 3.1: References showing application of big data analytics using smart meter data for DR.

| Learning Technique / DR Enhancement Category | Supervised | Unsupervised | Reinforcement |
|---|---|---|---|
| DR potential impact assessment | [113] [114] [117] | [115] [116] | |
| Customer categorisation for DR applications | [118] [119] [122] [121] | [120] [121] | |
| DR targeting for customer participation | [123] | [124] (proposed) | |
| Enhancement of renewable energy integration through DR | [125] | | |
| DR implementation in smart grids | [126] (proposed) | | [127] [113] [128] |

using smart meter data are divided into five categories as follows:

1. DR potential impact assessment.

2. Customer categorisation for DR applications.

3. DR targeting for customer participation.

4. Enhancement of renewable energy integration through DR.

5. DR implementation in smart grids.

Table 3.1 shows the references of big data analytics applications on smart meter data by category of DR enhancement.

**DR Potential Impact Assessment**

DR programs in many cases involve distribution companies offering identical incentives to participating consumers. Since consumers generally have different electricity consumption patterns driven by their life-style and electricity devices, it is important to estimate each consumers' DR potential using their smart meter data. Extracting features that adequately represents each users DR potential can enhance the effective and efficient actualisation of DR programs. A framework was proposed to evaluate the potential of real-time DR using a Gaussian mixture model to learn the probabilistic usage behaviours of customers based on their historical smart meter data [115]. A benefits-optimisation approach to both users and utilities can be used for estimating DR potential. In [113], a profit maximization model was proposed for DR using probabilistic Bayesian model on historic user consumption data, electricity dynamic price and local weather temperature. A data driven framework was proposed in [114] using customer electricity consumption data for quantifying the economic potential of residential DR. The variation in consumers reduction pattern during DR events was analysed using latent variables in statistical forecasting methods [114].

In quantifying the DR contributions of customers during DR events, it is important to accurately determine the baseline demand assuming no DR events. A $k$-means clustering method was proposed for calculating the baseline load for residential users [116]. The cluster-based method was applied on historical consumption data to group consumers into representative clusters, where each cluster center represents the groups' baseline load. Customers' DR contribution was quantified in [117] by using an explicit-duration hidden Markov model with differential observations to detect and estimate individual demand responsive load from the aggregate smart meter data.

#### Customer Categorisation for DR Applications

To efficiently implement DR programs and enhance its uptake at utility scale, it is important to understand customers demand characteristics and group them in clusters based on their characteristics. The time-based Markov model was used to formulate the electricity consumption behaviour dynamics[2] and customers were segmented into clusters using clustering by fast search and find of density peaks (CFSFDP) [118], which is a density-based clustering technique. For the applications with large smart meter datasets, a distributed approach based on a divide-and-conquer framework was proposed for the CFSFDP. Finite mixture modelling (FMM) was proposed to cluster users using smart meter data based on the defined attributes describing their daily, seasonal and annual consumption patterns with the optimal cluster number derived using the Bayesian information criterion [119]. In [122], a real-time and online randomized approximation algorithm was proposed to characterise and cluster large smart meter data for DR applications in smart grids. In [120], high dimensional smart meter data was characterised as a linear combination of partial usage patterns using the *k*-singular value decomposition technique. SVM was then applied to classify the customers into residential customers and small and business enterprise based on the derived patterns. In [121], a framework was proposed to classify new customers without consumption data to load profiles based on commercial, government and open data. Six load profiles were initially derived by applying spectral clustering algorithm using weighted kernel principal component analysis on 6,000 customer smart meter data. Random forest classification technique was then used to classify new customers based on the probability that they belong to any of the six derived clusters.

### DR Targeting for Customer Participation

Efficient and effective DR targeting is important for the successful deployment of DR programs. Deriving insights about users demand characteristics using their smart meter data are expected to enhance DR targeting in smart grids. Analysing users' smart meter data at utility scale will require the use of state-of-the-art data analytics techniques. A scalable customer selection procedure formulated as a stochastic Knapsack problem was proposed in [123] for DR targeting using heuristic algorithms. The customer selection procedure

[2] dynamics here, refer to the transitions between consumption levels in adjacent periods

was formulated as a Integer Linear program and a polynomial time approximation scheme was utilised to minimize curtailment error between targeted and achieved demand reduction during a DR event [129]. In this thesis, a $k$-medoid clustering algorithm and dynamic time warping distance measure is proposed to cluster characterised user demand profile for DR targeting [124].

#### Enhancement of Renewable Energy Integration through DR

Smart meter data analytics can be employed in enhancing renewable energy integration in smart grids through enhanced DR implementation. In [130], an algorithm was proposed to overcome solar PV intermittency using heat load models derived from smart meter data. An incentive-based DR program was proposed in [131] to mitigate wind and solar power generation uncertainty in micro-grids. In [125], the fuzzy c-means clustering algorithm was used to group users based on daily demand pattern for ToU DR with an objective to maximize the penetration of renewable energy.

#### DR implementation in smart grids

Data analytics techniques has been recently applied to the development of both price based DR and incentive based DR. An optimal pricing decision mechanism was proposed in [127] with the aid of Q-learning algorithm by learning customers consumption behaviour with respect to changing electricity cost. In [132], a pricing strategy was proposed for DR using reinforcement learning algorithm to learn users response functions. An energy management system was formulated as a reinforcement learning problem in [128] with grid signals coming in the form of either price or incentive based DR. In this thesis, a long short-term memory recurrent neural network is proposed to estimating customers baseline for incentive based DR [126]. A control-based baseline approach is also proposed that takes into consideration the demand rebound effect.

### 3.3.2 DR research focus area and summary of proposed techniques

In this thesis, focus is made on enhancing the performance of DR through 2 key areas:

1. Customer targeting for peak load reduction.
2. Customer demand baseline estimation.

#### Customer targeting for peak load reduction

For the successful roll out of DR programs, the need to identify and target customers with high likelihood of participation is paramount. The availability of smart meter data provides a starting point for which key features can be formulated and extracted to enable direct targeting of customers. An unsupervised learning approach is proposed to understand similarities and clusters of customers based on their demand profile. The approach includes the use of the *k*-medoid technique and the dynamic time wraping distance measure. The rationale behind proposing an Unsupervised method for this problem boils down to the nature of the problem for which no input and output examples are available. Clustering the characterised customer demand profile uncovers groupings for which the demand profile can be analysed as to their suitability for peak load reduction. The distance measure proposed takes into consideration the variability in similar customers demand phase from time to time. The work is presented in chapter 4.

#### Customer demand baseline estimation

Customer baseline estimation primarily quantify the level of a participating DR customers contribution to demand reduction during a DR event. The baseline estimation gives a basis to which DR benefits is allocated to participating customers as well as the utility company. The need for fairness in benefit allocation makes the need for an accurate estimation method important. Two machine learning techniques is proposed in this thesis to estimate customer baseline which are supervised learning approach in chapter 5 and an unsupervised learning approach in chapter 6. The proposed supervised learning approach proposed in chapter 5 is a LSTM RNN technique trained on historical non DR days customer demand data. In chapter 6, a clustering technique (unsupervised learning) is utilised as part of the control-based model proposed for estimating demand baseline.

# Chapter 4

# Smart Meter Data Characterisation and Clustering for Peak Demand Targeting in Smart Grids

The increasing popularity of smart meters deployed at customer sites provides a vital opportunity for network operators to effectively target customers with DR programs aimed at peak demand reduction. Defining the right features for customers smart meter data is the first critical step of achieving an effective data driven DR solution. Firstly, a characterisation model is proposed for peak load targeting of consumers. The proposed feature characterisation describes customers' demand variations over one day period with consumption levels ranging from 0 to 1. A $k$-medoid clustering algorithm and a dynamic time warping (DTW) distance measure is proposed to cluster the characterised smart meter data. The COP index cluster validation technique is used to derive the optimal number of clusters. The proposed model is applied to the publicly available Irish smart meter data, and results show a robust targeting of customers for peak load reduction.

In addition to targeting customers for peak load reduction, the second part of this chapter focuses on targeting customers for local peak reduction. Local peak in this chapter is defined as lower peak points on a customers demand profile. Local peaks are of particular interest especially for DR programs aimed at flattening the aggregate demand curve and reducing the need for peaking generators for short period spans. Fig. 4.1 shows the description of a customers normalised daily demand profile with peak period ($p_p$) and local peaks ($p_{lp1}, p_{lp2}, p_{lp3}, p_{lp4}$) displayed. A novel set of features is proposed for targeting customers for local peak load reduction. The analysis of the proposed methodology shows an effective process of targeting customers based on the potential of each customers to contribute to local peak reduction.

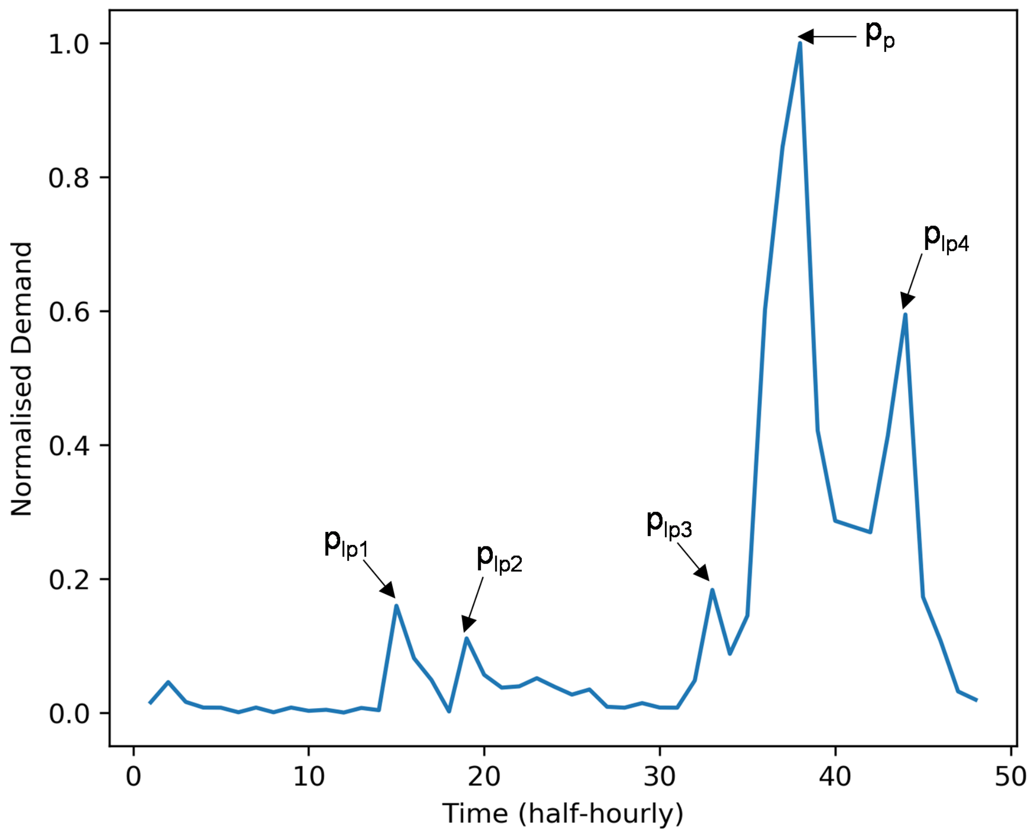


Figure 4.1: Demand profile showing peak and local peaks.

## 4.1 Introduction

The increasing demand from residential, commercial and industrial customers poses a challenge to electricity supply and smart grids. One of the key obstacle is meeting demand during peak period. Traditionally, the increase in peak demand will normally be met with increased generation, transmission and distribution capacity. This solution however is inefficient as the peak period demand being met by investing in capacity upgrades only spans through a short period of time relative to the demand for the day. DR provides an opportunity for the peak demand increase to be curtailed or deferred to other periods of lower demand. In enabling an effective DR solution to rising demand, an efficient and effective process of identifying and targeting customers for DR programs becomes paramount. The availability of smart meters at customers site aids the effectiveness of DR programs as it provides the opportunity of using a data-driven approach in meeting the set objective. Defining the right features for customers smart meter data is the first critical step of achieving an effective data driven DR solution.

In characterizing smart meter data for DR, it is vital to define the right attributes to represent the data, as this can affect the quality of knowledge extracted from data analytics techniques. Common approaches used in characterizing smart meter data include dimensionality reduction and granularity scale reduction. One of the main advantages of these approaches is to reduce the size of the smart meter data for quick processing and knowledge extraction. Chelmis *et al.* [133] used principal component analysis (PCA) to characterize smart meter data for clustering. PCA linearly maps the smart meter data to a lower-dimensional space such that the variance of the transposed data is maximized [134]. Chen *et al.* [135] also implemented the PCA dimension reduction algorithm on smart meter data and modeled a price-based DR scheme using the clustered result of the derived components. Independent component analysis (ICA) was used to characterize customers' smart meter data with *k*-means clustering technique to group the derived representative components [136]. Ikeda and Nishi [137] utilized sparse coding learning techniques to characterize smart meter data and applied hierarchical clustering technique for customer grouping. The sparse coding approach was also used to characterize smart meter data in [138]. Wang *et al.* [120] characterized smart meter data using non-negative K-SVD sparse representative techniques. In their work, customer load profiles were decomposed into a linear combination of partial usage patterns and classified into residential customers, and small and medium enterprises customers using the support vector machine classification method. A granularity scale reduction approach was utilized by Haben *et al.* [119], where smart meter data was characterized using their relative average demand over four periods of the day, mean standard deviation over the year, seasonal score and a weekend versus weekday score as attributes. Wang *et al.* [118] also reduced daily smart meter data into four periods and focused their analysis on users' consumption dynamics by studying the transition of consumption levels between adjacent time periods of like-days.

The use of dimensionality reduction techniques and granularity scale reduction approaches in characterizing smart meter data can undermine accounting for the level of each individual customer's contribution to peak load. This challenge is evident in the information loss encountered in reconstructing the original demand profile from the reduced data. In order to address this issue, a shape-based characterization method of customers' smart meter data is proposed to specifically describe each customers demand variability. The main contributions from this chapter are as follows:

- The characterization of smart meter data is proposed based on their per-unit variability for peak demand targeting.
- A shape-based representative clustering technique ($k$-medoid clustering with DTW distance measure) is proposed to group customers for peak demand targeting.
- The COP index cluster validation technique is used to derive the optimal number of clusters.
- The performance of the proposed methodology is evaluated using the publicly avail-

able Irish smart meter data [139] and the suitability of the derived clusters is shown for peak load reduction.

- A novel set of features is proposed for targeting customers for local peak load reduction.
- A demonstration of customer targeting for local peak load reduction is presented.

The rest of the chapter is organized as follows: in Section II, the proposed methodology and results for peak load targeting of customers is presented. Section III presents the methodology and results for local peak load targeting of customers. The conclusion of the chapter is presented in Section IV.

## 4.2 Customer Targeting for Peak Load Reduction

A data mining framework is proposed for characterizing and identifying consumers for peak load reduction in smart grids using smart meter data collected from customers. Fig. 5.3 shows the main components of the proposed framework. The attributes are defined, and the smart meter data are characterized. A $k$-medoid partitional clustering algorithm is proposed to group the characterized smart meter data based on the defined attributes. For the clustering, a shape-based similarity measure is proposed to evaluate the distance between customers' demand profiles. The optimal cluster number is derived using the COP index. Each component of the framework is explained in the rest of the section.

### 4.2.1 Feature Development

Representative demand profiles are selected to describe the energy consumption characteristics of each consumer. The focus of the analysis is on the demand variability of each consumer. Hence, the demand profile is normalised within the range 0 and 1 as follows:

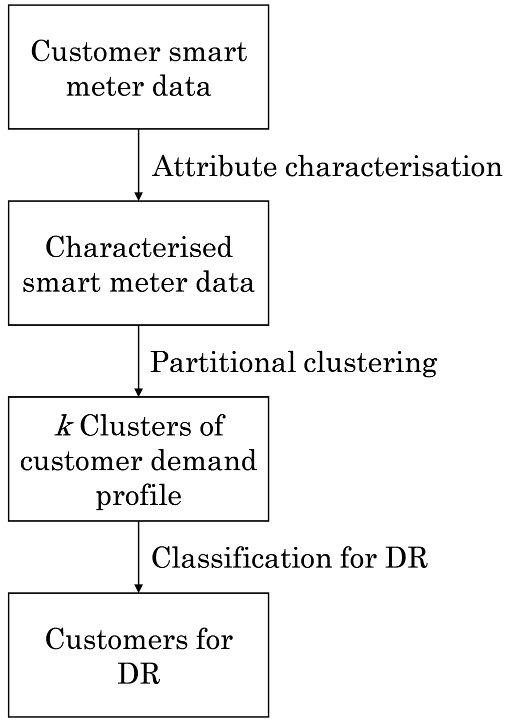


Figure 4.2: Framework for customer DR targeting.

$$p_i^{'} = \frac{p_i - p_{min}}{p_{max} - p_{min}}, \tag{4.1}$$

where $p_i^{'}$, $p_i$, $p_{min}$, $p_{max}$ represent the normalised, actual, minimal and maximal energy consumption of user $i$ over a period of time, respectively (48 data points representing 30-minutes interval through the day). This representation of each consumer's smart meter data defines the attribute characterization. The main reason for taking this representation approach is to clearly identify customers by their per unit demand variability and peak demand contribution.

### 4.2.2 $k$-Medoid Partitional Clustering Algorithm

$k$-medoid partitional clustering with DTW distance measure is proposed for clustering the characterized smart meter data. $k$-medoid assigns a set of $N$ time series objects to $k$ groups such that the intra-cluster distance between objects of the same group is minimized while the inter-cluster distance is maximized. The k-medoid clustering algorithm has the following procedures [37]:

- $k$ time series data is randomly initialized from the set of $N$ customer demand profile as medoids.
- The remaining demand profiles $(N-k)$ is distributed to the nearest medoid using the DTW distance measure.
- For each cluster, randomly select a non-representative time series object, $o_{ran}$.
- Calculate the total cost, $S$ of swapping the representative cost with $o_{ran}$.
- If the total cost of swapping is less than 0, i.e., $S<0$, $o_{ran}$ becomes the representative object, i.e., medoid.
- This is computed iteratively until there is no change.

### 4.2.3 Euclidean and Manhattan Distance Measure

In calculating the distance between time series data, the two most commonly used distance measure is the Euclidean and Manhattan. The general formula for calculating the Euclidean and Manhattan distances between two points in space is given by 4.2 and 4.3 where $n$ represents the dimension of the series.

$$d_{euc} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}, \tag{4.2}$$

$$d_{mann} = \sum_{i=1}^{n} |p_i - q_i|. \tag{4.3}$$

### 4.2.4 Dynamic Time Warping

DTW is a shape-based distance measure technique developed for time series clustering. It measures the distance similarity between each time series by finding the optimal non-linear alignment between them [140]. Finding shape similarity through traditional distance measures (e.g., Euclidean) for two time series functions $f(t)$ and $g(t)$ means distances are measured between adjacent periods of the time series as shown as follows:

$$d(f(t), g(t)) = \sqrt{\sum_{i=1}^{n} [f(i) - g(i)]^2}, \tag{4.4}$$

where $i = (1, 2, 3, \cdots, n)$.

The traditional distance measure presents a relatively higher distance score for time series data of similar shape but different phases or time delay. Time delay is typical of similar consumption patterns as electrical devices are generally switched on at different periods of the day. An example is a situation where two consumers with similar consumption pattern switch on a heating device at 6:00 and 6:30, respectively. This situation means similar time series data can end up in different clusters. To calculate the DTW for two time series $f(t)$ and $g(t)$, expressed in (4.5) and (4.6), respectively as

$$f(t) = f(1), f(2), f(3) \cdots, f(n), \tag{4.5}$$

$$g(t) = g(1), g(2), g(3) \cdots, g(n), \tag{4.6}$$

the distance $d(1)$ and $d(n)$ is first computed according to (4.7) and (4.8) as the first and last points of the warping path must be aligned, i.e.,

$$d(1) = |f(1) - g(1)|, \tag{4.7}$$

$$d(n) = |f(n) - g(n)|. \tag{4.8}$$

The distance $d(i)$ as shown in (4.9) selects the minimum distance between the point $i$ in $f(t)$ to points $i-w, i-w+1, \cdots, i, \cdots, i+w$ in $g(t)$ where $w$ is defined as the time window constraint. Since the first and last points of $f(t)$ and $g(t)$ must be aligned, for points in $g(t)$, $i-w > 1$ and $i+w < n$, we have

$$\begin{aligned} d(i) = min(&|f(i)-g(i-w)|, |f(i)-g(i-w+1)|, \\ &\cdots, |f(i)-g(i)|, \cdots, |f(i)-g(i+w)|). \end{aligned} \tag{4.9}$$

The time window constraint provides a phase range for which the distances between $f(t)$ and $g(t)$ is computed. The DTW presented in (4.10) is the total distance along the optimal path as shown in Fig. 4.3 and is derived using dynamic programming.

$$D = \left(\sqrt{\sum_{t=1}^{n} d(t)^2}\right). \tag{4.10}$$

### 4.2.5 Optimal Cluster Number Selection: COP Index

COP index is defined based on its properties that is, Context-independent, Optimality and Partiality [141]. The COP index cluster validation technique is used to derive the optimal number of clusters. The COP index is defined as the ratio of the intra-cluster variance (cohesion) to the inter-cluster variance (separation) [141]. The intra-cluster variance is the average DTW distance between the time series object in a cluster and its medoid while the inter-cluster distance is the DTW distance between the furthest time series object in the clusters. For dataset $X$ containing $N$ time series objects i.e. $X = \{x_1, x_2, x_3, \cdots, x_N\}$ and partitioned into $k$ disjoint clusters $C = \{c_1, c_2, c_3, \cdots, c_k\}$, the COP index is derived using (4.11) below,

$$COP(C) = \frac{1}{N} \sum_{c_k \in C} |c_k| \frac{\frac{1}{|c_k|} \sum_{x_i \in c_k} D(x_i, c_{km})}{min_{x_i \notin c_k} max_{x_j \in c_k} D(x_i, x_j)}, \tag{4.11}$$

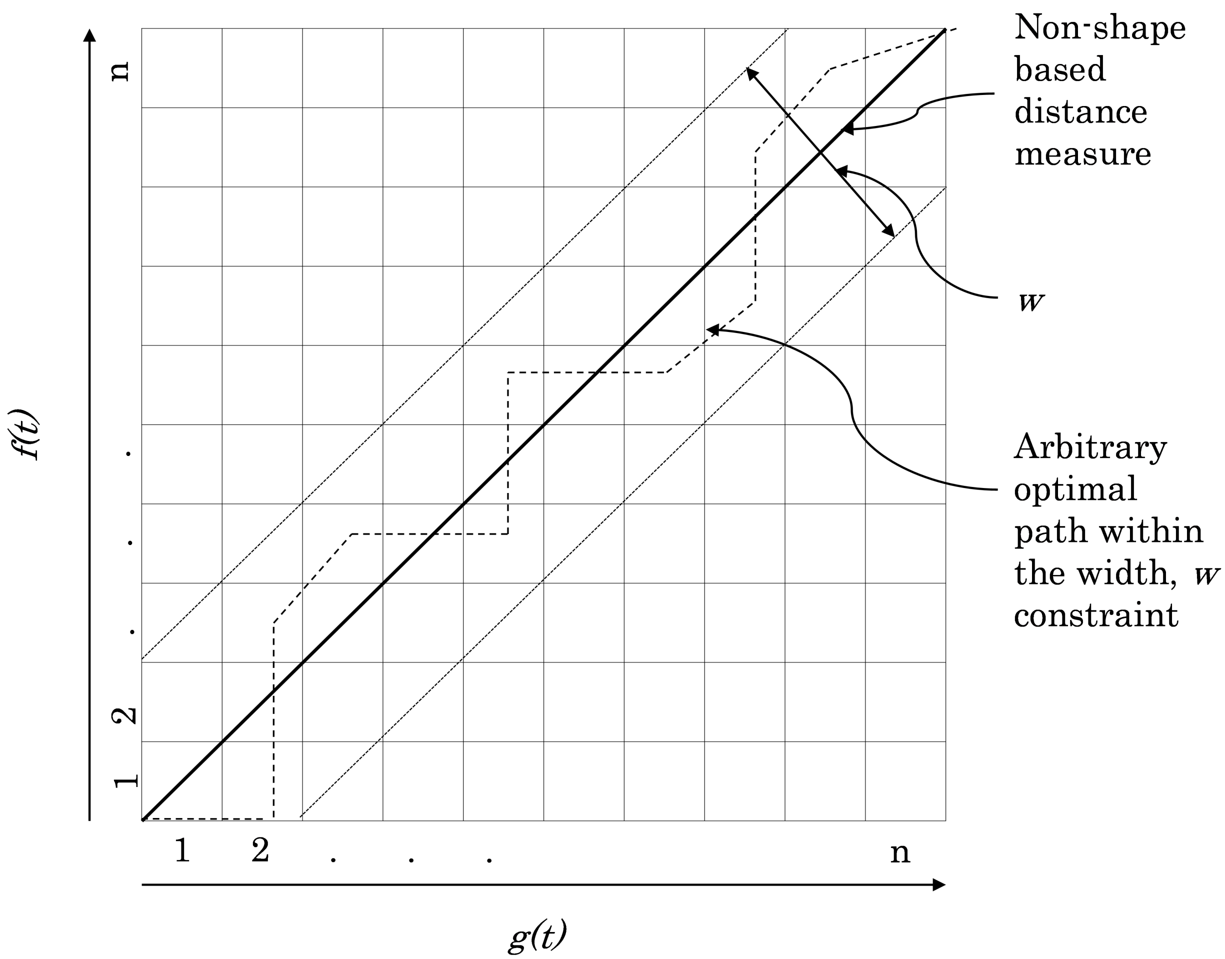


Figure 4.3: Dynamic time warping path.

where $c_k$ is cluster $k$ and $c_{km}$ is the medoid of cluster $k$. The optimum cluster number for the dataset is selected from a range of $K$ for which the COP index is minimum.

### 4.2.6 Identifying Clusters for Peak Demand Reduction

A peak demand potential factor ($\tau$) is proposed. This factor is presented in 4.12,

$$\tau = \frac{\sum_{i=t-w/2}^{t+w/2}(p_i)/(w+1)}{\sum_{i=1}^{T}(p_i)/T}, \tag{4.12}$$

where $w$ is the DTW window and even, $t$ time of peak demand and $T$ is the time period for the profile.

Firstly, the average of the normalised demand between $t - w/2$ and $t + w/2$ is computed. The computed average is divided by the overall average of the normalised demand profile of the cluster. The higher the $\tau$ value, the more the potential for peak load reduction for the cluster.

Compared to the traditional method of estimating peak load reduction potential i.e. peak to average ratio, the $\tau$ takes into consideration nearest neighbour demand levels. Given the high variability in demand especially from residential customers within clusters, the proposed $\tau$ presents a more representative peak load reduction potential for the constituent customers as this captures the within cluster peak load variation around the defined window $w$.

### 4.2.7 Results and Discussion

The dataset considered in this study comprises of 99 customer smart meter data taken from the Electric Ireland and Sustainable Energy Authority of Ireland smart meter dataset [139]. The *dtwclust* package, developed for the R statistical software [142] was employed to implement the smart meter data clustering. A representative daily load profile of 30 minutes' granularity was selected to represent the typical demand profile of each customer. A window size of 2 was selected for the DTW distance measure to allow for a 1-hour time lag in appliance usage differences among customers with similar demand characteristics. The smart meter data is processed using the proposed feature characterization and the $k$-medoid clustering algorithm with DTW distance measure is applied.

In order to determine the optimal number clusters, the COP index is derived for $k$ values ranging from 2 to 9. Fig. 4.4 shows the plot of the COP index for the ranges of $k$ values. As the COP index is minimum when $k$ is 8, the optimal cluster number is selected as 8.

The aggregate demand for the 99 customers is presented in Fig. 4.5 and the peak demand is at period 39 for which the aim is to target customers for peak demand reduction.

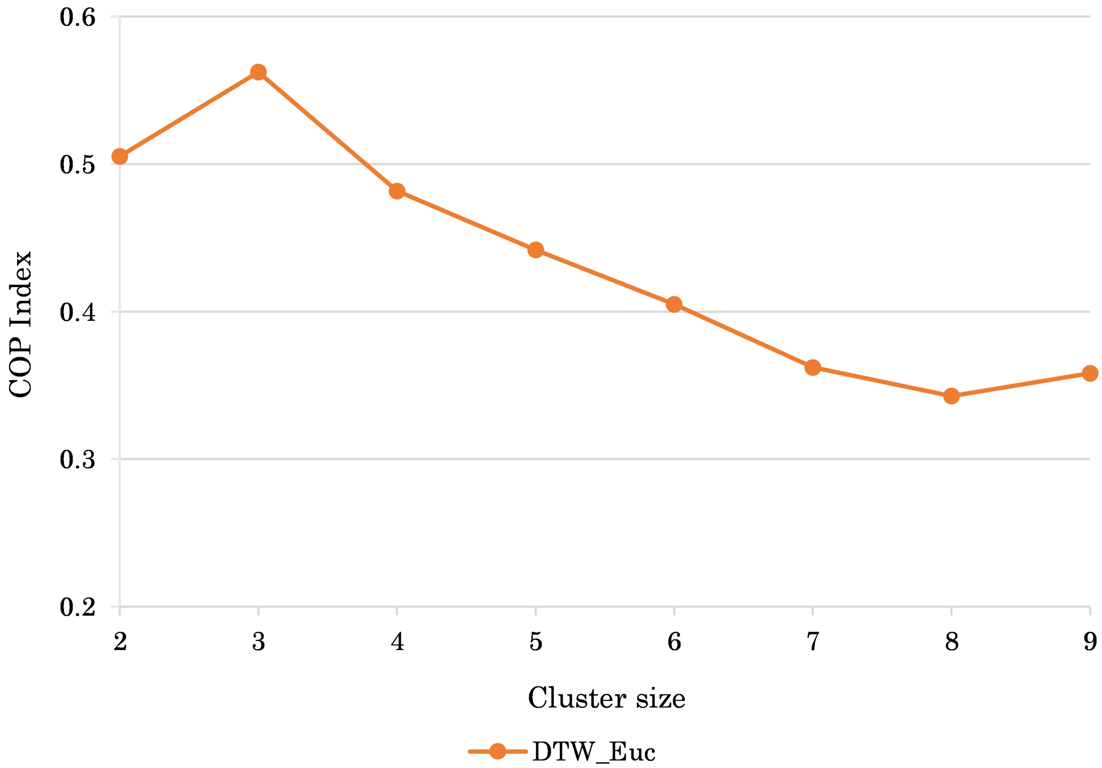


Figure 4.4: COP index cluster validity.

Fig. 4.6 shows the medoid of the 8 derived clusters. The peak demand potential factor is computed for each for each of the clusters and presented in table 4.1 below.

From Table 4.1, it can be observed that customers in clusters 2, 3, 4, 5 and 8 show potential for peak load reduction. Cluster 3 shows the highest potential owning to the most $\tau$ value of 5.21.

Fig. 4.7 shows the distribution of customers among the clusters with cluster 3 showing 19% of customers at a medoid $\tau$ level of 5.21. Clusters 2 and 8 which has a medoid $\tau$ level of 2.32 and 2.18 respectively have 20% and 30% of the customers.

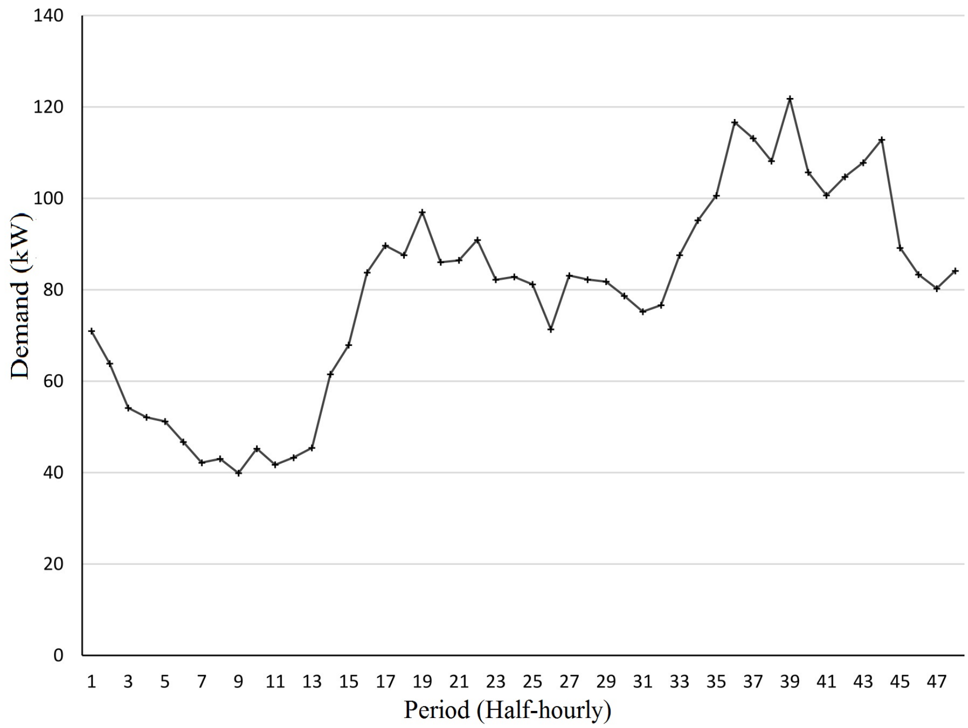


Figure 4.5: Aggregate customer demand.

**Effect of Distance Measure on Clustering Performance - Case for Manhattan Distance Measure for Large Customer Demand Profile Clustering**

The cluster performance is compared using 4 distance measure approaches. These are highlighted below

- DTW distance measure with Euclidean
- DTW distance measure with Manhattan

Table 4.1: Cluster medoids' peak demand potential factor.

| Medoid number | $\tau$ |
|---|---|
| 1 | 0.43 |
| 2 | 2.32 |
| 3 | 5.21 |
| 4 | 1.13 |
| 5 | 1.61 |
| 6 | 0 |
| 7 | 0.14 |
| 8 | 2.18 |

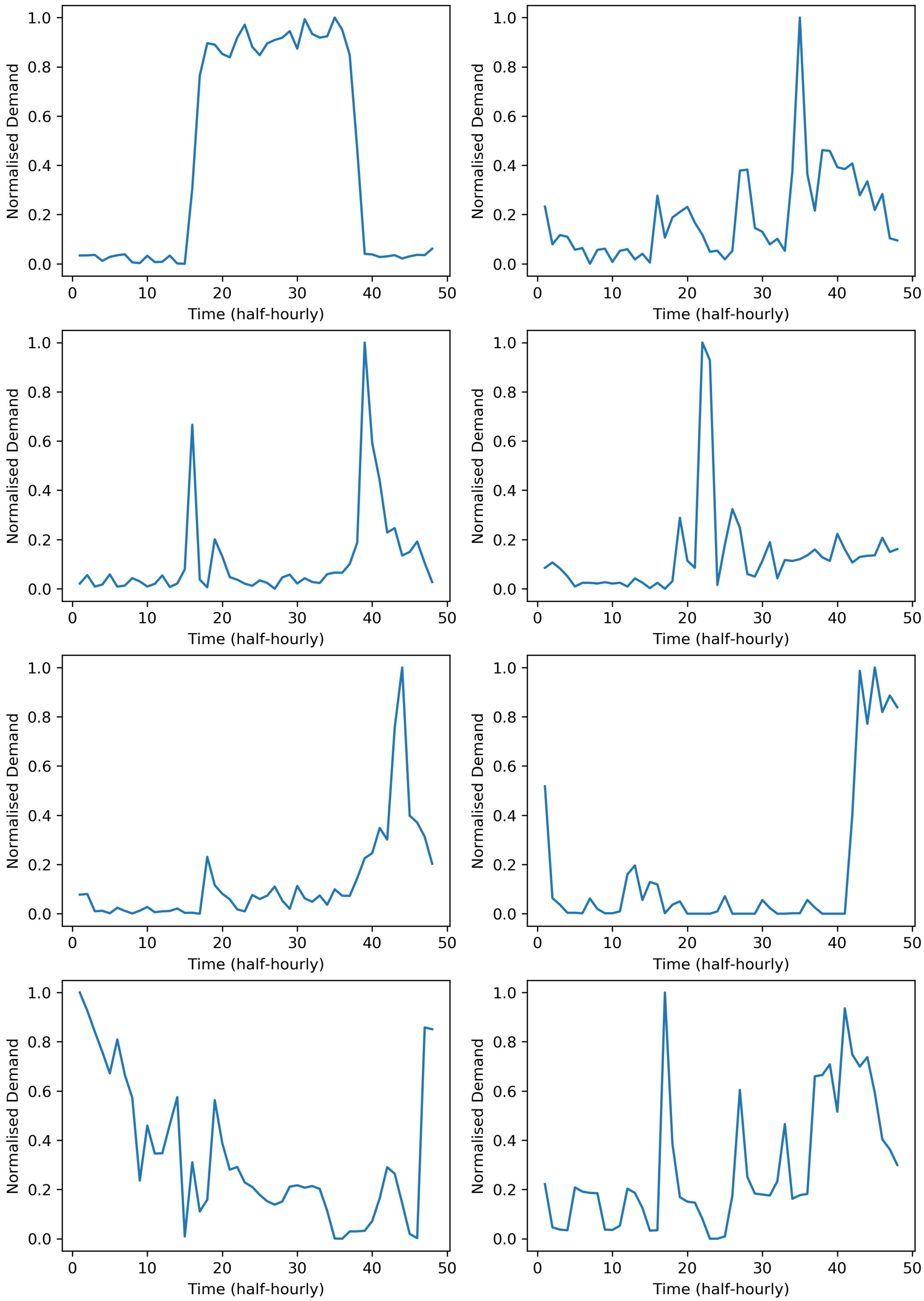


Figure 4.6: Customer clusters.

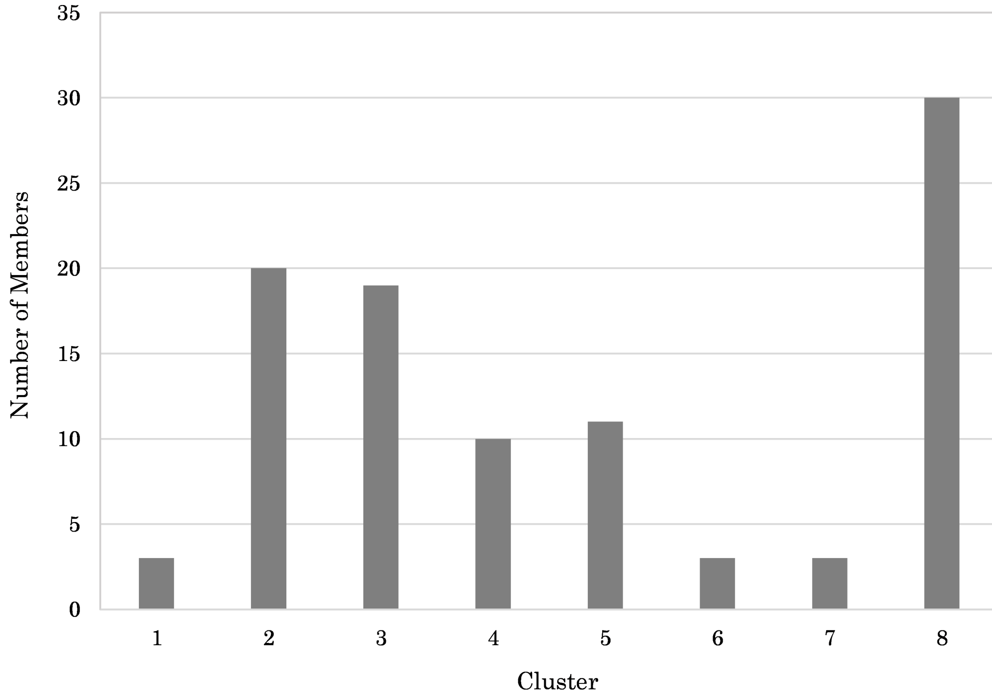


Figure 4.7: Customer cluster distribution.

- Direct Euclidean measure
- Direct Manhattan measure

The DTW distance measure for both Euclidean and Manhattan gave an optimal cluster number of 8 compared to 7 obtained in the direct Euclidean and Manhattan distance measure approach. The COP index plot is presented in Fig. 4.8.

The distribution of the distance between all the cluster members and their medoids is summarised in Fig. 4.9.

The 2 DTW distance approaches show similar distribution of distances across all cluster members. This observation applies to the direct distance measure approaches as well. Given the less computational expensive nature of the Manhattan distance measure (in comparison to Euclidean distance measure) and the similarity observed in the distance distribution across all cluster members, the DTW Manhattan distance measure is proposed for clustering large smart meter datasets.

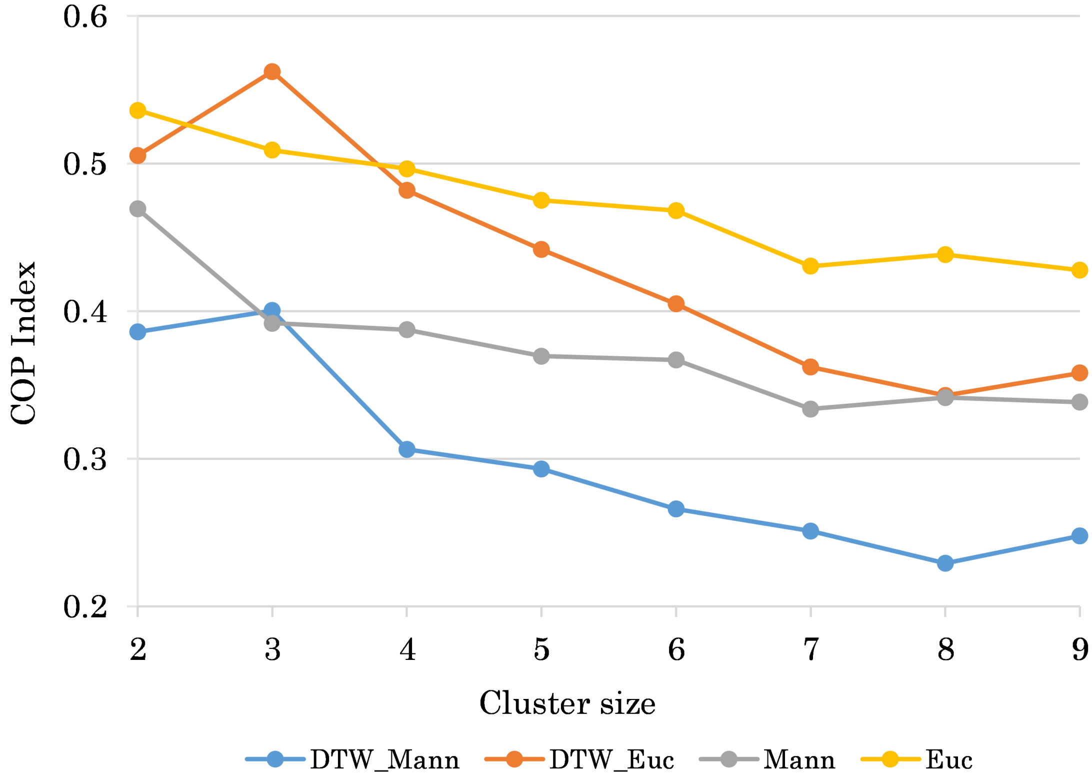


Figure 4.8: COP index validity measure for DTW Euclidean; DTW Manhattan; direct Euclidean; and direct Manhattan.

#### Effect of Shape Based Distance Measure on Peak Targeting

The $\tau$ value of each of the derived clusters in the direct distance method is compared to that of the DTW distance method. Table 4.3 below shows the $\tau$ value for each of the clusters for all the four distance measure approach.

To analyse the performance of the distance approaches, 30 customers is taken to be targeted for peak load reduction from the pool of 99 customers. Customers are firstly pooled from clusters with the highest $\tau$ medoid value up to the lowest until the target number of customer is met. Direct Manhattan distance measure is used as a baseline to compare the average distance of the cluster members to their medoids. Table 4.3 shows the average Manhattan distance for each of the clusters in the 4 distance measure approach. Any of the 4 distance approaches could be used as a baseline and Manhattan was utilised because of its simplicity.

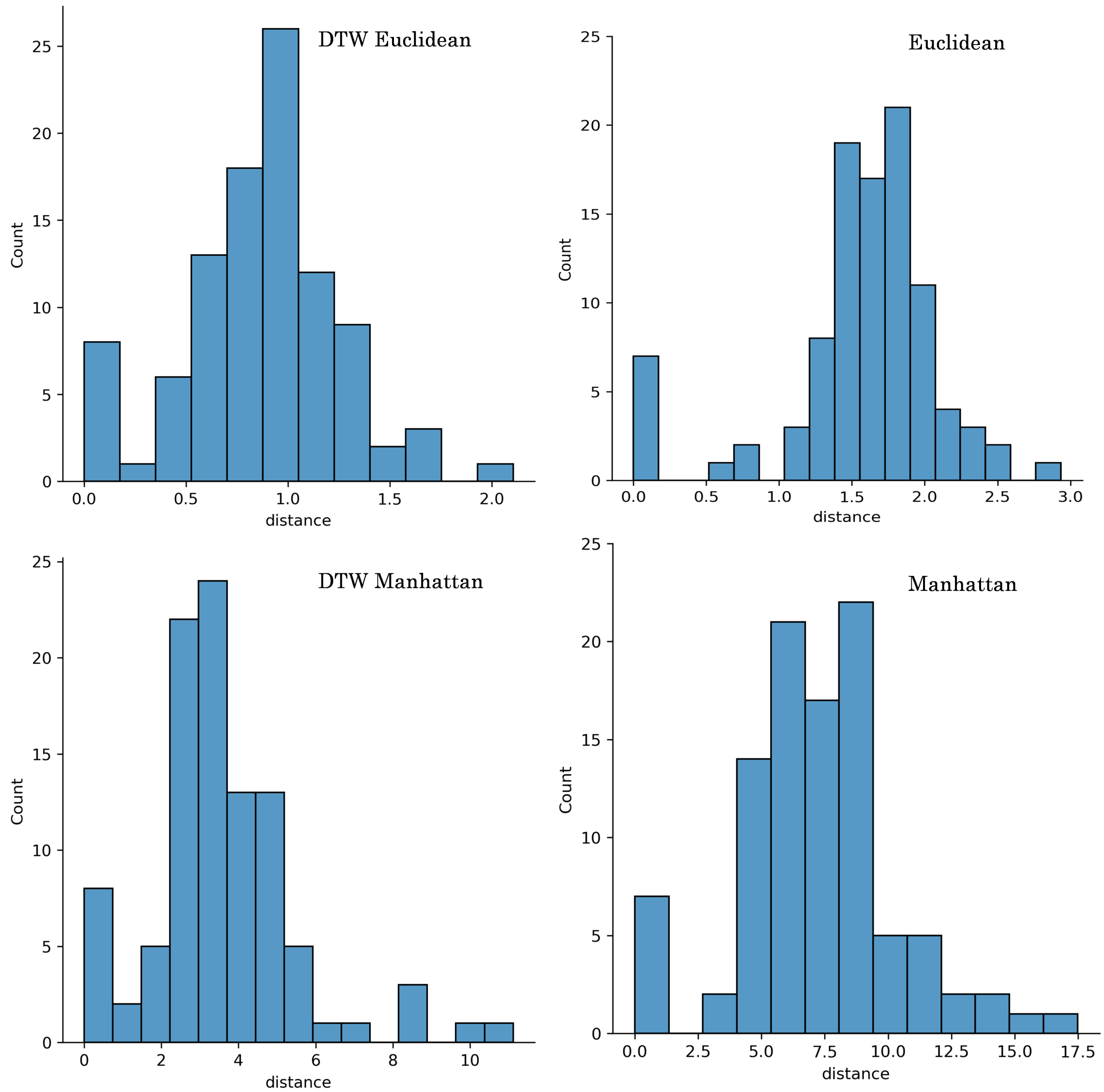


Figure 4.9: Distance distribution of all 4 distance measure approaches.

Table 4.4 shows the member distribution in each of the clusters for the 4 distance approaches.

In meeting the 30 customers targeting objective for peak load reduction, taking the DTW with Euclidean distance measure approach, the 19 customers in cluster 3 is selected first as its medoid has the highest $\tau$ value of 5.21. The remaining 11 customers is derived from cluster 2 as it has the second highest $\tau$ value of 2.32. In comparison to the direct Manhattan distance measure, all 30 customers can be selected from the 40 customers

Table 4.2: Cluster medoids' peak demand potential factor for each distance measure.

| Medoid number | $\tau$ (DTW_Euc) | $\tau$ (DTW_Mann) | $\tau$ (Euc) | $\tau$ (Mann) |
|---|---|---|---|---|
| 1 | 0.43 | 1.61 | 0.66 | 1.61 |
| 2 | 2.32 | 1.23 | 4.3 | 4.3 |
| 3 | 5.21 | 2.05 | 1.27 | 5.21 |
| 4 | 1.13 | 1.13 | 0.83 | 0.83 |
| 5 | 1.61 | 3.08 | 1.37 | 1.37 |
| 6 | 0 | 2.08 | 2.08 | 1.93 |
| 7 | 0.14 | 0.14 | 0.14 | 0.14 |
| 8 | 2.18 | 2.14 | | |

Table 4.3: Average base Manhattan distance measure for each clusters.

| Medoid number | $d_{Mann}$ (DTW_Euc) | $d_{Mann}$ (DTW_Mann) | $d_{Mann}$ (Euc) | $d_{Mann}$ (Mann) |
|---|---|---|---|---|
| 1 | 7.68 | 25.4 | 9.47 | 25.4 |
| 2 | 7.53 | 8.82 | 7.55 | 7.55 |
| 3 | 5.41 | 6.62 | 6.72 | 7.32 |
| 4 | 7.67 | 7.67 | 8.88 | 8.88 |
| 5 | 5.13 | 9.07 | 10.14 | 10.14 |
| 6 | 5.23 | 14.32 | 14.32 | 18.36 |
| 7 | 10.8 | 10.8 | 10.8 | 10.8 |
| 8 | 10.42 | 11.34 | | |

Table 4.4: Cluster member size for each distance measure.

| Medoid number | $d_{Mann}$ (DTW$_{Euc}$) | $d_{Mann}$ (DTW$_{Mann}$) | $d_{Mann}$ (Euc) | $d_{Mann}$ (Mann) |
|---|---|---|---|---|
| 1 | 3 | 3 | 6 | 3 |
| 2 | 20 | 18 | 21 | 26 |
| 3 | 19 | 23 | 39 | 40 |
| 4 | 10 | 8 | 15 | 12 |
| 5 | 11 | 19 | 5 | 7 |
| 6 | 3 | 16 | 10 | 8 |
| 7 | 3 | 3 | 3 | 3 |
| 8 | 30 | 9 | | |

in cluster 3 with $\tau$ value of 5.21 however, the base distance is 7.32 compared to DTW Euclidean of 5.41 and 7.53 for clusters 3 and 2, respectively. In accessing the performance of the distance approaches, it is important to take into consideration both the $\tau$ value and the cluster base distance as $\tau$ gives the peak DR potential and cluster base distance gives an indication of how close the cluster members are to their medoid.

## 4.3 Local Peak Load Targeting of Customers

During local peak periods, peaking generators are often utilised to meet the short period of demand spikes. These generators are usually gas powered and expensive. A cost effective and environmentally friendly alternative is to utilise customers flexibility to curtail or shift demand to achieve a flattening of the aggregate demand profile. Targeted allocation of storage to customers in order to enable them as prosumers in a competitive market can also benefit from the proposed methodology especially where customers demand profile show high demand during DR periods to maximise on-site utilisation of storage.

This section presents a novel approach and method to targeting customers for local peak reduction by proposing two features to classify customers based on their potential for contributing to DR.

### 4.3.1 Feature Development

A two-dimensional parameter is proposed as features for each customers. Firstly, the parameters $\alpha$ and $\beta$ is defined in (4.13) and (4.14) below as

$$\alpha = \frac{p_{lp} - p_{lavg}}{p_{lavg}}, \tag{4.13}$$

$$\beta = \frac{p_{lavg}}{p_{avg}}, \tag{4.14}$$

where $p_{lp}$, $p_{lavg}$, and $p_{avg}$ are local peak demand, average demand within local window and average demand for the day, respectively.

Fig. 4.10 shows a visual description of how $\alpha$ and $\beta$ is derived. Given the objective of enabling DR at a local peak $p_{lp}$, a local peak window ($\psi$) is defined.

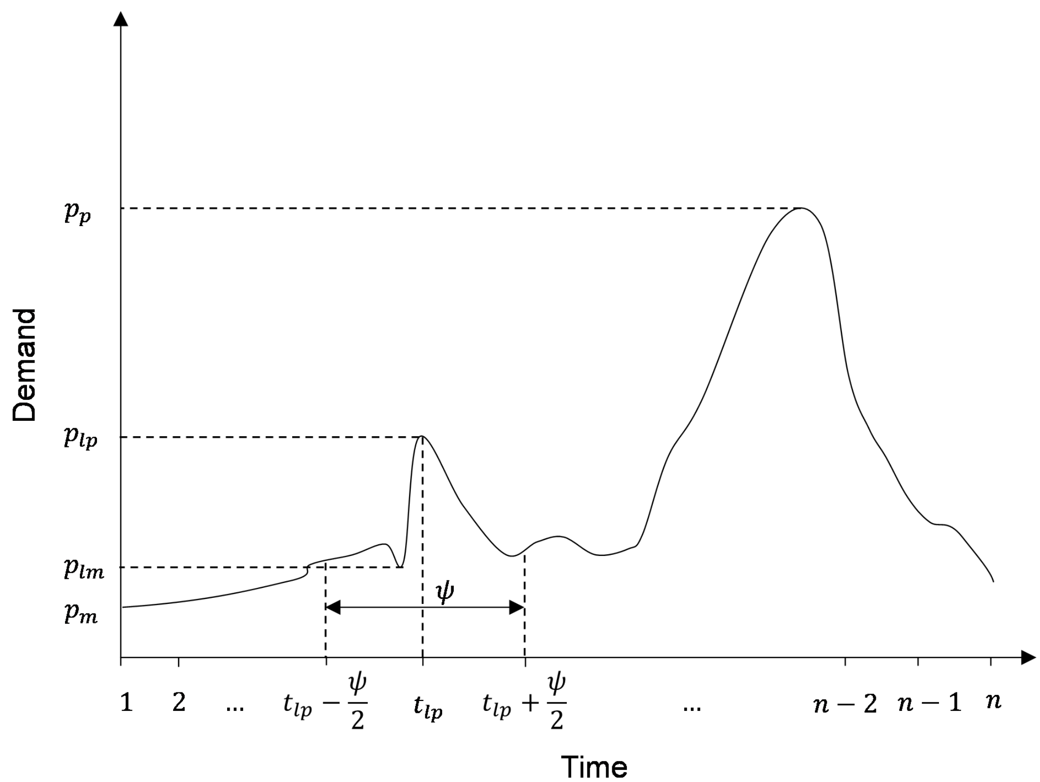


Figure 4.10: Demand profile illustration showing local peak, local peak window and actual peak.

In selecting a value for $\psi$, it is important to note the length of the DR event period. Given a DR event at a local peak time $t_{lp}$, neighbourhood demand points both prior to and post DR event time makes up the DR event span. For a half-hourly granular demand data, a $\psi$ value of 4 is proposed. This value captures an hour prior to and after DR.

As $\psi$ tend towards $n(T)$ (i.e. 48 in the case of half-hourly), the $\alpha$ feature tends toward 0 and $\beta$ tends toward local peak to daily average demand ratio.

### 4.3.2 Demand Response Resource Estimation for Local Peak Reduction

To determine the estimate of the DR resource for each customer, 4 quadrants is defined about the $\alpha$ and $\beta$ feature value of the aggregate demand. The quadrants are presented in Fig. 4.11 with points representing the $\alpha$ and $\beta$ values of the aggregate and all customers.

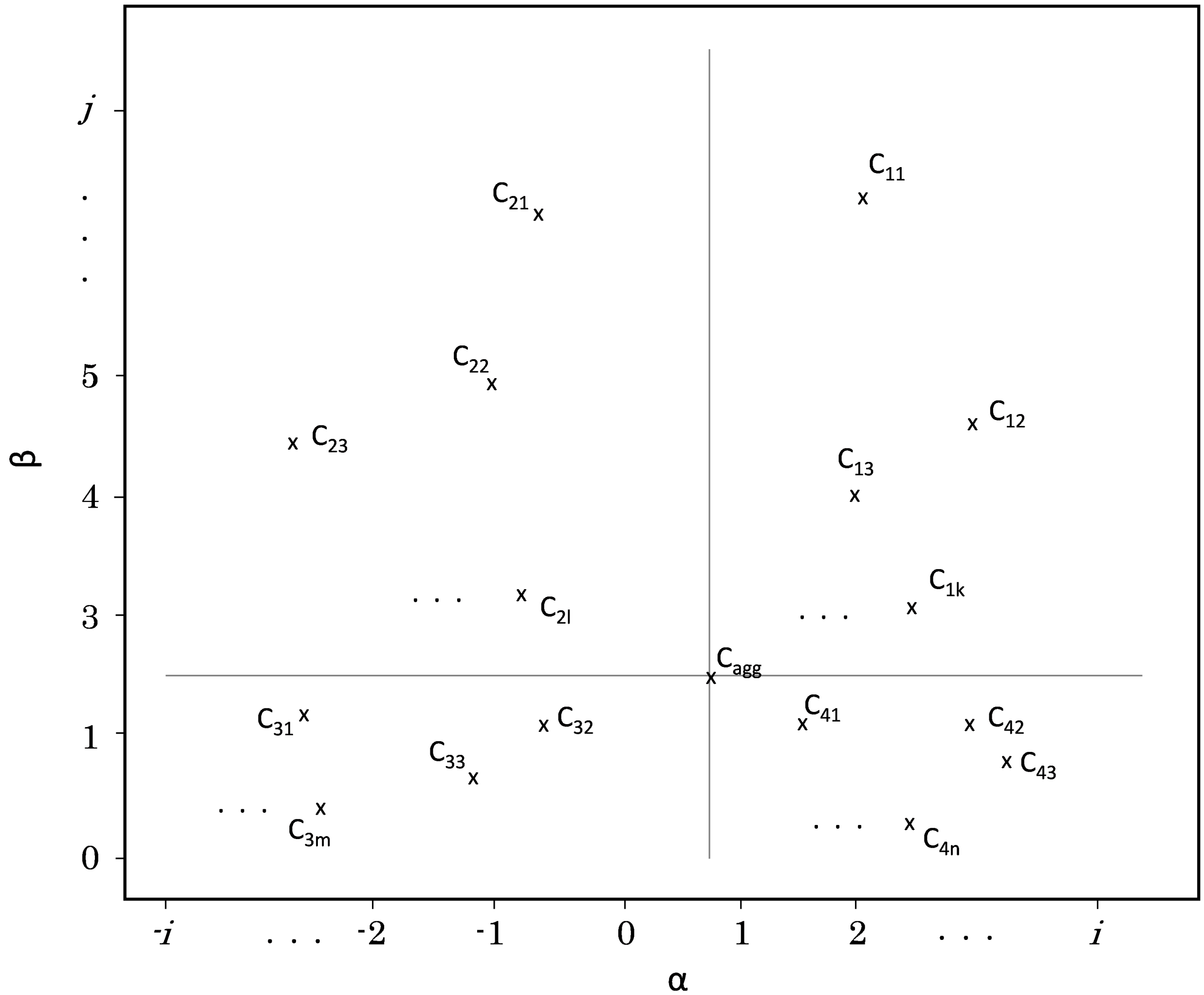


Figure 4.11: Quadrant plot of $\beta$ versus $\alpha$ parameter for customers and their aggregate demand.

The higher the $\alpha$ value, the more the potential to shift demand within the DR event window. This resource relationship stems from the magnitude of difference between demand at local peak and the average demand for the DR window. A positive $\alpha$ value for a customer implies demand at local peak is higher than the within DR event average demand thereby creating opportunity for the customer to directly flatten their within DR event demand. A negative $\alpha$ however implies a lower demand at local peak for the customer compared to their average demand during the DR event window thereby meaning that reducing demand at the peak demand won't enable flattening of the within DR event demand for the customer.

The higher the $\beta$ value, the more the potential to shift demand outside the DR event window for the purpose of meeting a demand profile flattening objective. A $\beta$ value less than 1 implies the average demand within DR event window is less than the overall average demand meaning low potential for demand profile flattening.

The significance of demand shift within DR event compared to shifting outside of DR event comes from the diverse shifting constraint of different types of load. An example is a heating or cooling load's temperature set point controlled for a short period during the local peak having its demand deferred to a period which is still within the DR event window. Compared to a washing machine where a washing activity can be directly deferred to a period entirely outside of the DR event.

The DR resource potential of each quadrant is described in the items below.

- Quadrant I: In this quadrant, customer's $\alpha$ and $\beta$ values are higher than that of the aggregate demand. This scenario implies a higher potential for within DR event levelling as well as high potential for demand shifting for the objective of flattening the demand profile.

- Quadrant II: Quadrant II shows customers with $\beta$ values higher than that of the aggregate demand but $\alpha$ value lower than that of the aggregate demand. This scenario showcase a high potential for demand shifting with the objective of flattening the demand profile while showing a low potential of levelling demand within DR event.

- Quadrant III: Here, both $\alpha$ and $\beta$ values are lower than that of the aggregate demand.

This scenario implies a low potential of levelling demand within DR event as well as shifting demand with the objective of flattening the demand profile.

- Quadrant IV: The values for $\beta$ in this quadrant are lower than that of the aggregate demand but $\alpha$ value is higher than that of the aggregate demand. This situation implies a low potential for demand shifting with the objective of flattening the demand profile while the within DR event window has a high potential of levelling demand.

### 4.3.3 Results and Discussion

Fig. 4.12 shows the aggregate demand for all 99 customers. The peak demand is at period 37. For this study, the local peak at period 16 is considered.

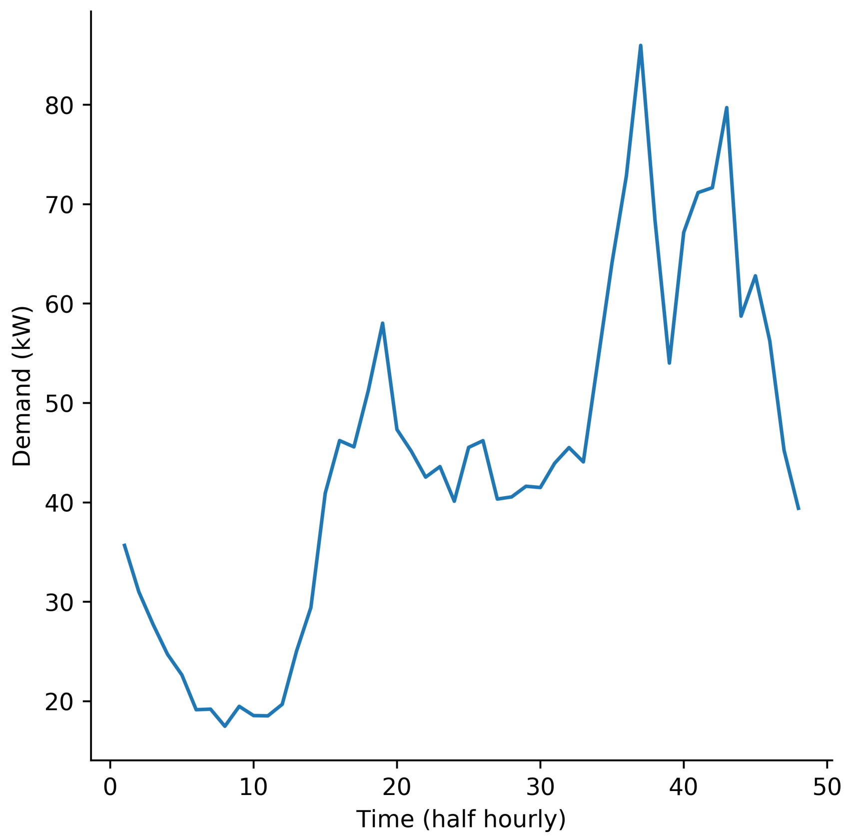


Figure 4.12: Aggregate demand for 99 customers on day 253.

The parameters $\alpha$ and $\beta$ is computed for both the aggregate demand profile and each customers profile a day prior to DR day. The local window $\psi$ is taken as 4 for this study.

Fig. 4.13 shows the scatter plot of $\alpha$ versus $\beta$ for all 99 customers on day 253 (the like day before DR event). The aim is to target 30 customers for local peak reduction based on their feature parameters.

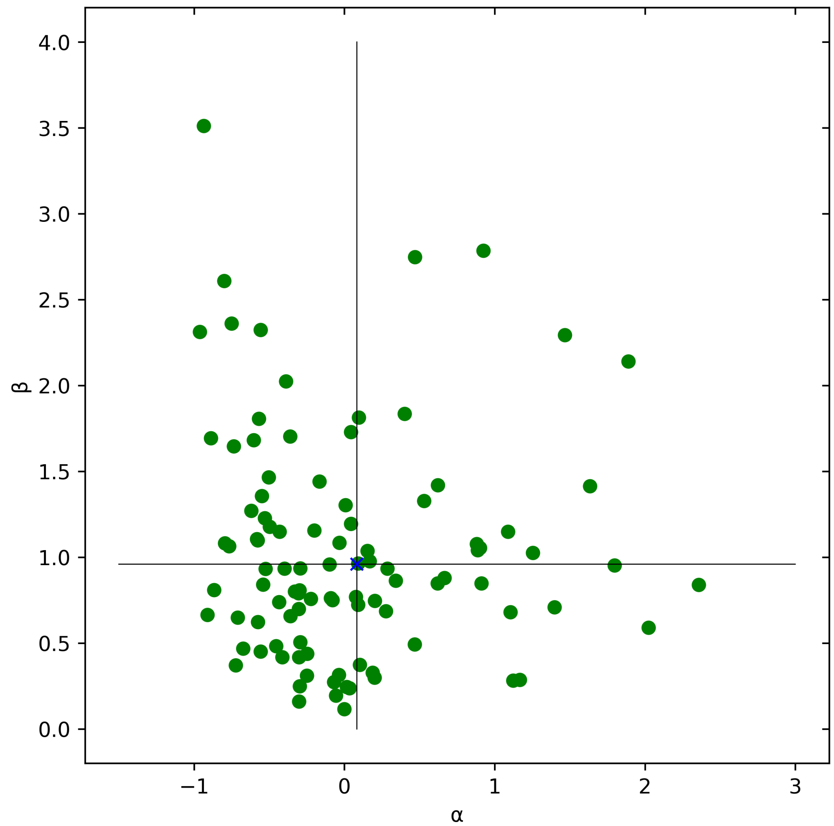


Figure 4.13: Plot of $\beta$ versus $\alpha$ parameter for 99 customers.

Based on the plot in Fig. 4.13, a rule based selection is proposed for customer targeting based on the quadrant location of the customer and distance from the quadrant origin (i.e. $\alpha$ and $\beta$ dimension for customer aggregate). The distance between each customers to the quadrant origin is computed according to (4.15) and ranked in descending order,

$$featuresum = \alpha + \beta. \tag{4.15}$$

Customers are first taken from quadrant I to II, IV and lastly III. II is ordered prior to

IV as preference is given to potential contributing to the flattening of the customer profile ahead of flattening within DR event window. In the case of within quadrants, customers are selected from the highest $featuresum$ value to the lowest.

Fig. 4.14 shows the normalised demand profile of a customer in each quadrant and how they compare to the cumulative demand profile.

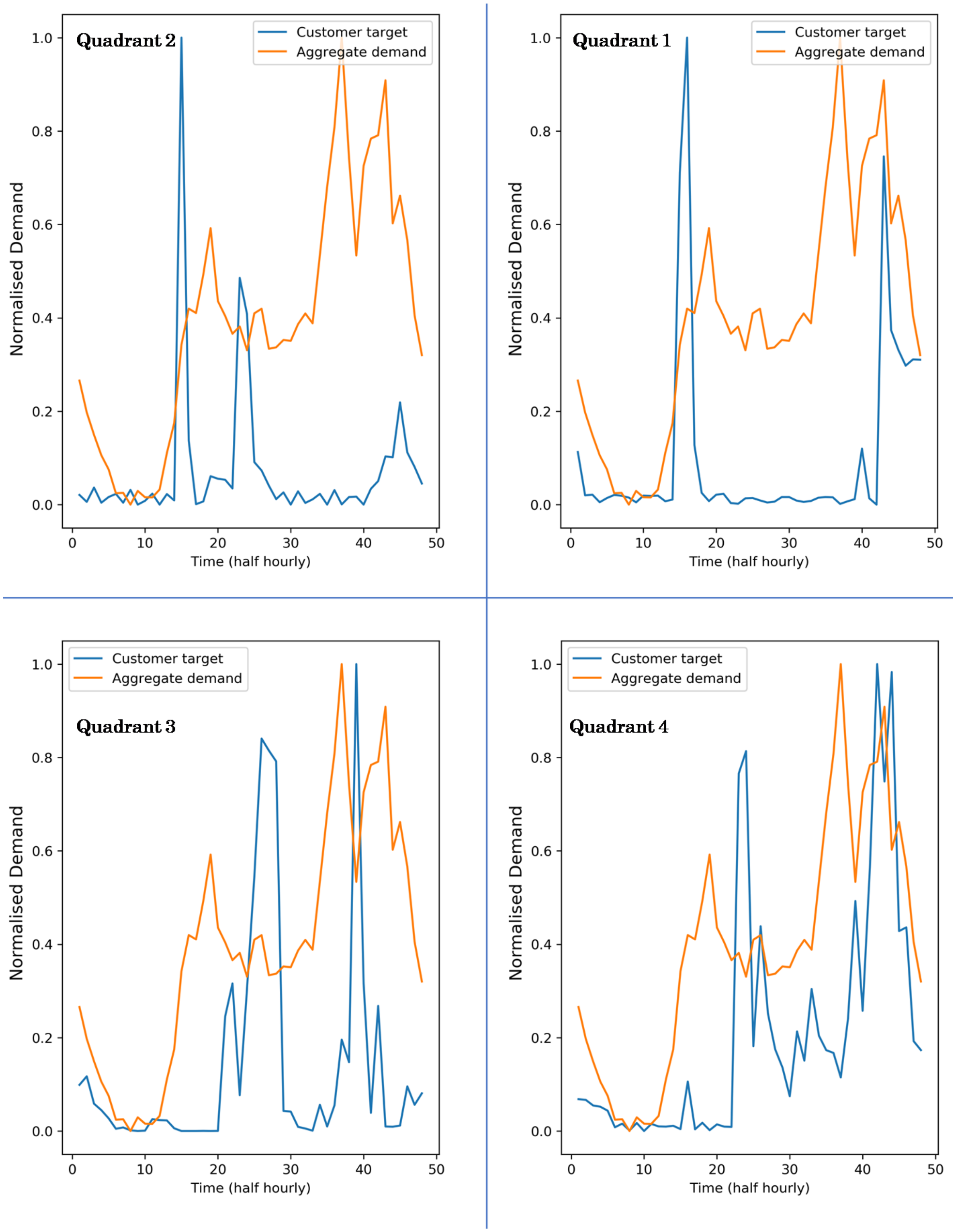


Figure 4.14: Plot of customers normalised profile with normalised aggregate profile for each quadrant.

## 4.4 Conclusion

In this chapter, a novel characterization method for customers' smart meter data was proposed for peak load targeting. Customer's daily demand profile is normalised to present the per-unit variation of their demand. A representative partitional clustering algorithm (*k*-medoid with DTW distance measure) is proposed to group customers with respect to their daily demand variability. DTW distance measure was utilized to measure the similarity between the characterized smart meter data. The results showed a distinct grouping of customers into clusters for which they can be targeted for peak load reduction.

In addition to peak demand targeting, a novel methodology for targeting customers for local peak demand reduction is also proposed. Two parameters were proposed and designed to rank customers based on their potential to contribute to DR. Scope for future work is assessing the performance of the proposed targeting method with traditional methods using real customer DR recruitment data from utilities (traditional targeting methods include recruiting customers based on factors like monthly bill and aggregate demand).

# Chapter 5

# Customer Baseline Load Estimation for Incentive-Based Demand Response Using Long Short-Term Memory Recurrent Neural Network

The transition to an intelligent, reliable and efficient smart grid with a high penetration of renewable energy drives the need to maximise the utilisation of customers DR potential. The availability of smart meter data means this potential can be more accurately estimated and suitable DR programs can be targeted to customers for load shifting, clipping and filling. To quantify customers contribution to demand reduction during DR events, the need to accurately estimate a customer's demand baseline is very important. This importance stems from accurately and fairly allocating DR benefits in form of compensation to customers, to understanding the DR resources available to utilities for effective allocation of grid resources. Methods for estimating customer baseline can be similar in concept to those of load forecasting, however, the time frame for baseline estimation spans from a past to present time frame. Load forecasting strictly focuses on future time frame. The main purpose for customer baseline estimation is to quantify what a customer would have demanded should they not have participated in DR as opposed to estimating what their future demands might be in the case of load forecasting.

In this chapter, focus is made on estimating customer demand baseline for incentive-based DR. A long short-term memory recurrent neural network methodology is proposed for customer baseline estimation using previous like days data as during DR events period for model training. The proposed method is tested on the publicly available Irish smart meter data and results shows a significant increase in baseline estimation accuracy when compared to traditional baseline estimation methods.

## 5.1 Introduction

The need to reduce the carbon footprint all through the electricity supply chain and the ever increasing integration of renewable energy sources makes demand response (DR) key to ensuring a stable, reliable and high quality power supply. The non-dispatchable characteristics of renewable energy makes DR important as flexibility in customers demand can be exploited to follow available renewable energy resources. DR refers to a customer side effort to reduce demand in order to follow available supply which could be limited as a result of energy resources or transmission and distribution capacity constraints. There are two main categories of DR programs which are price-based DR and incentive-based DR [5] [7]. Price-based DR involves using time-varying electricity prices to change customers demand characteristics while incentive-based DR involves the utility offering customers incentives to reduce demand during a DR event. Depending on a customers demand characteristics, various DR programs can achieve different amount of demand reduction during DR events.

With the rising popularity of smart meters, smart meter data showing customers consumption can be utilised to derive intelligence for the optimal deployment of DR programs to customers. Some recent research on DR has focused on deriving intelligence from customer smart meter data for enhancing DR program implementation in smart grids [124] [143] [144]. Considering the two main load classes which are industrial and residential loads, price based DR has seen more application for residential customers while incentive-based DR is much popular with industrial customers. One of the main reasons for this is the relative ease of estimating an industrial customers demand baseline compared to the highly variable nature of residential customers demand characteristics. The availability of smart meters at residential customers' location means the data can be explored to estimate a more representative baseline estimate of their demand during DR events, thereby enhancing their suitability for incentive-based DR. The main importance of baseline estimation is that customers demand reduction during DR events can be accurately estimated especially for accurate incentive payment and accurate knowledge of the aggregate DR potential of users i.e., showing the indication of how much DR resource is for a particular event. One of the key benefits smart meter data can provide for baseline estimation is the means to reduce opportunities for baseline manipulation. Baseline estimation models can harness historical granular customer demand data to estimate customer baselines thereby making manipulation (e.g. increasing demand levels a few days prior to DR event day) uneconomical.

Baseline estimation methods proposed in literature can be classified under averaging, regression, statistical and machine learning approaches. Examples of averaging approaches include direct average of X previous days; average of the highest X of Y days (HighXofY); average of the middle X of Y days (MidXofY); and average of the lowest X of Y days (LowXofY) [145]. These methods are widely used in the industry with Mid4of6 and High4of5 used in PJM interconnection [146]. A support vector regression (SVR) method was proposed in [147] for baseline estimation. A clustering approach was implemented in [148] with self organizing map and *k*-means methodology proposed for estimating customer baseline. In [149], a density based clustering method together with k-means partitional clustering method was proposed to find representative baseline for customers. A statistical approach was proposed in [150] using a customer control group selection algorithm for estimating customer demand baseline. In [151], a probabilistic baseline estimation method was proposed for estimating customers consumption baseline using Gaussian process.

In this chapter, a machine learning approach using the long short-term memory recurrent neural network (LSTM RNN) methodology for customer baseline estimation during DR event period is proposed. Historic demand data within the DR event time span is used to train the LSTM model and proposed estimation takes in data of a defined number of previous like-days for estimating customers baseline for the DR event period. The motivation for this proposed methodology is the improvement in forecasting results obtained using the LSTM RNN methodology for demand forecasting [152] [153] as well as battery state of charge estimation [154] compared to traditional methods.

The contributions of this chapter are as follows:

- The proposition and implementation of the LSTM RNN methodology for event based customer baseline estimation.
- The justification of mean percentage error (MPE) as a means to measure baseline estimation accuracy compared to mean absolute percentage error (MAPE) and root mean square error (RMSE).
- The comparison of the proposed customer baseline estimation results with the Low4of5, Mid4of6 and High4of5 methods.

## 5.2 Recurrent Neural Network

Among the main classes of artificial neural networks, RNNs has proven to be best suited for sequence learning such as time series forecasting problems [155]. Compared to feed-forward neural networks (FNNs), where signals travel in one direction from the input to the output via the hidden layer, RNNs allow signals to go back and forth hence allowing information in past data to be exploited for future data estimation. Fig. 5.1 shows the architecture of a RNN compared to a FNN and shows the differences in the directional flow of the data.

Given a time series input $x = \{x_1, x_2, \cdots, x_t\}$, the hidden state and output sequence is derived using 5.1 and 5.2 respectively,

$$h_t = f(W_{hx}x_t + W_{hh}h_{t-1} + b_h), \tag{5.1}$$

$$y_t = g(W_{yh}h_t + b_y), \tag{5.2}$$

where $W_{hx}$, $W_{hh}$ and $W_{yx}$ are the input-hidden, hidden-hidden and hidden-output weight matrix, respectively. $b_h$ and $b_y$ represent the bias of the hidden and output layer, respectively. The activation layer of the input and output layer is represented as $f(\cdot)$ and $g(\cdot)$.

### 5.2.1 Long Short-Term Memory

One of the main challenges with RNN is the exploding and vanishing gradient problem [156]. This challenge occurs as a result of the loss function decaying exponentially with time. To address this challenge, LSTMs include a memory cell and gates at the hidden layers. Fig. 5.2 shows the architecture of an LSTM RNN.

The LSTM block consists of an input gate ($i$), output gate ($o$), forget gate ($f$) and attached memory cells ($s$) [152]. The input $x_t$ and previous output $h_{t-1}$ to the LSTM block determine the decisions of the input, output and forget gates whether to switch on or off [153]. The input gate controls what to keep in the internal state $s_t$, while the output gate decides the part of the internal state $s_t$ to pass to the output $h_t$. The forget gate however decides what part of the previous state $s_{t-1}$ needs to be forgotten. The equations (5.3), (5.4), (5.5), (5.6), (5.7), (5.8) and (5.9) below presents the operation of the gates as well as the states and output formulation.

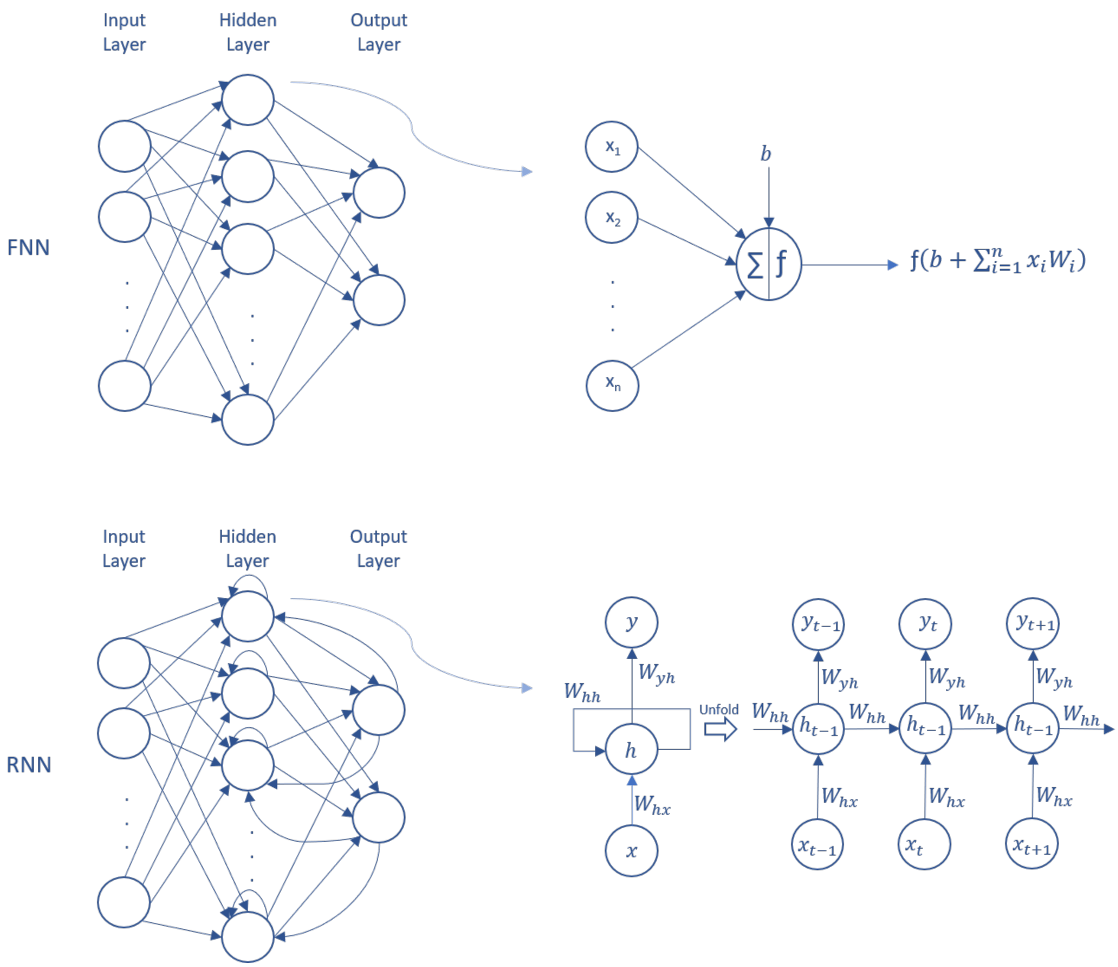


Figure 5.1: Comparison of the RNN and FNN architecture.

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fs}s_{t-1} + b_f), \tag{5.3}$$

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + W_{is}s_{t-1} + b_i), \tag{5.4}$$

$$u_t = g(W_{sx}x_t + W_{sh}h_{t-1} + b_s), \tag{5.5}$$

$$s_t = u_t i_t + s_{t-1} f_t, \tag{5.6}$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{os}s_{t-1} + b_o), \tag{5.7}$$

$$h_t = o_t \ell(s_t), \tag{5.8}$$

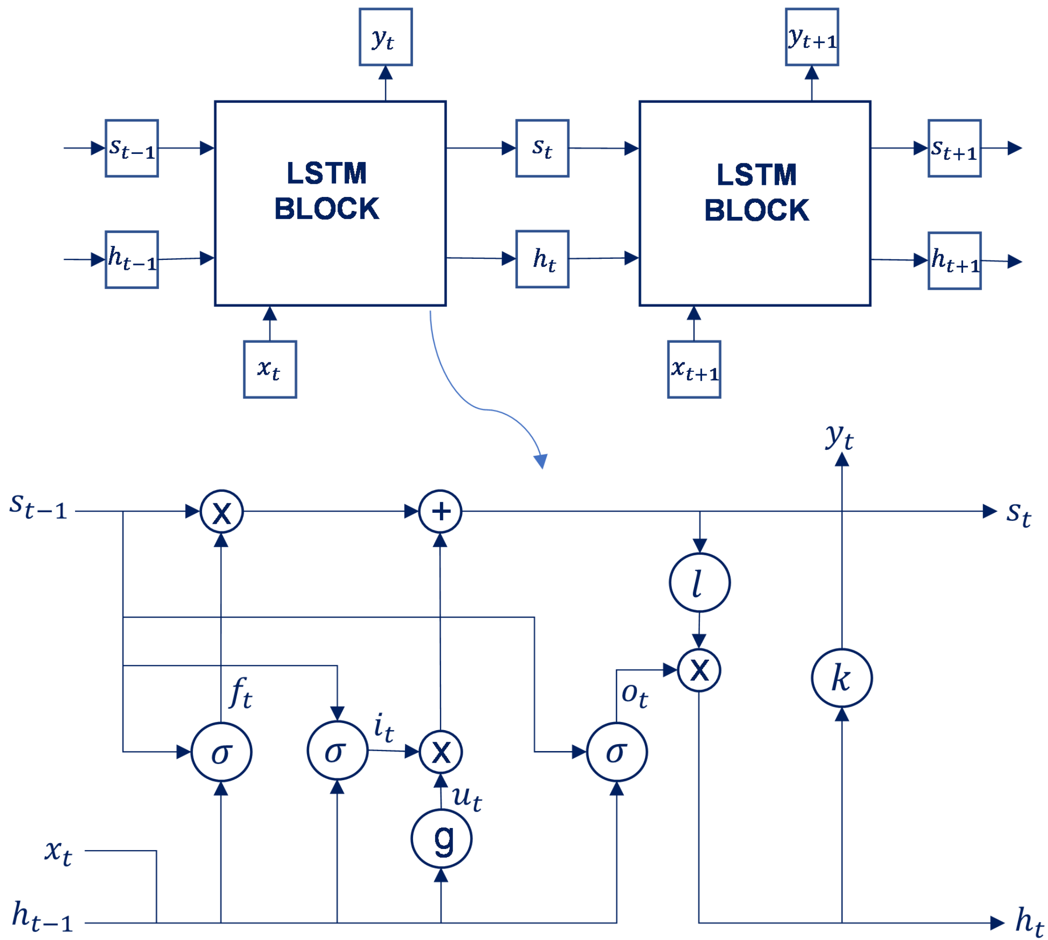


Figure 5.2: LSTM RNN architecture.

$$y_t = k(W_{yh}h_t + b_y), \tag{5.9}$$

where $W_{fx}$, $W_{ix}$, $W_{fx}$, $W_{sx}$, $W_{ox}$; $W_{fh}$, $W_{ih}$, $W_{sh}$, $W_{oh}$; $W_{yh}$; $W_{fs}$, $W_{ic}$, $W_{os}$; $b_f$, $b_i$, $b_s$, $b_o$, $b_y$ are the input weight matrices; recurrent weight matrices; hidden output weight matrix; weight matrices of peephole connections and bias vectors, respectively.

## 5.3 Customer Baseline Estimation

An event based customer baseline estimation method is proposed for incentive-based DR. We propose the LSTM RNN technique for estimating the baseline demand during event period. Fig. 5.3 shows the framework of our proposed methodology. We characterize users demand for the event period of $y$ previous non DR like-days[1] by normalising it using the MinMaxScaler presented in 5.10 below,

$$d_t^* = \frac{d_t - d_{min}}{d_{max} - d_{min}} * (max - min) + min, \tag{5.10}$$

where $d_t^*$ is the normalised data, $d_t$ is the demand at time $t$, $d_{min}$ and $d_{max}$ is the minimum and maximum demand for the event period, respectively, and $min$ and $max$ represents the range of the feature. The derived feature is then fed into the LSTM network as input matrix $X$. The input feature range of -1 and 1 is proposed for $min$ and $max$, respectively, as it enhances an effective measure for backpropagation as proposed in [157].

The proposed DR event estimation model is presented in the subsequent subsections.

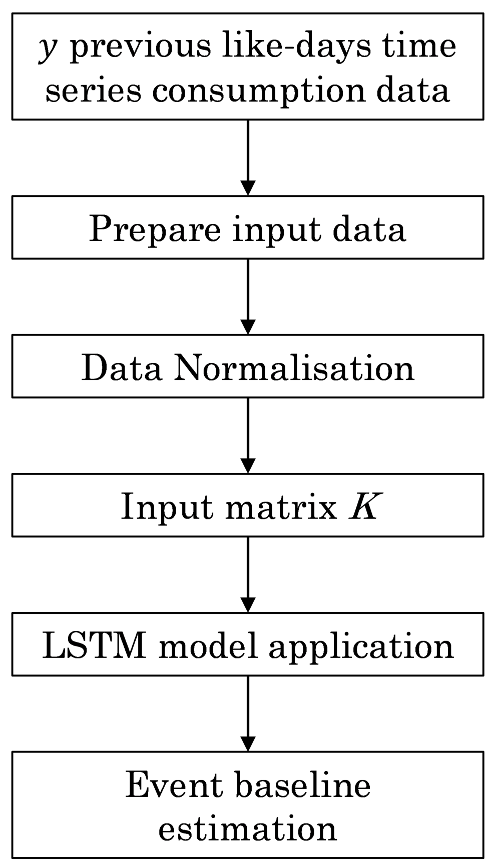


Figure 5.3: Framework for customer DR baseline estimation.

[1]like-days mean previous week days or weekend days

Prior to the data preparation stage, like-days can alternatively be selected based on characteristics like temperature and weather. This selection approach is based on the premise that days with similar temperature or weather have similar demand characteristics. Given the distinct difference in demand profile between weekday and weekend; coupled with the high level of similarities within like-days as presented in [158], like-days by day type is chosen for this analysis.

### 5.3.1 Proposed Model

Firstly, the input data is prepared, normalised and fed into the trained LSTM model for baseline estimation. Let $C$ represent a set of $k$ customers as shown in (5.11),

$$C = \{c_1, c_2, c_3, \cdots, c_k\}, \tag{5.11}$$

with each enrolled for incentive-based DR.

The aim is to estimate the demand baseline for a customer $c_n$ during a DR event period between $t_s$ and $t_e$. Let $D_{a,b}$ as shown in (5.12),

$$D_{a,b} = \{d_{0,1}, d_{0,2}, \cdots, d_{0,48}, \cdots, d_{y,1}, d_{y,2}, \cdots, d_{y,48}\} \tag{5.12}$$

represent the demand series for customer $c_n$ where $a$ denotes the days from event day 0 to previous like-day $y$ as shown in (5.13) and $b$ denotes the daily half-hourly demand data as shown in (5.14) below:

$$a = \{0, 1, 2, \cdots, y\}, \tag{5.13}$$

$$b = \{0, 1, 2, \cdots, 48\}. \tag{5.14}$$

Given a DR day with reduction event between $t_s$ and $t_e$ where $t_s$ represents the event start time and $t_e$ represents the event end time for a day's demand. The objective is to estimate the customer baseline demand between times $t_s$ and $t_e$ where $t_s, t_e = \{1, 2, 3, \cdots, 48\}$ and $t_e > t_s$. The input data to the model is given as (5.15),

$$X = D^*_{a,b} = \{d_{y,t_s}, d_{y,t_{s+1}}, \cdots, d_{y,t_e}, d_{y-1,t_s}, d_{y-1,t_{s+1}}, \cdots, d_{y-1,t_e},$$
$$\cdots, d_{2,t_s}, d_{2,t_{s+1}}, \cdots, d_{2,t_e}, \cdots, d_{1,t_s}, d_{1,t_{s+1}}, \cdots, d_{1,t_e}\}, \tag{5.15}$$

while the output, that is, the estimated baseline is given as (5.16) below

$$Q(t) = \{q_{t_s}, q_{t_{s+1}}, q_{t_{s+2}}, \cdots, q_{t_e}\}. \tag{5.16}$$

### 5.3.2 Performance Metrics

In order to measure the performance of our proposed methodology for customer baseline estimation, we propose using the mean percentage error (MPE), mean average percentage error (MAPE) and root mean square error (RMSE) to measure the closeness of the estimate to the true customer demand for an event period. The MPE is defined as the average of the percentage error between the baseline estimate and the true customer demand during the event period as shown in 5.17,

$$MPE = \frac{100}{y} \sum_{t=t_s}^{t_e} \frac{d_t - q_t}{d_t}, \tag{5.17}$$

where $d_t$ and $q_t$ is the true baseline and estimate baseline, respectively.

MAPE however measures performance based on the absolute value of the error difference between the estimated and true baseline. MAPE is presented in 5.18:

$$MAPE = \frac{100}{y} \sum_{t=t_s}^{t_e} |\frac{d_t - q_t}{d_t}|. \tag{5.18}$$

RMSE is the mean of the sum of the square error where the error is the difference between the estimate and the true baseline. RMSE is presented in 5.19 below:

$$RMSE = \sqrt{\frac{\sum_{t=t_s}^{t_e}(d_t - q_t)^2}{y}}. \tag{5.19}$$

## 5.4 Result and Discussion

For our analysis, customer smart meter data from the publicly available Irish Commission for Energy Regulation (CER) is used [139]. Fig. 5.4 shows the demand profile of one of the customers which we use as a case study.

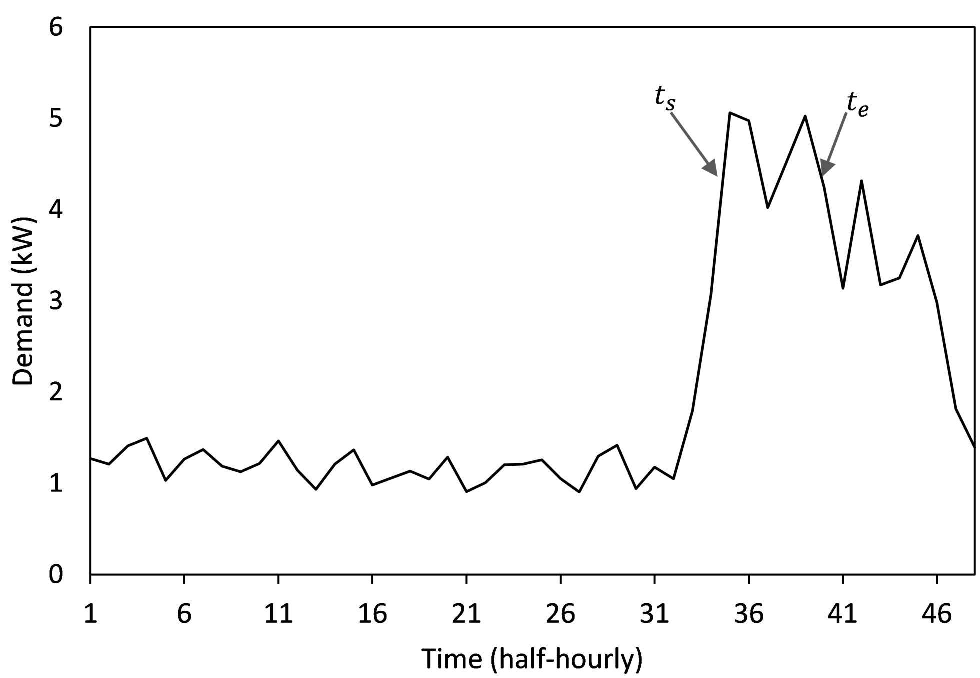


Figure 5.4: Customer demand profile.

Given a DR event period between time period 34 and 40, we estimate the demand baseline for this period. We implement the proposed baseline estimation methodology using 10 previous like-days i.e., $y = 10$. Firstly, we run experiments on our model by varying the LSTM hyper-parameters which are the number of neurons and epochs. Table

Table 5.1: DR estimation model error summary with varying epochs.

| Epoch numbers | MPE(%) | MAPE(%) | RMSE |
|---|---|---|---|
| 300 | 28.95 | 45.44 | 2.08 |
| 600 | 22.07 | 33.59 | 1.54 |
| 900 | 13.93 | 41.44 | 1.95 |
| 1200 | 7.36 | 34.80 | 1.78 |
| 1500 | 5.32 | 28.83 | 1.44 |

Table 5.2: DR estimation model error summary with varying neuron numbers.

| Neuron numbers | MPE(%) | MAPE(%) | RMSE |
|---|---|---|---|
| 100 | 28.22 | 41.10 | 1.93 |
| 200 | 21.06 | 46.17 | 2.09 |
| 300 | 27.54 | 40.10 | 1.92 |
| 400 | 7.97 | 26.11 | 1.26 |
| 500 | 5.32 | 28.83 | 1.44 |

5.1 presents the MPE, MAPE and RMSE of the LSTM model with varying number of epochs and neuron numbers fixed at 500.

Increasing number of epochs translated to decreasing MPE value indicating the closeness of the aggregate demand of the LSTM estimate to the true baseline. However, MAPE and RMSE did not follow the trend when epoch number was increased from 600 to 900. MAPE and RMSE use absolute values in its error calculation and this does not take into consideration the bias between both the estimated demand and true demand per unit time. In baseline estimation, the important indication for accuracy is the closeness of the aggregate demand during event period to the aggregate true demand given a non DR event. This situation makes MPE the ideal factor to measure the performance of a customer baseline estimation for incentive-based DR. We also vary the neuron numbers with epoch number fixed at 1500. Table 5.2 presents the MPE, MAPE and RMSE of the LSTM model with varying number of neurons.

The error results presented shows a decreasing MPE for increasing number of neurons. This trend shows a closer cumulative demand of the baseline estimate to the true baseline as the number of neuron increases. Also, the consequence of using more neurons is an increase in both computational time and resources. The MAPE and RMSE also in this case, does not follow a decreasing trend with increasing number of neurons as bias is not considered in the error calculation. Fig. 5.5 shows the baseline estimation result of the LSTM model using different neuron numbers during event time compared to the true baseline.

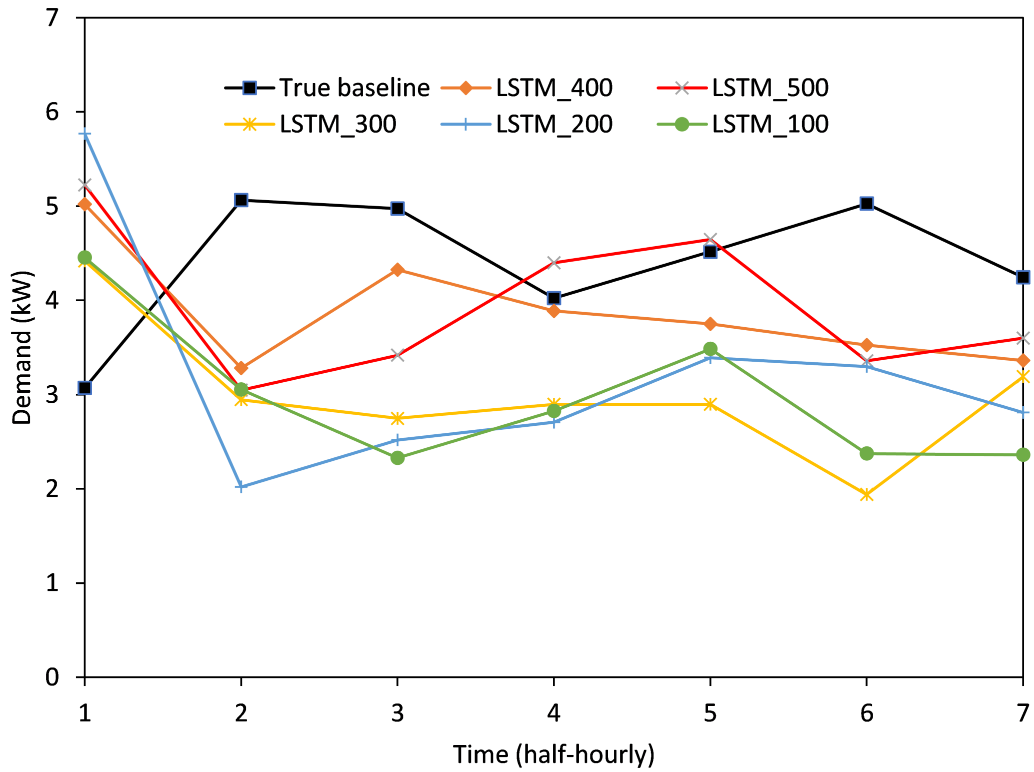


Figure 5.5: Baseline estimation with varying neuron numbers.

Based on the results of the varying hyper-parameter for the proposed LSTM model, we select neuron number as 500 and epoch number as 1500 and compare the estimation results with traditional baseline methodology i.e., Low4of5, Mid4of6 and High4of5. Fig. 5.6 presents the result of the LSTM estimation with the baseline estimate in red and input data to model in blue.

The proposed LSTM baseline estimation model result compared to traditional methods and the true baseline is presented in Fig. 5.7.

Table 5.3 presents the MPE, MAPE and RMSE of the LSTM methodology compared to the Low4of5, Mid4of6 and High4of5 baseline estimation methods. Both the MAPE

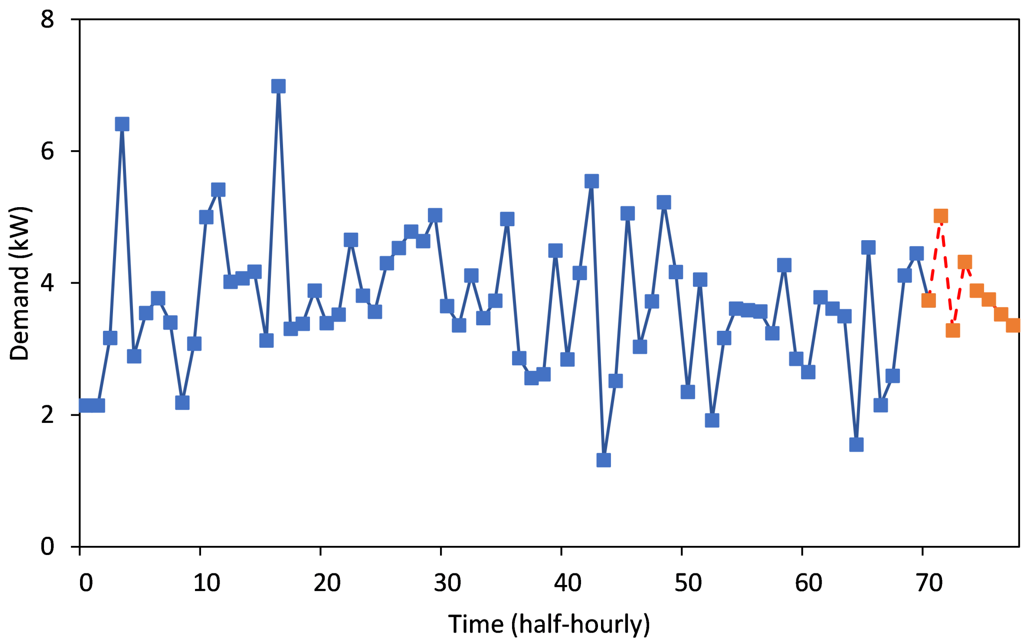


Figure 5.6: Baseline estimation showing input data and output of the LSTM model.

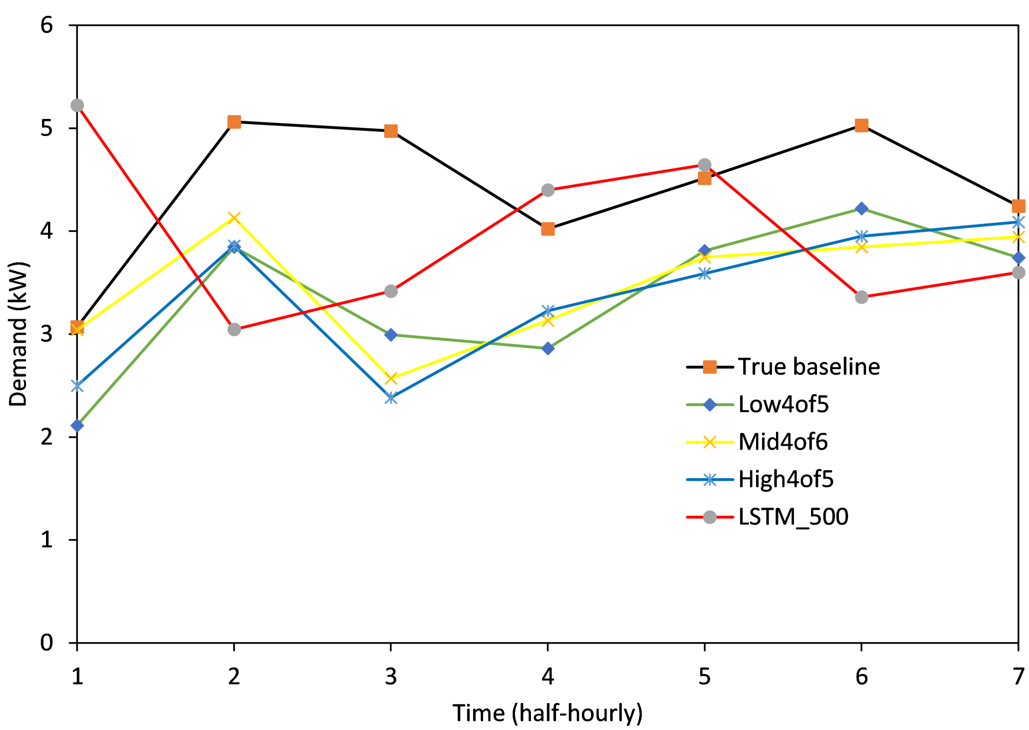


Figure 5.7: Baseline estimation using LSTM compared to traditional methods.

and RMSE shows a higher error value for the LSTM method compared to the Low4of5, Mid4of6 and High4of5.

Table 5.3: DR estimation model error summary.

| Estimation model | MPE(%) | MAPE(%) | RMSE |
|---|---|---|---|
| LSTM Model | 5.32 | 28.83 | 1.44 |
| Low4of5 | 23.91 | 23.91 | 1.14 |
| Mid4of6 | 19.63 | 19.63 | 1.17 |
| High4of5 | 22.84 | 22.84 | 1.26 |

These error values is however different from the MPE which shows the LSTM method having a significantly much lower value i.e., 5.32% compared to 23.91%, 19.63% and 22.84% for Low4of5, Mid4of6 and High4of5, respectively. The MPE gives an indication of how close the estimate of the cumulative demand during DR event is to the true baseline as can be shown in Fig. 5.8.

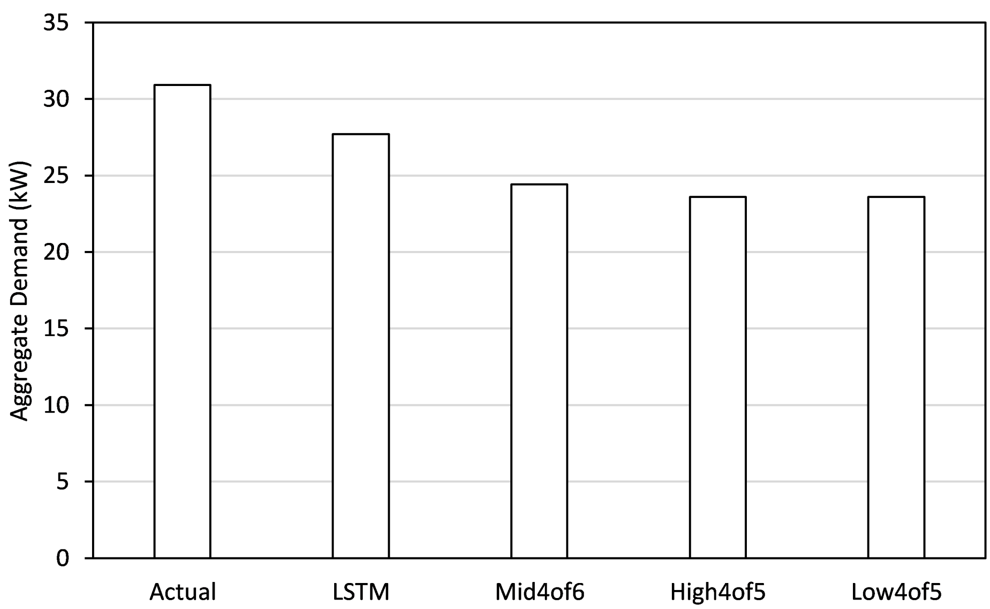


Figure 5.8: Comparison of aggregate demand estimates using LSTM to traditional methods.

## 5.5 Conclusion

In this chapter, a novel customer baseline estimation methodology is proposed. The LSTM RNN method is proposed for estimating customer demand specifically during DR event.

To minimise the computational expense, the model is trained on the event period only compared to the whole daily period as used in various methods proposed in literature. The proposed method proved to be more accurate as the cumulative demand of the estimation was much closer to the true baseline compared to traditional methods. One scope for future work is to assess the performance of the proposed model in mitigating against baseline manipulation. The proposed future work will also include an assessment of other baseline methods performance with regards to how they mitigate against baseline manipulation.

# Chapter 6

# Control Based Residential Customer Baseline Estimation: Taking into Consideration the Demand Response Rebound Effect

This chapter proposes a novel customer baseline estimation methodology that takes into consideration the demand rebound effect. In the earlier work done in chapter 5, an LSTM model was proposed for customer baseline estimation based on a historic like-days data. One key element of DR is the demand rebound effect where demand reduction during a DR event is shifted to other time period. The need to better estimate customer baseline makes capturing the demand rebound in the estimation methodology vital. To address this, a control based DR estimation methodology is proposed. The proposed methodology is applied to estimate the baseline of residential customers given the challenge they pose due to their high variability with results compared to earlier methods proposed in literature. The model is like-wise applicable to industrial and commercial customers and differences between various methods are minimal given the relatively low variation in these load types.

## 6.1 Introduction

Customer demand baseline refers to the amount of electricity a customer would have consumed if they had not participated in DR. An accurate estimation of customers demand baseline is very key as this gives the basis for a fair financial compensation to customers especially for incentive-based DR. Aside accurately estimating rewards for participating customers, customer baseline estimation can help operators track the level of DR resources available as well as have a true picture of their responsiveness from time to time. In estimating a customers demand baseline, it is important to take into consideration the DR's effect on customers demand profile.

The effect of demand shift from DR event period to other period of a customers demand profile can result into what is known as demand rebound. The amount of demand rebound exhibited by a particular customer depends on 2 main factors.

- The amount of demand reduction during DR event.
- The proportion of the reduced demand shifted compared to demand clipped.

The increasing availability of smart meters gives an opportunity for enhanced and improved baseline estimation. This chapter presents a novel method of estimating customer baseline using the demand profile of similar non DR participating customers. The method not only estimates customer baseline but also estimates demand rebound, amount of demand shifted as well as amount of demand curtailed.

## 6.2 Customer Baseline Load Estimation Method Classification

The proposed methodologies in literature for customer baseline estimation is classed under two main groups based on the source of the demand estimation data which are non-control group methods and control group methods. Non-control group methods employs using the approach where the baseline is determined using the DR customer's historic demand data and other external influencing factors like weather, type of day and special events day (holiday, national sports event and so on). Control group methods differ in that the DR customer's baseline is determined using the demand data of similar non DR participating customers. Fig. 6.1 shows the group and subgroups of existing methodologies for customer baseline estimation in literature.

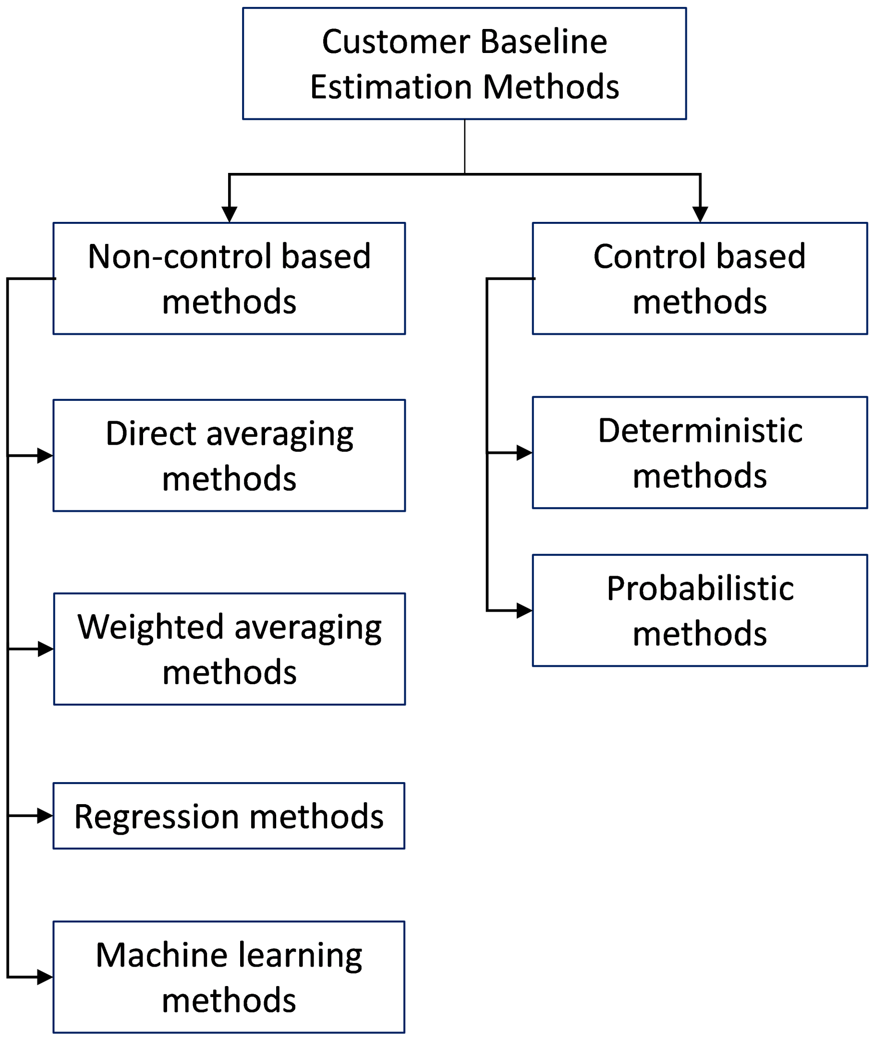


Figure 6.1: Methodologies for customer baseline estimation.

The customer baseline subgroups methods are discussed in the subsections below.

### 6.2.1 Non-Control Group Methods

Non-control group methods estimates DR customers baseline using data from their previous days demand and can include coupling with external factors like weather, type of day (weekday or weekend) and historical trends. Based on methods proposed in literature, non-control groups include direct averaging techniques, weighted averaging techniques, regression and machine learning methods.

#### Direct Averaging Methods

Direct averaging methods are based on finding the direct average of previous non DR days of a participating customers as the customers baseline demand. Examples of direct averaging methods include previous *n* like day average, LowXofY, MidXofY and HighXofY.

#### Weighted Averaging Methods

Weighted averaging techniques involves assigning a weighting factors to previous non DR like days before finding the average. The rationale of this method is that nearest non-DR days to DR days have a higher weighting factor with further days having a lower weighting factor.

#### Regression Methods

Regression methods makes use of influencing factors like weather. In this case questions like how does demand vary with the temperature, humidity and season is answered coupled with historic customers demand. Based on this, the baseline can be estimated using the derived regression equation showing the relationship between variables like temperature, humidity and customer demand.

#### Machine Learning Methods

Some machine learning models has been proposed for baseline estimation. In [126], a LSTM RNN technique was proposed for estimating residential customer's baseline demand using historical non-DR data. Another application of machine learning was in [159] where Neural nets was employed for baseline estimation. A baseline model utilising self organizing map and *k*-means clustering was proposed in [148].

### 6.2.2 Control Group Methods

Control group customer baseline load (CBL) estimation methodologies employs the use of non participating DR customers to estimate the baseline of participating DR customers. Clustering methods are employed to group DR customers into groups based on the similarity of their consumption profiles. The CBL is estimated using non-DR customers data present in the DR customers cluster. Proposed control group methods can be further grouped into deterministic and probabilistic methods.

#### Deterministic Method

Deterministic control group methods employ techniques to directly estimate a baseline for the customer using non-DR customers consumption data showing similar characteristics. Direct clustering approach was used in [160] and [149] with the baseline estimation derived from non DR members of clusters where DR customers belongs.

#### Probabilistic Method

For probabilistic methods, aside deriving the baseline from non participating DR customers, an element of uncertainty is added to the estimated baseline. A probabilistic clustering based residential baseline estimations methodology was proposed in [161]. Probabilistic baseline estimation was also proposed in [162] for baseline estimation.

## 6.3 Need to Consider the Demand Rebound Effect

During a DR event, user's demand reduction can either be in form of a curtailment or load shift. Fig. 6.2 shows the rebound effect of DR.

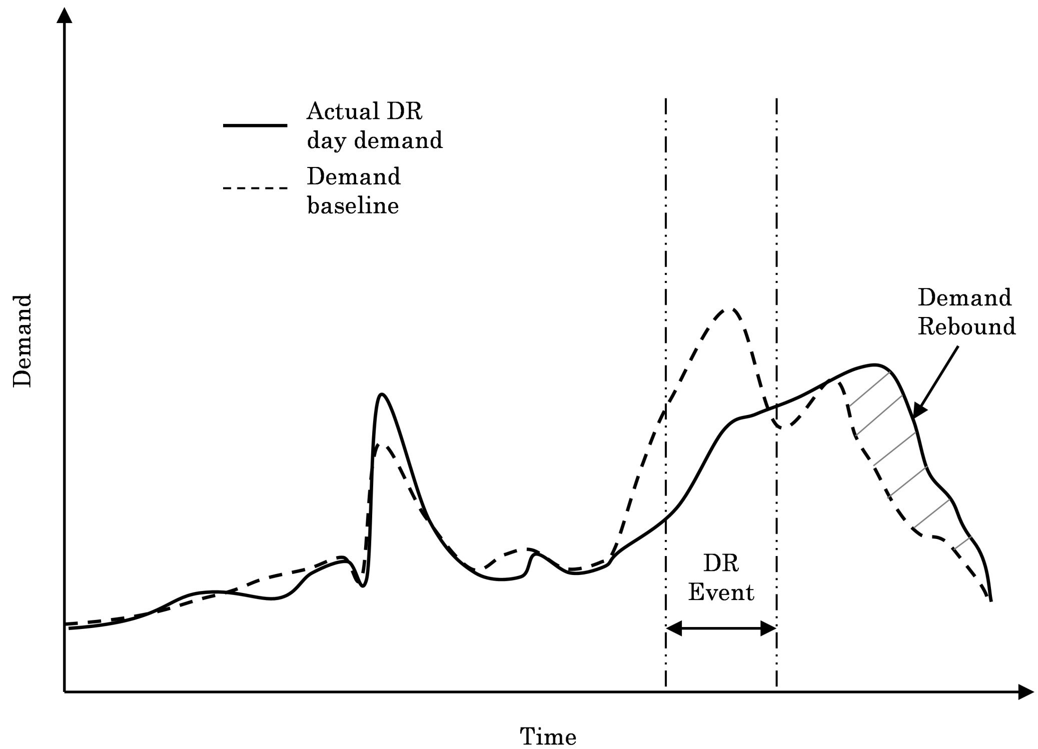


Figure 6.2: Customer Rebound Effect.

One of the key benefit of considering the demand rebound effect is how it helps to further improve the accuracy of baseline estimation. Having an estimate of the demand rebound can account for demand shift outwith the DR event period and hence improve the estimation of the customer's baseline. The demand rebound estimate can be utilised as a constraint feature for control based methods as non DR customers can be filtered depending on if the demand reduction estimate captures both demand curtailment and shift.

## 6.4 Proposed Control Based Method with DR Rebound

A control based baseline estimation methodology that takes into consideration the demand rebound effect is proposed. Fig. 6.3 presents the framework of the proposed estimation method. The key steps presented in implementing the proposed methodology are discussed in the subsections below.

### 6.4.1 *k*-Medoid Partitional Clustering with LB_Keogh Distance Measure

The *k*-medoid partitional clustering using the LB_Keogh (lower bound) distance measure is proposed for clustering the customers demand profile. Compared to the clustering methodology proposed in chapter 4, the LB_Keogh distance measure approach has performance benefits over the standard DTW as demonstrated in [163]. The more the time series required to be clustered the more the computational expense and the LB_Keogh distance measure performs well in this situation as shown in previous time series clustering studies [164].

### 6.4.2 LB_Keogh Distance Measure

In comparison to the standard DTW distance measure earlier used for *k*-medoid clustering in chapter 4, LB_Keogh takes a less computational expensive approach in calculating distances between time series. The LB_Keogh is a lower bounding measure proposed by Keogh *et al.* [163] that minimises the amount of points in the time series for which distances is computed. The process for computing the LB_Keogh distance measure is highlighted below.

1. Given two time series $Q$ and $C$, an upper and lower bound sequence is derived for $Q$ to envelop the time series as shown in Fig. 6.4. Where $w$ is defined as the envelope

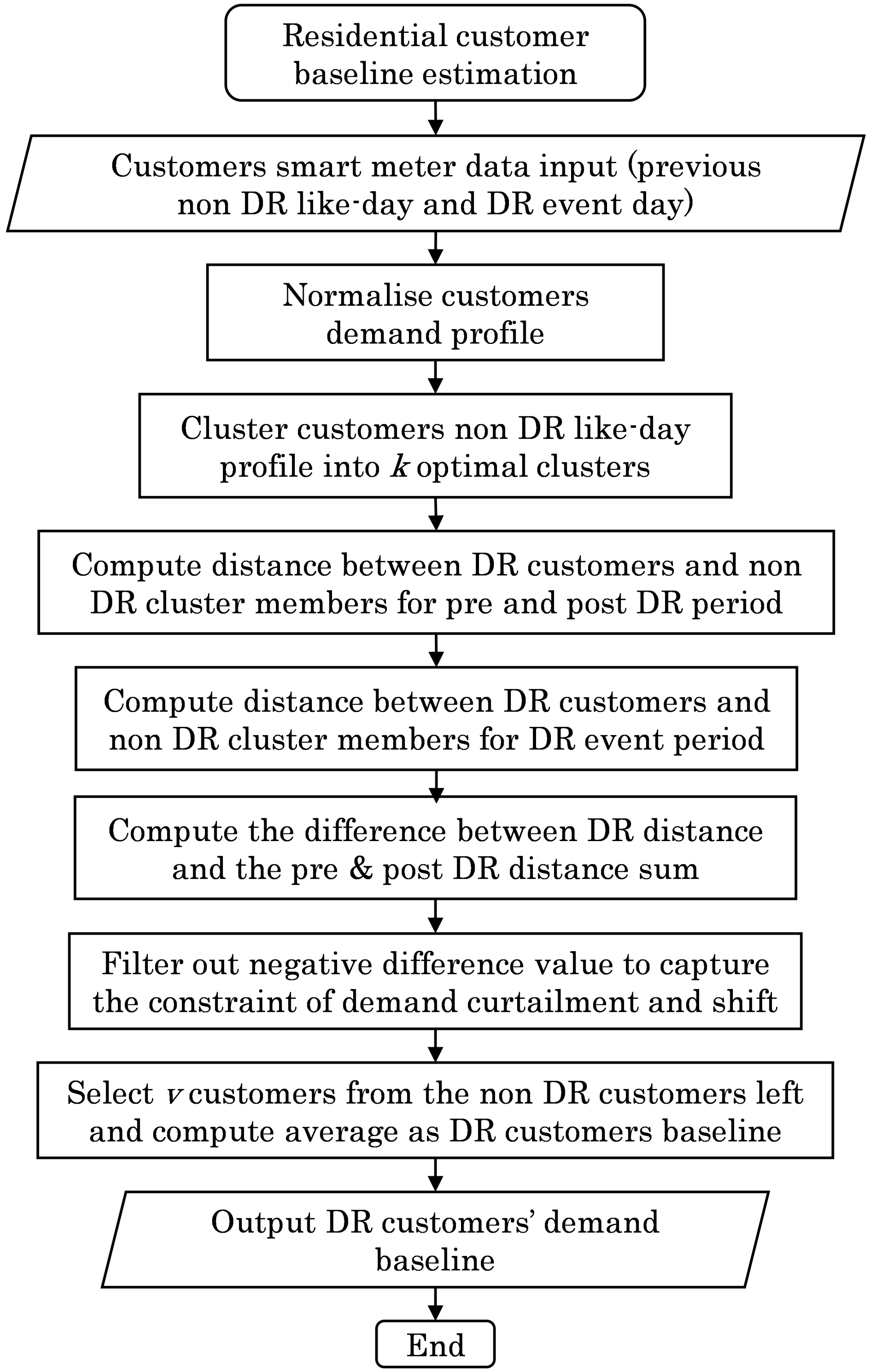


Figure 6.3: Proposed Methodology.

reach, $U$ and $L$ is derived according to (6.1) and (6.2) below:

$$U_i = max(q_{i-w} : q_{i+w}), \tag{6.1}$$

$$L_i = min(q_{i-w} : q_{i+w}). \tag{6.2}$$

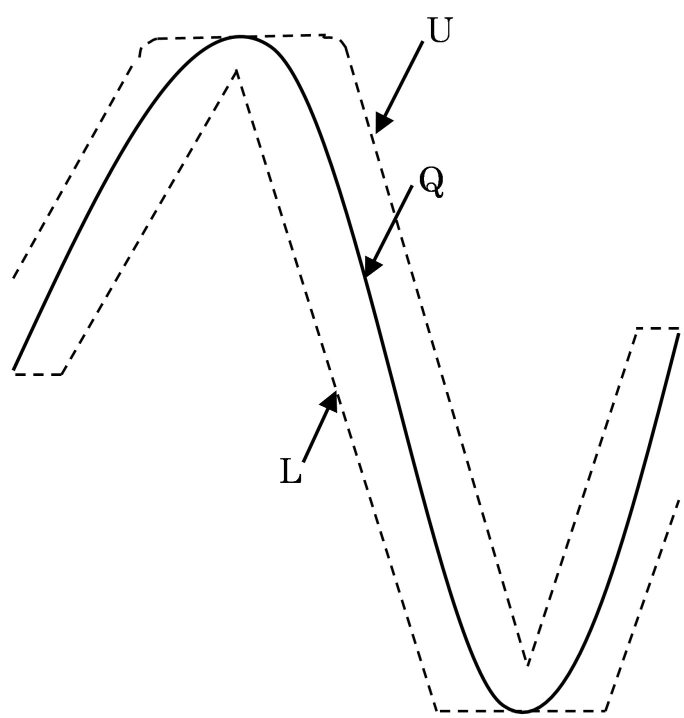


Figure 6.4: Illustration of upper and lower bound sequence envelope for time series $Q$.

2. The LB_Keogh is derived using (6.3) below:

$$\text{LB\_Keogh(Q,C)} = \sqrt{\sum_{i=1}^{n} \begin{cases} (c_i - \text{U}_i)^2 \text{ if } c_i > \text{U}_i \\ (c_i - \text{L}_i)^2 \text{ if } c_i < \text{L}_i \\ 0 \text{ otherwise} \end{cases}} . \tag{6.3}$$

As illustrated in Fig. 6.5, areas where C is within the upper and lower bounds $U$ and $L$ have distance set as 0, while distance is computed for areas outside the envelope.

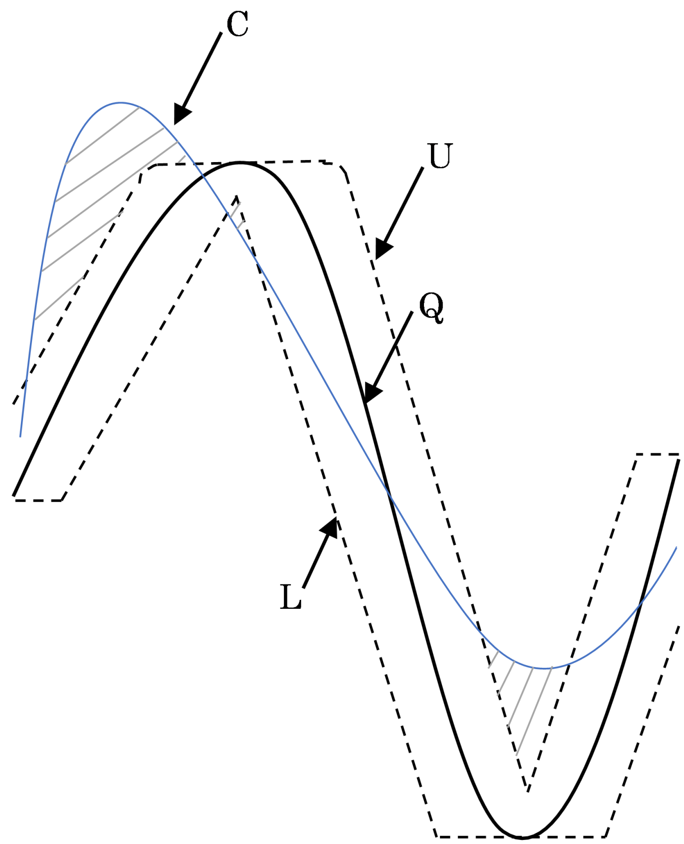


Figure 6.5: Illustration of LB_Keogh distance measure between $Q$ and $C$ with upper and lower bound sequence envelope for time series $Q$

### 6.4.3 Customer Baseline Estimation Algorithm

A baseline estimation algorithm taking into consideration the rebound effect is proposed. In order to achieve this, customers demand profile is partitioned into 3 sections which are pre DR period, DR period, and post DR period. The non absolute cumulative distance is computed between the DR customer and the non DR customers within its cluster for each individual partitions. A 3 step approach is proposed to estimate the baseline of customers.

1. Distance calculation between DR customer and similar non DR customers

2. Demand rebound constraint and non DR customers filtering

3. Shift parameter ($s$) calculation and baseline estimation

### Distance calculation between DR customer and similar non DR customers

Given $m$ number of non DR customers in a cluster $l$ where $l$ can be between 1 and $k$, the pre DR, DR and post DR non absolute distances between DR customer $c$ and a non DR customer $i$ is given by $d_{pre-DR}$, $d_{DR}$, and $d_{post-DR}$, respectively according to (6.4), (6.5) and (6.6) below:

$$d_{pre-DR} = \sum_{j=1}^{t_s} (d_{cj} - d_{ij}), \tag{6.4}$$

$$d_{DR} = \sum_{j=t_s}^{t_e} (d_{cj} - d_{ij}), \tag{6.5}$$

$$d_{post-DR} = \sum_{j=t_e}^{t_n} (d_{cj} - d_{ij}), \tag{6.6}$$

where $t_s$ and $t_e$ represents the start and end times of the DR period, $t_n$ represent the final timestamp of the daily profile (e.g. 48 for a half-hourly demand profile).

### Demand rebound constraint and non DR customers filtering

To accommodate the demand rebound effect in baseline estimation, the sum of $d_{pre-DR}$ and $d_{post-DR}$ for DR customer $c$ and each non DR customer $i$ is computed as shown in (6.7) below:

$$d_{diff} = d_{pre-DR} + d_{post-DR}. \tag{6.7}$$

Customers showing a positive value for $d_{diff}$ is further selected with negative ones filtered out. The baseline candidates is thereby reduced from an initial $m$ non DR customers to $u$ where $m >= u$. This constraint ensures that the sum exceeds what the representative baseline of the customer will be thereby capturing load shift as well as load clipping.

**Shift parameter ($s$) calculation and baseline estimation**

The shift parameter ($s$) is derived as the difference between $d_{post-DR}$ and $d_{diff}$ as shown in (6.8) below:

$$s = d_{DR} - d_{diff}. \tag{6.8}$$

To account for demand shifting from a DR period to either the pre-DR and/or post-DR period, the $s$ must be positive. A defined $v$ number of non DR customers satisfying the positive $s$ condition from the total $u$ non DR customers is used to estimate the baseline. A direct average of these $v$ candidates is proposed.

## 6.5 Results and Discussion

### 6.5.1 Proposed Estimation Methodology

The dataset used in this chapter comprises of 1200 customer smart meter data. The data is normalised and the proposed $k$-medoid clustering algorithm with LB_Keogh distance measure applied. In order to determine the optimal cluster number for the clustering, the COP index is derived for $k$ values ranging from 2 to 15. As shown in Fig. 6.6, the optimal cluster number $k$ for the dataset is 12 as it shows the COP index at its lowest for the input range.

Fig. 6.7 and Fig. 6.8 presents the 12 clusters derived with the member distribution presented in Fig. 6.9.

The aggregate demand of the 1200 customers is shown in Fig. 6.10 with the peak load at half-hourly period 37. A demand response event period from 36 to 40 is used for this study.

A DR customer with ID number 819 is selected for this study. The demand profile for customer 819 is presented in Fig. 6.11. On clustering the 1200 customers, customer ID

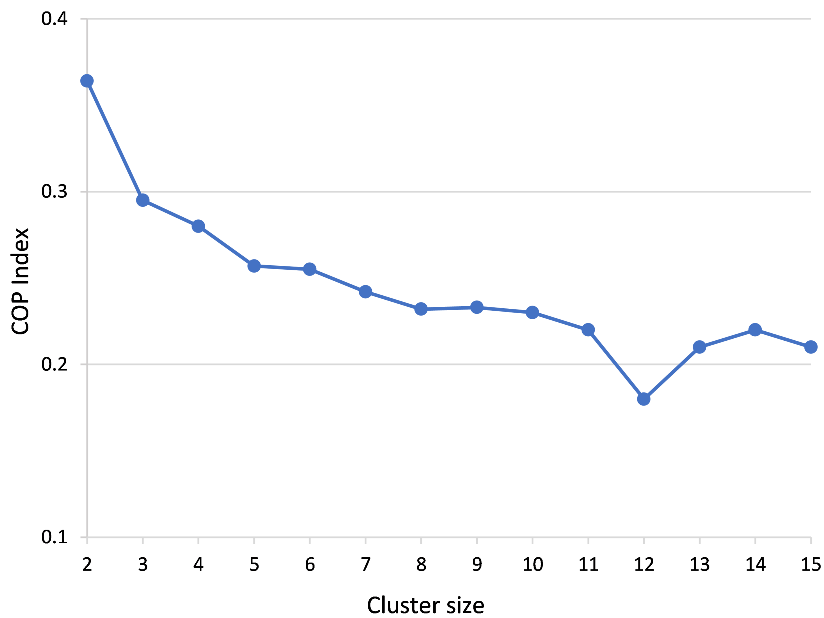


Figure 6.6: COP Index.

819 belongs to cluster 12 along with 177 other customers. Of the 178 customers in cluster 12, 167 are non DR customers with 11 customers participating in DR. Firstly, the other DR customers are filtered leaving 167 candidates for the process of estimating customer ID 819's demand baseline.

Fig. 6.12 shows the scatter plot of the non absolute pre DR distance versus post DR distance, where the non absolute distance is calculated between customer ID 819 and all non DR customers in cluster 12. Points in quadrant I represents non DR customers with both their pre DR and post DR demand lower than that of customer ID 819. Points located in quadrant III shows the opposite. Given the defined constraint from the methodology that the sum of the non absolute pre DR and post DR distance must exceed 0, all customers in quadrant III is filtered out. Customers in quadrant I and IV with the sum of pre DR and post DR distance difference less than 0 is also filtered out.

The non absolute distance between customer ID 819 and the non DR customers for the DR period is also computed i.e. $d_{DR}$. The non DR customers with positive $d_{DR}$ is filtered out. The main reason for this is to capture a reduction in demand during the DR event.

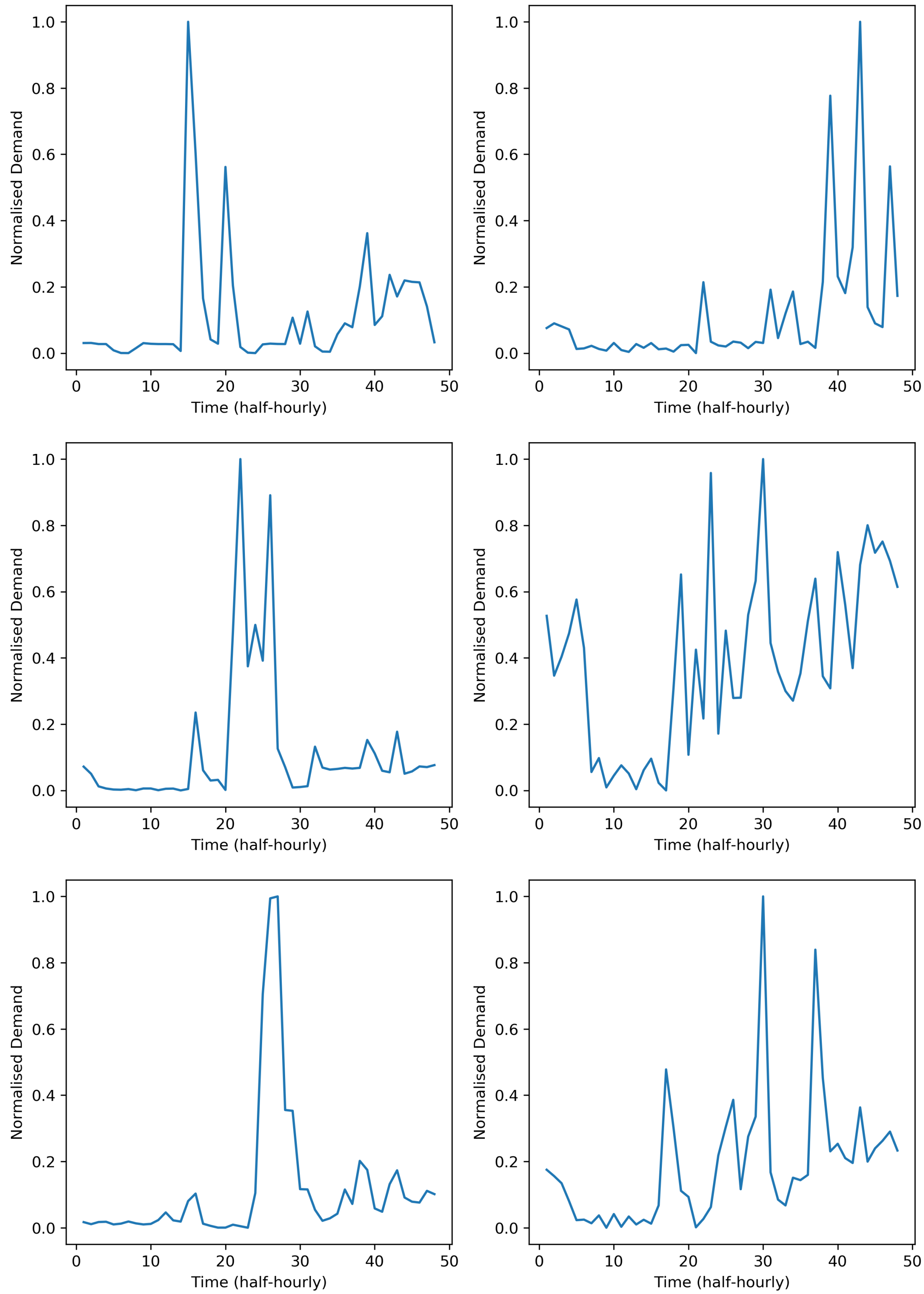


Figure 6.7: Cluster medoids 1-6.

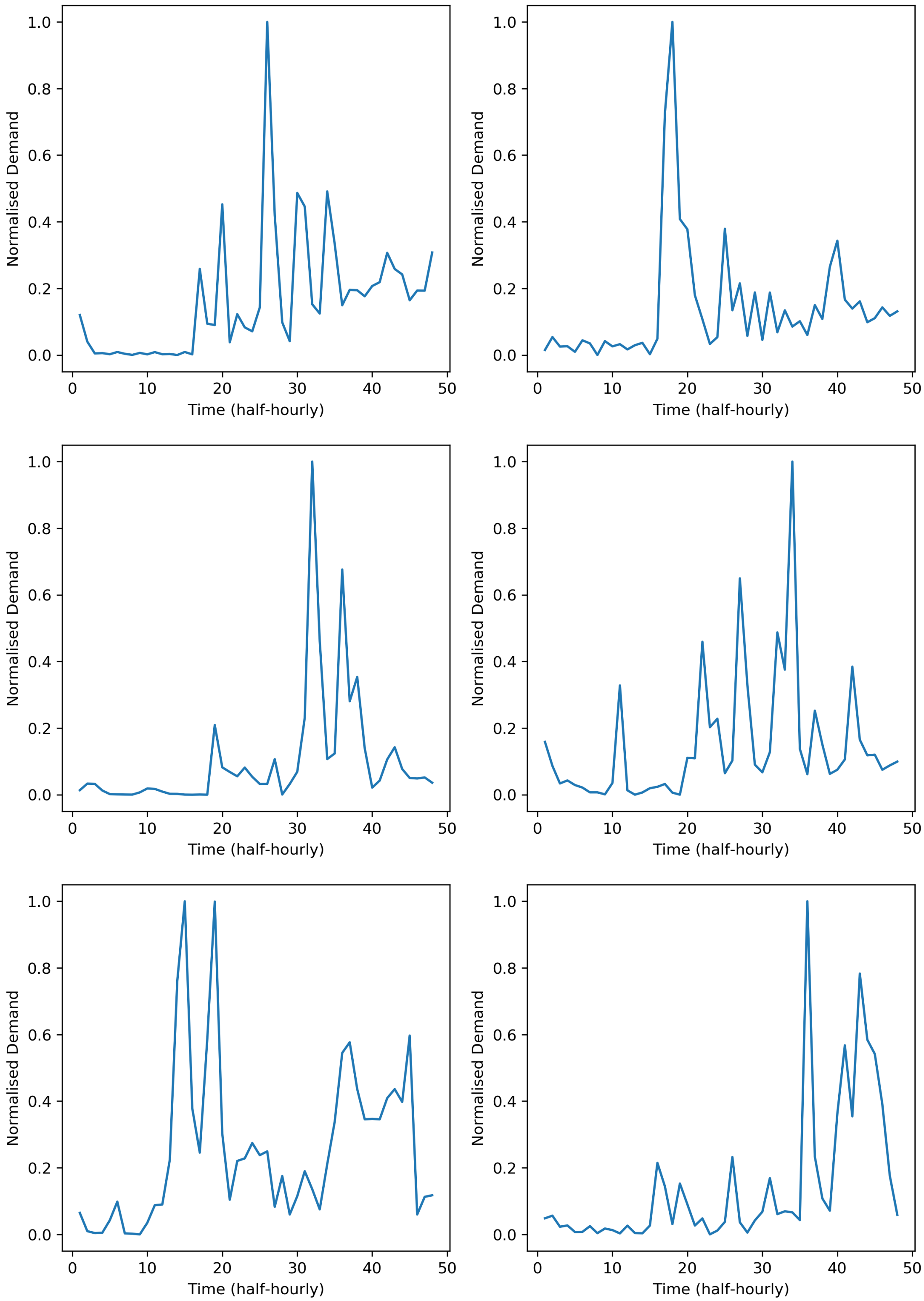


Figure 6.8: Cluster medoids 7-12.

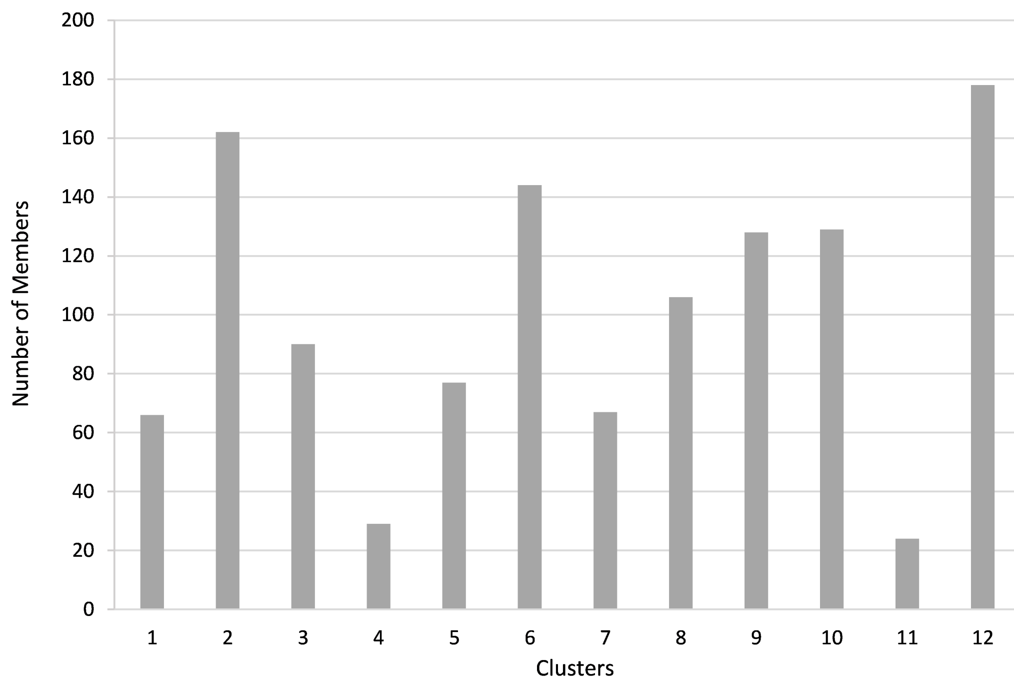


Figure 6.9: Customer cluster distribution.

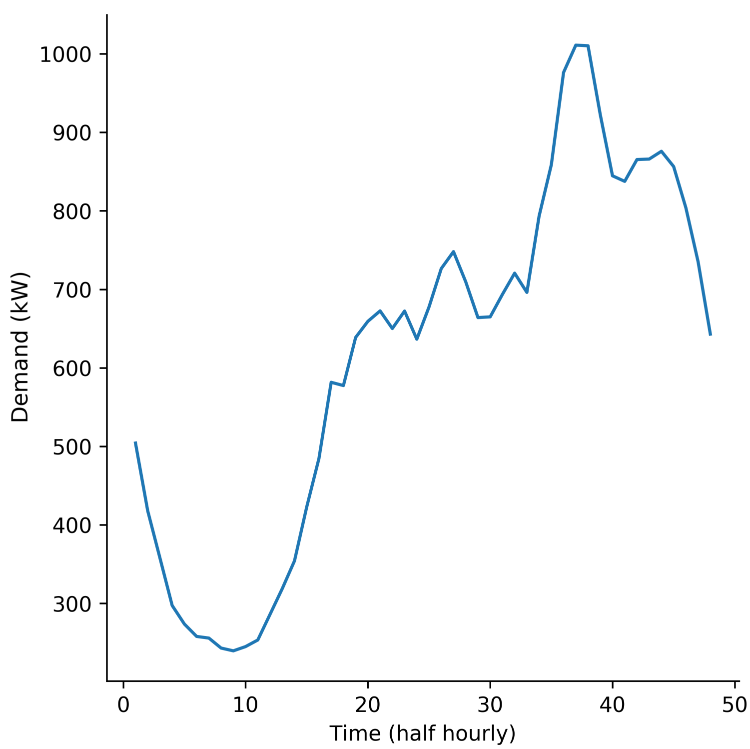


Figure 6.10: Aggregate demand for the 1200 customers.

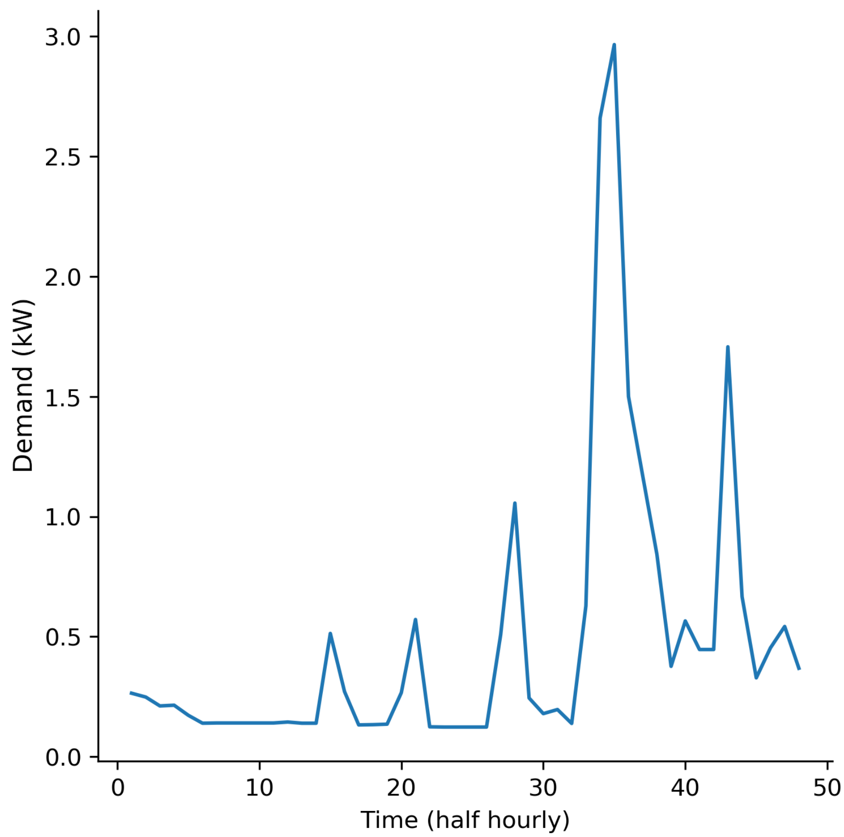


Figure 6.11: Customer 819 demand profile.

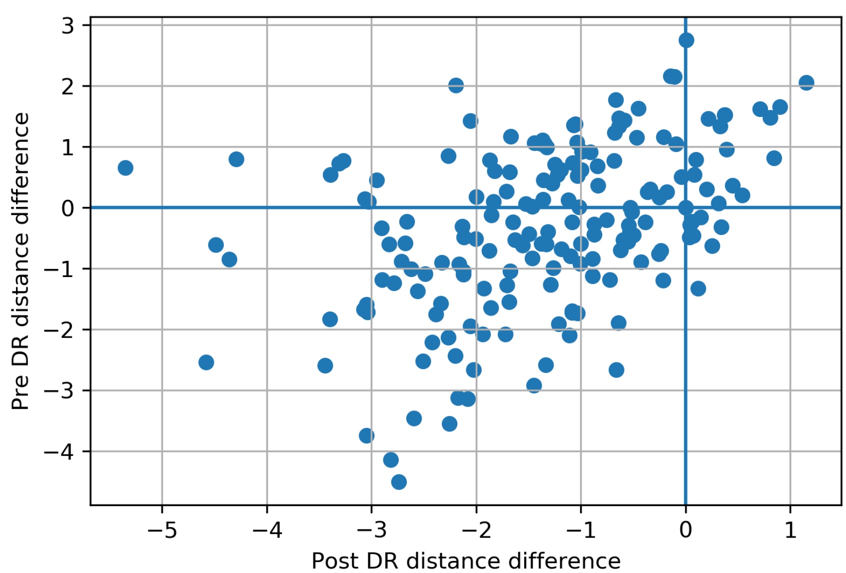


Figure 6.12: Scatter plot of non absolute pre DR distance versus post DR.

Fig. 6.13 shows the scatter plot of $d_{diff}$ versus $d_{DR}$.

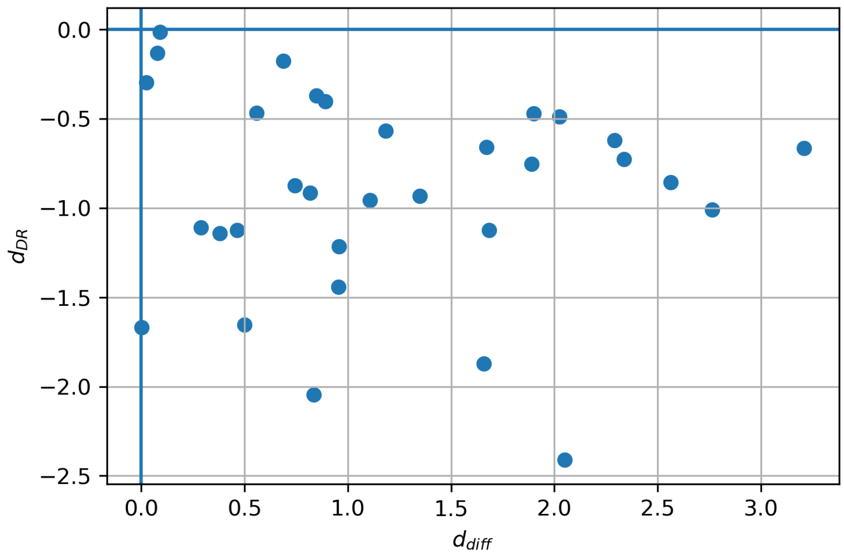


Figure 6.13: Scatter plot of $d_{DR}$ versus $d_{diff}$.

The baseline is derived from the average of up to 5 nearest points to the origin (0,0) of Fig. 6.13. Fig. 6.14 shows the plot of customer ID 819's demand profile and baseline. The baseline shows an estimated demand reduction of 1 with demand shift estimated as 0.85 and demand clipped as 0.15.

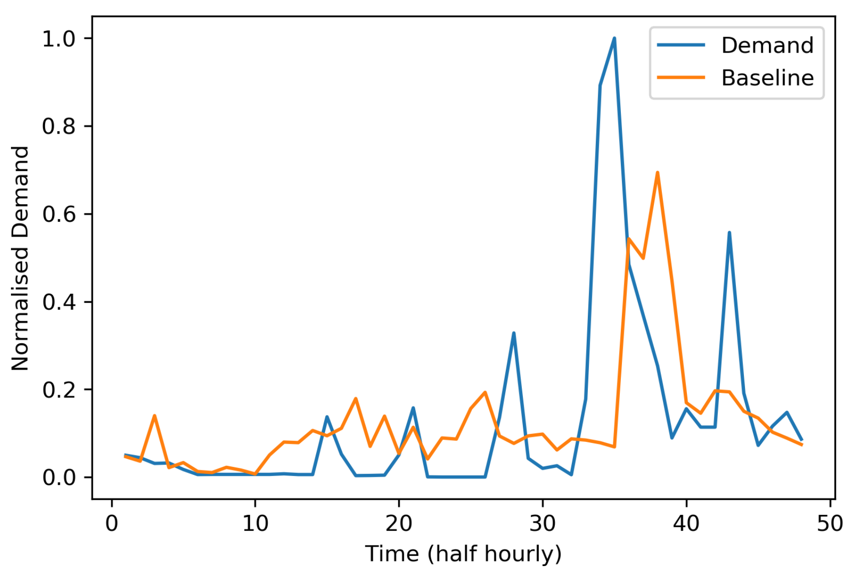


Figure 6.14: Baseline estimation for customer ID 819.

The amount of demand reduction resulting from DR shifting and DR clipping is derived as follows

1. Firstly, the total demand reduction is estimated. The reduction is computed by finding the difference between the actual demand and the baseline estimate during DR event as shown in (6.9),

$$d_{DR} = \sum_{j=t_s}^{t_e} (d_{bas,j} - d_{act,j}), \tag{6.9}$$

   where $d_{DR}$, $d_{bas,j}$, $d_{act,j}$, $t_s$, $t_e$ are demand reduction during DR event, actual demand at time $j$, estimate baseline at time $j$, DR event start time (36) and DR event end time (40), respectively.

2. Demand shift is derived by subtracting the pre DR and post DR sum of the baseline from the actual demand pre DR and post DR sum as shown in (6.10) below:

$$d_{shift} = \sum_{j=1}^{t_s} (d_{act,j} - d_{bas,j}) + \sum_{j=t_e}^{48} (d_{act,j} - d_{bas,j}). \tag{6.10}$$

3. Demand reduced during DR due to clipping is estimated by subtracting the demand shift from the total demand reduction during the DR event as shown in (6.11) below:

$$d_{clip} = d_{DR} - d_{shift}. \tag{6.11}$$

The method is applied to 6 other DR customers randomly selected from clusters 2, 6, 7, 9, 10 and 11 to estimate their baseline. Table 6.1 shows the customer ID, cluster number, demand reduction, demand shifted and demand clipped estimate for each of the DR customers. The plot showing both actual and estimated baseline for each of the 6 customers is shown in Fig. 6.15.

Table 6.1: Demand shift and clip estimate from estimated baseline.

| Customer ID | Cluster | Demand Reduction | Demand Shifted | Demand Clipped |
|---|---|---|---|---|
| 42 | 6 | 0.74 | 0.37 | 0.37 |
| 480 | 10 | 0.80 | 0.44 | 0.36 |
| 758 | 2 | 1.17 | 0.57 | 0.60 |
| 236 | 7 | 0.70 | 0.13 | 0.57 |
| 159 | 9 | 0.80 | 0.33 | 0.47 |
| 923 | 11 | 0.99 | 0.49 | 0.50 |

### 6.5.2 Result Comparison with Baseline Estimation Approaches Utilising Clustering in Literature

The result of the proposed methodology is compared to the following approaches utilising clustering proposed in literature.

- Cluster average: Here, the average of the non DR customers is used to represent the baseline of the DR customers. Zhang *et al.* in [116] demonstrated the effectiveness of this method in estimating customers baseline demand. This approach was also proposed in [149] for estimating DR customers baseline.

- Nearest pre & post DR demand: The non DR customer with the pre DR and post DR nearest to the DR customer is used to estimate the baseline for the DR customer. This approach was utilised as part of the framework for the baseline estimation methodology proposed in [165].

- 5 Nearest average: The closest 5 non DR customers to the DR customer is averaged and used to represent the DR customers' baseline. A similar approach was proposed in [149] although instead of a specific number of non DR customers, a percentage of the total cluster members is proposed to determine the number of nearest non DR customers to be averaged.

- Cluster medoid: In this case, the representative cluster medoid is assigned as the baseline of the DR customers.

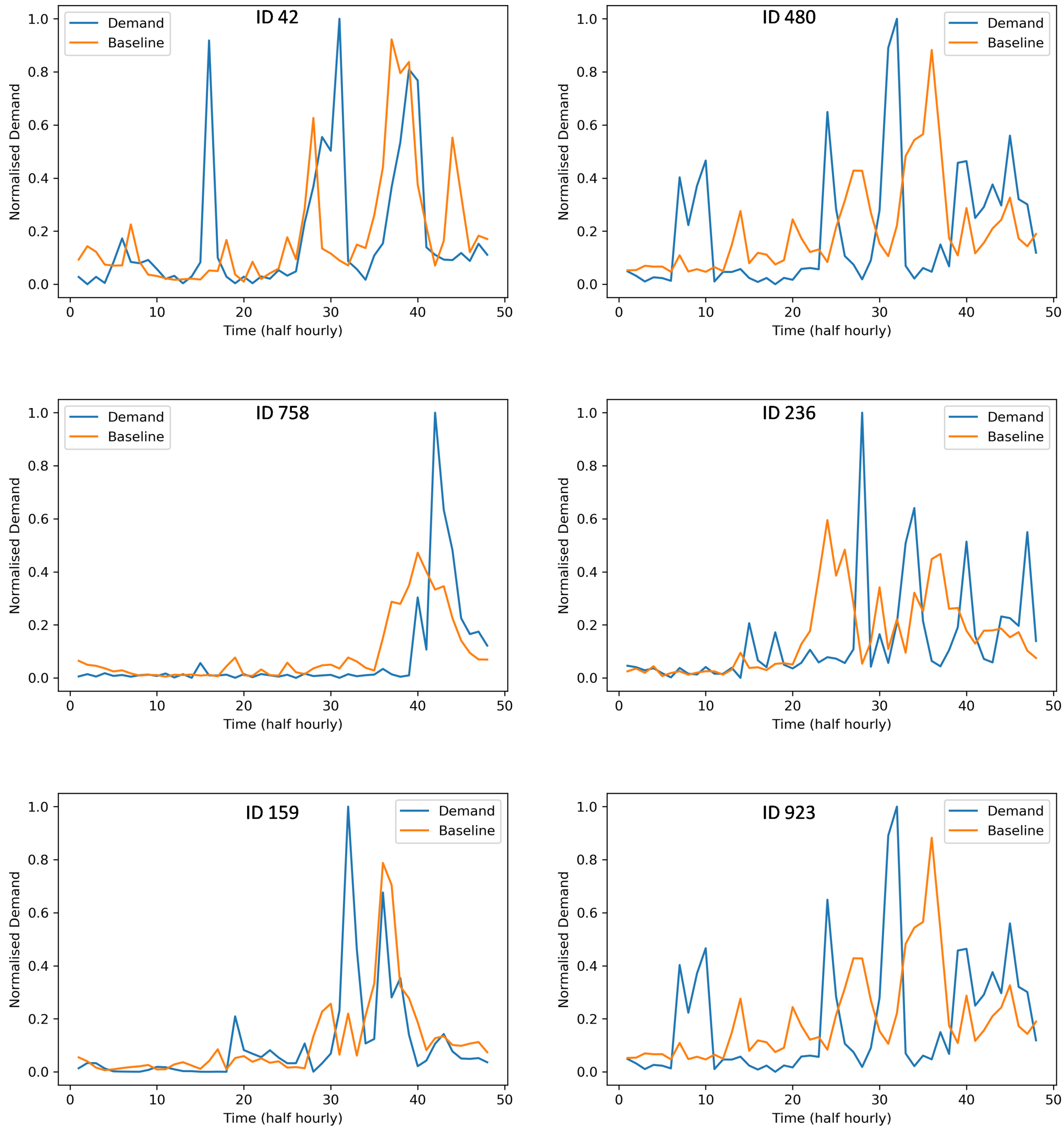


Figure 6.15: Baseline estimation for customer IDs 42, 480, 758, 236, 159 and 923.

Fig. 6.16 shows the comparison of each of the methodologies estimate and how they compare with the actual demand for customer ID 819.

Table 6.2 also presents the estimated demand shifted and demand clipped from the derived baseline. Demand reduction in the table shows the estimated demand reduced by

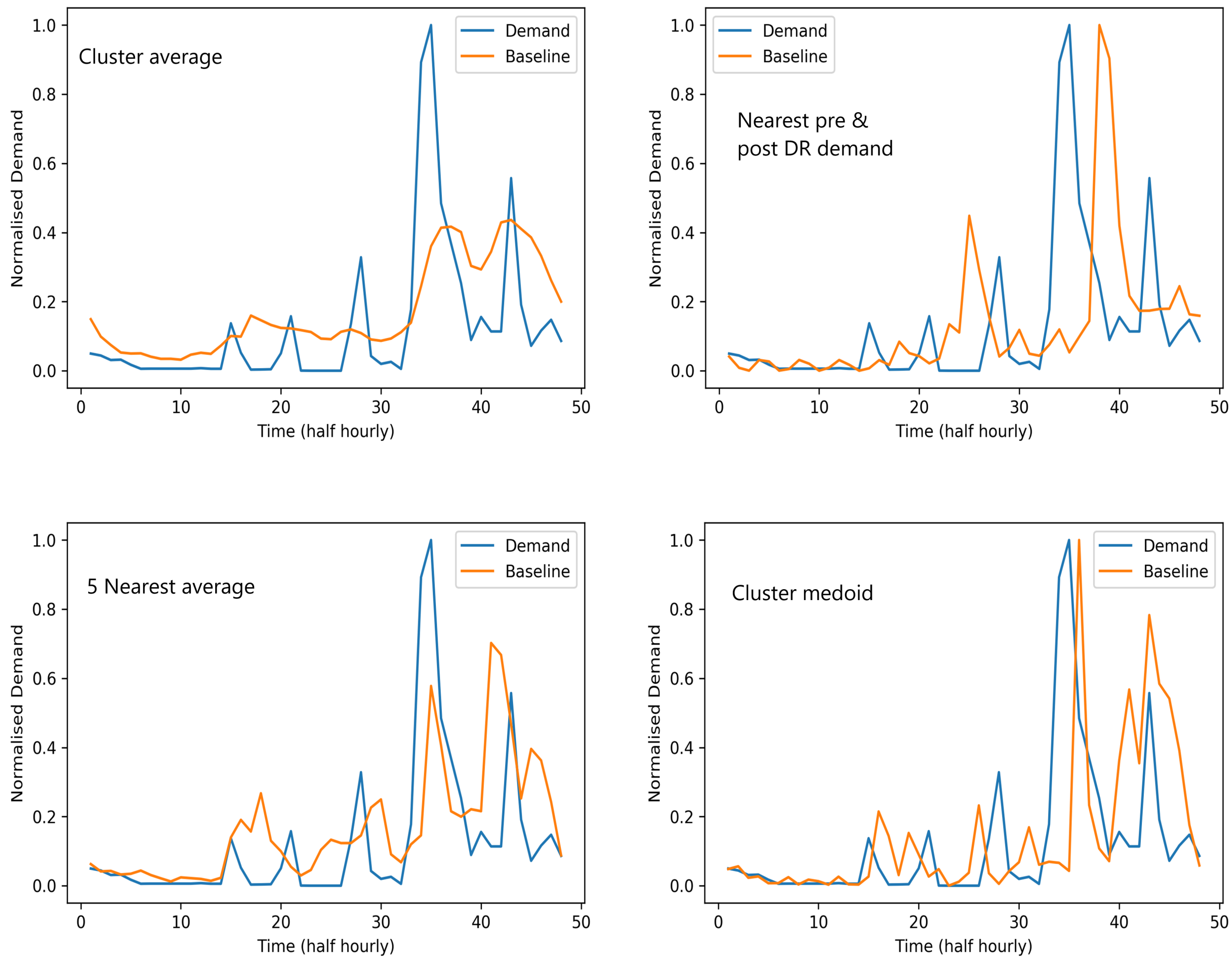


Figure 6.16: Results from other baseline estimation approaches with clustering application.

the customer during the DR period. This reduction consists of demand shifted which is an estimate of demand moved from DR period to a non-DR period and demand clipped which is an estimate of demand cut (non shift component of demand reduction).Compared to the proposed methodology, only the nearest pre and post DR demand approach showed a demand reduction with positive demand curtailed and shifted. A negative value for demand reduction means demand increased during event period for the DR customer. Negative demand shift means demand was shifted into the DR event period while negative demand clip means no demand clip but rather increase in demand.

Despite the positive value of demand reduction, demand shift and demand clip ob-

Table 6.2: Demand shift and clip estimate using other baseline estimation approaches.

| Method | Demand Reduction | Demand Shifted | Demand Clipped |
|---|---|---|---|
| Cluster averaging | 0.48 | -1.74 | 2.22 |
| Nearest pre and post DR demand | 1.22 | 0.96 | 0.26 |
| 5 nearest average | -0.09 | -2.16 | 2.07 |
| Cluster medoid | 0.42 | -0.63 | 1.05 |

tained from the nearest pre and post DR demand approach, there is no guarantee that this will be the case for other DR customers. The proposed method ensures a constraint to avoid negative demand in the estimation hence, ensuring demand reduction is captured whether in form of shifting or clipping.

## 6.6 Conclusion

In this chapter, a novel control based customer baseline estimation methodology was proposed. The methodology takes into account the demand rebound effect in the baseline estimation process. As far as the literature surveyed for customer baseline estimation, this is the first methodology that not just only estimate a DR customers baseline but also estimates the amount of demand reduction clipped as well as shifted to non DR periods.

# Chapter 7

# Conclusion and Future Work

In this chapter, the conclusion and areas for future work is presented.

## 7.1 Thesis Conclusions

Three novel methodology is proposed for enhancing demand response using customers smart meter data. Below is a list showing the summary for each of the proposed methodologies and the area of DR enhancement applied to.

- In chapter 4, a methodology was proposed for targeting customers for peak and local peak load reduction. The peak targeting methodology utilised the *k*-medoid clustering algorithm and dynamic time warping distance measure to group customers based on their demand profile. Results shown from clustering shows that the proposed method performs in targeting customers compared to the direct distance clustering method. Based on comparative analysis of different distance approach for the DTW, it is proposed that for large customers dataset, Manhattan distance can be used in place of Euclidean distance as the clustering showed similar performance in line with the objective of customer targeting for peak load reduction. In addition to peak demand targeting, a novel methodology for targeting customers for local peak demand reduction is proposed. The proposed method uses a 2 dimensional parameter based on customers potential to contribute to DR to rank customers.

- A customer baseline estimation methodology was proposed in chapter 5. The method is based on the LSTM RNN machine learning technique. To minimise the computational expense, the model is trained on the event period only compared to the whole daily period as used in various methods proposed in literature. Estimation results presented showed a better performance in baseline estimation when compared to traditional estimation methods.

- In chapter 6, a novel control based customer baseline estimation methodology was proposed. The methodology takes into account the demand rebound effect in the baseline estimation process. As far as the literature surveyed for customer baseline estimation, this is the first methodology that not just only estimate a DR customers baseline but also estimates the amount of demand reduction clipped as well as shifted to non DR periods.

## 7.2 Future Work

The following key direction highlighted below are possible future work:

- Targeting customers for integrated storage solutions and optimal EV charging with the objective of flattening aggregate demand profile. This area is an extension to targeting customers not just for peak load reduction but also for achieving desirable aggregate demand profile from the suppliers end using storage and EV load. The transitioning of vehicle mobility to electric will bring challenges to load management and a targeted approach coupled with customers demand data can help manage the potential strain a wide spread adoption of EVs will cause.

- Extending the LSTM model training to other periods of customers demand profile can help with a modular approach to estimating customer baseline for the whole demand profile. This approach can help create insight into periods where the load is shifted to as well as amount of demand clipped as proposed in chapter 6. Also, considering the effect of weather and other external influencing factors like holidays, adding features derived from these external data can help improved the estimation accuracy of the proposed model.

- Drilling through the customer demand profile up to the appliance level can help enhance the estimation accuracy for baseline as well as demand clipped and shifted. Studying the estimation performance with appliance level data is an interesting area for future studies

# Bibliography

[1] European Environment Agency, “Sectoral grennhouse gas emissions by IPCC sector,” https://www.eea.europa.eu/data-and-maps/daviz/change-of-co2-eq-emissions-2, [Online; accessed 05-Jan.-2021].

[2] IEA, “Global energy  CO2 status report 2019,” https://www.iea.org/reports/global-energy-co2-status-report-2019, [Online; accessed 05-Jan.-2021].

[3] H. Gharavi and R. Ghafurian, “Smart grid: The electric energy system of the future,” *Proceedings of the IEEE*, vol. 99, no. 6, pp. 917–921, Jun. 2011.

[4] U.S. Department of Energy, “Benefits of demand response in electricity markets and recommendations for achieving them,” DOE, Tech. Rep., Feb. 2006.

[5] R. Deng, Z. Yang, M.-Y. Chow, and J. Chen, “A survey on demand response in smart grids: Mathematical models and approaches,” *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 570–582, Jun. 2015.

[6] M. H. Albadi and E. El-Saadany, “A summary of demand response in electricity markets,” *Electric Power Systems Research*, vol. 78, no. 11, pp. 1989–1996, May 2008.

[7] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, “A survey on demand response programs in smart grids: Pricing methods and optimization algorithms,” *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 152–178, Firstquarter 2015.

[8] D. Li and S. K. Jayaweera, “Uncertainty modeling and price-based demand response scheme design in smart grid,” *IEEE Systems Journal*, vol. 11, no. 3, pp. 1743–1754, Sep. 2017.

[9] Y. Kim, “Optimal price-based demand response of hvac systems in multi-zone office buildings considering thermal preferences of individual occupants,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 5060–5073, Nov. 2018.

[10] L. Park, Y. Jang, S. Cho, and J. Kim, “Residential demand response for renewable energy resources in smart grid systems,” *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3165–3173, Dec. 2017.

[11] Ontario Hydro, “Ontario hydro rates,” http://www.ontario-hydro.com/current-rates, [Online; accessed 05-Nov.-2017].

[12] UK Power, “Everything you need to know about Economy 7,” https://www.ukpower.co.uk/home_energy/economy-7, [Online; accessed 05-Nov.-2017].

[13] L. Zhao, Z. Yang, and W.-J. Lee, “The impact of time-of-use (TOU) rate structure on consumption patterns of the residential customers,” *IEEE Transactions on Industry Applications*, vol. 53, no. 6, pp. 5130–5138, Nov. 2017.

[14] K. Herter, “Residential implementation of critical-peak pricing of electricity,” *Energy Policy*, vol. 35, no. 4, pp. 2121–2130, Apr. 2007.

[15] Electricite de France, “Blue tariff,” https://particulier.edf.fr/en/home/energy-at-home/electricity/blue-tariff.html, 2015, [Online; accessed 04-Jul.-2017].

[16] VTT Technical Research Centre of Finland, “Energy efficiency and load curve impacts of commercial development in competitive markets-results from the EFFLOCOM pilots,” VTT, Tech. Rep., Jun. 2004.

[17] H. Aalami, M. P. Moghaddam, and G. Yousefi, "Modeling and prioritizing demand response programs in power markets," *Electric Power Systems Research*, vol. 80, no. 4, pp. 426–435, Apr. 2010.

[18] B. Chai, J. Chen, Z. Yang, and Y. Zhang, "Demand response management with multiple utility companies: A two-level game approach," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 722–731, Mar. 2014.

[19] K. M. Tsui and S.-C. Chan, "Demand response optimization for smart home scheduling under real-time pricing," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1812–1821, Dec. 2012.

[20] F. Kamyab, M. Amini, S. Sheykhha, M. Hasanpour, and M. M. Jalali, "Demand response program in smart grid using supply function bidding mechanism," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1277–1284, May 2016.

[21] P. Samadi, A.-H. Mohsenian-Rad, R. Schober, V. W. Wong, and J. Jatskevich, "Optimal real-time pricing algorithm based on utility maximization for smart grid," in *Proc. of First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Maryland, MD, Oct. 2010.

[22] R. Tulabing, R. Yin, N. DeForest, Y. Li, K. Wang, T. Yong, and M. Stadler, "Modeling study on flexible load's demand response potentials for providing ancillary services at the substation level," *Electric Power Systems Research*, vol. 140, pp. 240–252, Nov. 2016.

[23] K. Valogianni and W. Ketter, "Effective demand response for smart grids: Evidence from a real-world pilot," *Decision Support Systems*, vol. 91, pp. 48–66, 2016.

[24] P. Zikopoulos, C. Eaton, D. Deroos, T. Deutsch, and G. Lapis, *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 2011.

[25] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, Jan. 2014.

[26] L. O'Malley, "The evolving digital utility: The convergence of energy and IT," MaRS Market Insights, Tech. Rep., Aug. 2014.

[27] Department for Business, Energy and Industrial Strategy, "Smart meters: Quarterly report to end september 2017," BEIS, Tech. Rep., Nov. 2017.

[28] S. Sagiroglu and D. Sinanc, "Big data: A review," in *Proc. of IEEE International Conference on Collaboration Technologies and Systems (CTS)*, San Diego, CA, May 2013.

[29] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big data: Issues and challenges moving forward," in *Proc. of the 46th Hawaii International Conference on System Sciences (HICSS)*, Honolulu, HI, Jan. 2013.

[30] P. Buneman, "Semistructured data," in *Proc. of the 16th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, City, AZ, May 1997.

[31] N. Koseleva and G. Ropaite, "Big data in building energy efficiency: understanding of big data and main challenges," *Procedia Engineering*, vol. 172, pp. 544–549, 2017.

[32] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: a technology tutorial," *IEEE Access*, vol. 2, pp. 652–687, Jun. 2014.

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

[34] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. MIT press, 2012.

[35] N. C. Krishnan, "CSL 603: Machine learning," http://cse.iitrpr.ac.in/ckn/courses/f2017/csl603/w1n.pdf, [Online; accessed 26-Mar.-2018].

[36] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Informatica*, vol. 31, pp. 249–268, Oct. 2007.

[37] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[38] K. Yue, Q. Fang, X. Wang, J. Li, and W. Liu, "A parallel and incremental approach for data-intensive learning of bayesian networks," *IEEE transactions on cybernetics*, vol. 45, no. 12, pp. 2890–2904, Dec. 2015.

[39] K. Yue, H. Wu, X. Fu, J. Xu, Z. Yin, and W. Liu, "A data-intensive approach for discovering user similarities in social behavioral interactions based on the bayesian network," *Neurocomputing*, vol. 219, pp. 364–375, Sep. 2017.

[40] B. Liu, E. Blasch, Y. Chen, D. Shen, and G. Chen, "Scalable sentiment classification for big data analysis using Naive Bayes classifier," in *Proc. of IEEE International Conference on Big Data*, Silicon Valley, CA, Oct. 2013.

[41] X. Han, L. Xu, M. Ren, and W. Gu, "A Naive Bayesian network intrusion detection algorithm based on principal component analysis," in *Proc. of IEEE 7th International Conference on Information Technology in Medicine and Education (ITME)*, Huangshan, China, Nov. 2015.

[42] X. Liu, R. Lu, J. Ma, L. Chen, and B. Qin, "Privacy-preserving patient-centric clinical decision support system on naive bayesian classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 2, pp. 655–668, Mar. 2016.

[43] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[44] P. Rebentrost, M. Mohseni, and S. Lloyd, “Quantum support vector machine for big data classification,” *Physical Review Letters*, vol. 113, no. 13, p. 130503, Sep. 2014.

[45] X. Ke, H. Jin, X. Xie, and J. Cao, “A distributed SVM method based on the iterative MapReduce,” in *Proc. of IEEE International Conference on Semantic Computing (ICSC)*, Anaheim, CA, Feb. 2015.

[46] X. Wang, X. Liu, and S. Matwin, “A distributed instance-weighted SVM algorithm on large-scale imbalanced datasets,” in *Proc. of IEEE International Conference on Big Data*, Washington, DC, Oct. 2014.

[47] D. Singh, D. Roy, and C. K. Mohan, “DiP-SVM: Distribution preserving kernel support vector machine for big data,” *IEEE Transactions on Big Data*, vol. 3, no. 1, pp. 79–90, Mar. 2017.

[48] S. Nan, L. Sun, B. Chen, Z. Lin, and K.-A. Toh, “Density-dependent quantized least squares support vector machine for large data sets,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 1, pp. 94–106, Jan. 2017.

[49] B. Wang, B. Fang, Y. Wang, H. Liu, and Y. Liu, “Power system transient stability assessment based on big data and the core vector machine,” *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2561–2570, Sep. 2016.

[50] I. W. Tsang, J. T. Kwok, and P.-M. Cheung, “Core vector machines: Fast SVM training on very large data sets,” *Journal of Machine Learning Research*, vol. 6, pp. 363–392, May 2005.

[51] Y.-Y. Song and L. Ying, “Decision tree methods: applications for classification and prediction,” *Shanghai Archives of Psychiatry*, vol. 27, no. 2, pp. 130–135, Apr. 2015.

[52] A. Swetapadma and A. Yadav, "A novel decision tree regression-based fault distance estimation scheme for transmission lines," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 234–245, Feb. 2017.

[53] N. Siddique and H. Adeli, *Computational intelligence: synergies of fuzzy logic, neural networks and evolutionary computing*. John Wiley & Sons, 2013.

[54] A. Segatori, F. Marcelloni, and W. Pedrycz, "On distributed fuzzy decision trees for big data," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 1, pp. 174–192, Feb. 2018.

[55] H. A. Wisesa, M. A. Ma'sum, P. Mursanto, and A. Febrian, "Processing big data with decision trees: a case study in large traffic data," in *Proc. of International Workshop on Big Data and Information Security (IWBIS)*, Jakarta, Indonesia, Oct. 2016.

[56] J. Chen, T. Wang, R. Abbey, and J. Pingenot, "A distributed decision tree algorithm and its implementation on big data platforms," in *Proc. of IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Montreal, Canada, Oct. 2016.

[57] J. O. Rawlings, S. G. Pantula, and D. A. Dickey, *Applied regression analysis: a research tool*. Springer Science & Business Media, Apr. 2001.

[58] Z. Ghahramani, "Machine learning," Machine Learning Group, Department of Engineering, University of Cambridge, 2015.

[59] K. Yu, K. Aristodemou, F. Becker, and J. Lord, "Fast mode regression in big data analysis," in *Proc. of ACM International Conference on Big Data Science and Computing*, Beijing, Aug. 2014.

[60] M. Raissi, "Parametric gaussian process regression for big data," *arXiv preprint arXiv:1704.03144*, Apr. 2017.

[61] J. Fiosina and M. Fiosins, "Distributed non-parametric and semi-parametric regression on spark for big data forecasting," *Applied Computational Intelligence and Soft Computing*, Mar. 2017.

[62] Y. Gal, M. van der Wilk, and C. E. Rasmussen, "Distributed variational inference in sparse gaussian process regression and latent variable models," in *Proc. of International Conference on Advances in Neural Information Processing Systems*, Montreal, Canada, Dec. 2014.

[63] G. Smrithy, S. Munirathinam, and R. Balakrishnan, "Online anomaly detection using non-parametric technique for big data streams in cloud collaborative environment," in *Proc. of IEEE International Conference on Big Data*, Washington, DC, Dec. 2016, pp. 921–932.

[64] J. L. Powell, "Estimation of semiparametric models," *Handbook of econometrics*, vol. 4, pp. 2443–2521, Dec. 1994.

[65] D. Ruppert, M. P. Wand, and R. J. Carroll, *Semiparametric regression*. Cambridge University Press, Jul. 2003.

[66] A. R. Saboo, V. Kumar, and I. Park, "Using big data to model time-varying effects for marketing resource (re)allocation," *MIS Quarterly*, vol. 40, no. 4, pp. 911–939, Dec. 2016.

[67] E. X. Fang, M.-D. Li, M. I. Jordan, and H. Liu, "Mining massive amounts of genomic data: a semiparametric topic modeling approach," *Journal of the American Statistical Association*, vol. 112, no. 519, pp. 921–932, 2017.

[68] J. Luts, "Real-time semiparametric regression for distributed data sets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 2, pp. 545–557, Feb. 2015.

[69] Y. Goude, R. Nedellec, and N. Kong, “Local short and middle term electricity load forecasting with semi-parametric additive models,” *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 440–446, Jan. 2014.

[70] P. Dayan, M. Sahani, and G. Deback, “Unsupervised learning,” *The MIT encyclopedia of the cognitive sciences*, Oct. 1999.

[71] P. Berkhin, “A survey of clustering data mining techniques,” in *Grouping multidimensional data*. Springer, 2006, pp. 25–71.

[72] I. Kusuma, M. A. Ma’sum, N. Habibie, W. Jatmiko, and H. Suhartanto, “Design of intelligent k-means based on spark for big data clustering,” in *Proc. of IEEE Internal Workshop on Big Data and Information Security (IWBIS)*, Jakarta, Indonesia, Oct. 2016.

[73] J. Judith and J. Jayakumari, “Distributed document clustering analysis based on a hybrid method,” *China Communications*, vol. 14, no. 2, pp. 131–142, Feb. 2017.

[74] J. Yuan and Y. Tian, “Practical privacy-preserving mapreduce based k-means clustering over large-scale dataset,” *IEEE Transactions on Cloud Computing*, vol. PP, pp. 1–1, Jan. 2017.

[75] Y.-T. Zhu, F.-z. Wang, X.-H. Shan, and X.-Y. Lv, “K-medoids clustering based on mapreduce and optimal search of medoids,” in *Proc. of IEEE International Conference on Computer Science & Education (ICCSE)*, Vancouver, Canada, Aug. 2014.

[76] M. O. Shafiq and E. Torunski, “A parallel k-medoids algorithm for clustering based on mapreduce,” in *Proc. of IEEE International Conference on Machine Learning and Applications (ICMLA)*, Anaheim, CA, Dec. 2016.

[77] Y. Jeon, J. Yoo, J. Lee, and S. Yoon, “NC-link: A new linkage method for efficient hierarchical clustering of large-scale data,” *IEEE Access*, vol. 5, pp. 5594–5608, Apr. 2017.

[78] M. Bendechache, N.-A. Le-Khac, and M.-T. Kechadi, "Hierarchical aggregation approach for distributed clustering of spatial datasets," in *Proc. of IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, Barcelona, Spain, Dec. 2016.

[79] T.-S. Xu, H.-D. Chiang, G.-Y. Liu, and C.-W. Tan, "Hierarchical k-means method for clustering large-scale advanced metering infrastructure data," *IEEE Transactions on Power Delivery*, vol. 32, no. 2, pp. 609–616, Apr. 2017.

[80] J. Larsen, A. S. Have, and L. K. Hansen, "Probabilistic hierarchical clustering with labeled and unlabeled data," *International Journal of Knowledge-Based and Intelligent Engineering Systems*, vol. 6, no. 1, pp. 56–62, 2002.

[81] K. A. Heller and Z. Ghahramani, "Bayesian hierarchical clustering," in *Proc. of ACM 22nd International Conference on Machine Learning*, Bonn, Germany, Aug. 2005.

[82] A. Anthony and M. Desjardins, "Generative models for clustering: the next generation," in *Proc. of AAAI Spring Symposium*, Stanford, CA, Mar. 2008.

[83] W. Fan, H. Sallay, and N. Bouguila, "Online learning of hierarchical pitman-yor process mixture of generalized dirichlet distributions with feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 9, pp. 2048–2061, Sep. 2017.

[84] J. Fan, "OPE-HCA: An optimal probabilistic estimation approach for hierarchical clustering algorithm," *Neural Computing and Applications*, pp. 1–11, Aug. 2015.

[85] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. of Second International Conference on Knowledge Discovery and Data Mining*, Portland, OR, Aug. 1996.

[86] J. Shen, X. Hao, Z. Liang, Y. Liu, W. Wang, and L. Shao, "Real-time superpixel segmentation by dbscan clustering algorithm," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5933–5942, Dec. 2016.

[87] K. Khan, S. U. Rehman, K. Aziz, S. Fong, and S. Sarasvady, "DBSCAN: Past, present and future," in *Proc. of Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT)*, Bangalore, India, Feb. 2014.

[88] P. Liu, D. Zhou, and N. Wu, "VDBSCAN: varied density based spatial clustering of applications with noise," in *Proc. of IEEE International Conference on Service Systems and Service Management*, Chengdu, China, Jun. 2007.

[89] C. R. Valêncio, G. P. Daniel, C. A. De Medeiros, A. M. Cansian, L. C. Baida, and F. Ferrari, "VDBSCAN+: performance optimization based on GPU parallelism," in *Proc. of IEEE International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, Taipei, Taiwan, Dec. 2013.

[90] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: ordering points to identify the clustering structure," in *Proc. of ACM SIGMOD International Confernce on Management of Data*, Philadelphia, Jun. 1999.

[91] Z. Deng, Y. Hu, M. Zhu, X. Huang, and B. Du, "A scalable and fast optics for clustering trajectory big data," *Cluster Computing*, vol. 18, no. 2, pp. 549–562, Jun. 2015.

[92] M. M. A. Patwary, D. Palsetia, A. Agrawal, W.-K. Liao, F. Manne, and A. Choudhary, "Scalable parallel OPTICS data clustering using graph algorithmic techniques," in *Proc. of International Conference on High Performance Computing, Networking, Storage and Analysis*, Denver, CO, Nov. 2013.

[93] A. Hinneburg and D. A. Keim, "An efficient approach to clustering in large multimedia databases with noise," in *Proc. of the Fourth International Conference on Knowledge Discovery and Data Mining*, New York, Aug. 1998.

[94] A. Idrissi, H. Rehioui, A. Laghrissi, and S. Retal, "An improvement of DENCLUE algorithm for the data clustering," in *Proc. of IEEE 5th International Conference on Information & Communication Technology and Accessibility (ICTA)*, Marrakech, Morocco, Dec. 2015.

[95] G. Sheng, H. Hou, X. Jiang, and Y. Chen, "A novel association rule mining method of big data for power transformers state parameters based on probabilistic graph model," *IEEE Transactions on Smart Grid*, vol. 9, pp. 695–702, Mar. 2018.

[96] H.-Y. Chang, Z.-H. Hong, T.-L. Lin, W.-K. Chang, and Y.-Y. Lin, "IPARBC: an improved parallel association rule based on mapreduce framework," in *Proc. of IEEE International Conference on Networking and Network Applications (NaNA)*, Hakodate, Japan, Jul. 2016.

[97] R. Liu, K. Yang, Y. Sun, T. Quan, and J. Yang, "Spark-based rare association rule mining for big datasets," in *Proc. of IEEE International Conference on Big Data*, Washington, DC, Dec. 2016.

[98] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, May 1996.

[99] K. Chen, Z. He, S. X. Wang, J. Hu, L. Li, and J. He, "Learning-based data analytics: Moving towards transparent power grids," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 1, pp. 67–82, Mar. 2018.

[100] Z. H. Zhou, *Machine Learning*. Tsinghua University Press, 2016.

[101] E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, and J. Slootweg, "On-line building energy optimization using deep reinforcement learning," *IEEE Transactions on Smart Grid*, pp. 1–1, 2018.

[102] A. S. Polydoros and L. Nalpantidis, "Survey of model-based reinforcement learning: applications on robotics," *Journal of Intelligent & Robotic Systems*, vol. 86, no. 2, pp. 153–173, May 2017.

[103] P. Dayan and Y. Niv, "Reinforcement learning: the good, the bad and the ugly," *Current Opinion in Neurobiology*, vol. 18, no. 2, pp. 185–196, Apr. 2008.

[104] Q. Wei, D. Liu, Y. Liu, and R. Song, "Optimal constrained self-learning battery sequential management in microgrid via adaptive dynamic programming," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 2, pp. 168–176, Apr. 2017.

[105] Q. Wei, D. Liu, F. L. Lewis, Y. Liu, and J. Zhang, "Mixed iterative adaptive dynamic programming for optimal battery energy control in smart residential microgrids," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 4110–4120, May 2017.

[106] K. Al-Jabery, Z. Xu, W. Yu, D. C. Wunsch, J. Xiong, and Y. Shi, "Demand-side management of domestic electric water heaters using approximate dynamic programming," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 36, no. 5, pp. 775–788, May 2017.

[107] D. Liu, Y. Xu, Q. Wei, and X. Liu, "Residential energy scheduling for variable weather solar energy based on adaptive dynamic programming," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, pp. 36–46, Jan. 2018.

[108] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine learning*, vol. 3, no. 1, pp. 9–44, Feb. 1988.

[109] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*. MIT Press Cambridge, Nov. 2017.

[110] Y. Wang and M. Pedram, "Model-free reinforcement learning and Bayesian classification in system-level power management," *IEEE Transactions on Computers*, vol. 65, no. 12, pp. 3713–3726, Dec. 2016.

[111] R. Yousefian and S. Kamalasadan, "Design and real-time implementation of optimal power system wide-area system-centric controller based on temporal difference learning," *IEEE Transactions on Industry Applications*, vol. 52, no. 1, pp. 395–406, Jan. 2016.

[112] F. Rahimi and A. Ipakchi, "Demand response as a market resource under the smart grid paradigm," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 82–88, Jun. 2010.

[113] F.-L. Meng and X.-J. Zeng, "A profit maximization approach to demand response management with customers behavior learning in smart grid," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1516–1529, May 2016.

[114] D. Zhou, M. Balandat, and C. Tomlin, "A Bayesian perspective on residential demand response using smart meter data," *arXiv preprint arXiv:1608.03862*, Sep. 2016.

[115] Y. Bai, H. Zhong, and Q. Xia, "Real-time demand response potential evaluation: A smart meter driven method," in *Proc. of IEEE Power and Energy Society General Meeting (PESGM)*, Boston, MA, Jul. 2016.

[116] Y. Zhang, W. Chen, R. Xu, and J. Black, "A cluster-based method for calculating baselines for residential loads," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2368–2377, Sep. 2016.

[117] Z. Guo, Z. J. Wang, and A. Kashani, "Home appliance load modeling from aggregated smart meter data," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 254–262, Jan. 2015.

[118] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of electricity consumption behavior dynamics toward big data applications," *IEEE Transaction on Smart Grid*, vol. 7, no. 5, pp. 2437–2447, Sep. 2016.

[119] S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 136–144, Jan. 2016.

[120] Y. Wang, Q. Chen, C. Kang, Q. Xia, and M. Luo, "Sparse and redundant representations-based smart meter data compression and pattern extraction," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2142–2151, May 2016.

[121] D. Vercamer, B. Steurtewagen, D. Van den Poel, and F. Vermeulen, "Predicting consumer load profiles using commercial and open data," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3693–3701, Sep. 2016.

[122] R. Pal, C. Chelmis, M. Frincu, and V. Prasanna, "Towards dynamic demand response on efficient consumer grouping algorithmics," *IEEE Transactions on Sustainable Computing*, vol. 1, no. 1, pp. 20–34, Jan. 2016.

[123] 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, Sep. 2016.

[124] J. Oyedokun, S. Bu, Y. Xiao, and Z. Han, "Smart meter data characterization and clustering for peak demand targeting in smart grids," in *Proc. of IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, Sarajevo, Bosnia and Herzegovina, Oct. 2018.

[125] M. Ding, X. Wang, J. Wang, Z. Yang, H. Zhong, and J. Yang, “A dynamic period partition method for time-of-use pricing with high-penetration renewable energy,” in *Proc. of IEEE Conference on Energy Internet and Energy System Integration(EI2)*, Beijing, China, Nov. 2017.

[126] J. Oyedokun, S. Bu, Z. Han, and X. Liu, “Customer baseline load estimation for incentive-based demand response using long short-term memory recurrent neural network,” in *Proc. of IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, Bucharest, Romania, Sep. 2019.

[127] B.-G. Kim, Y. Zhang, M. Van Der Schaar, and J.-W. Lee, “Dynamic pricing and energy consumption scheduling with reinforcement learning,” *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2187–2198, Sep. 2016.

[128] Z. Wen, D. ONeill, and H. Maei, “Optimal demand response using device-based reinforcement learning,” *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2312–2324, Sep. 2015.

[129] S. R. Kuppannagari, R. Kannan, C. Chelmis, A. S. Tehrani, and V. K. Prasanna, “Optimal customer targeting for sustainable demand response in smart grids1,” *Procedia Computer Science*, vol. 80, pp. 324–334, Jun. 2016.

[130] N. K. Kandasamy, K. J. Tseng, and S. Boon-Hee, “Virtual storage capacity using demand response management to overcome intermittency of solar PV generation,” *IET Renewable Power Generation*, vol. 11, no. 14, pp. 1741–1748, Oct. 2017.

[131] G. Aghajani, H. Shayanfar, and H. Shayeghi, “Demand side management in a smart micro-grid in the presence of renewable generation and demand response,” *Energy*, vol. 126, pp. 622–637, Mar. 2017.

[132] A. Ghasemkhani and L. Yang, “Reinforcement learning based pricing for demand response,” in *Proc. of IEEE International Conference on Communications Workshops*. Kansas, USA: IEEE, May 2018.

[133] C. Chelmis, J. Kolte, and V. K. Prasanna, “Big data analytics for demand response: Clustering over space and time,” in *Proc. of IEEE International Conference on Big Data (Big Data)*, Santa Clara, CA, Oct. 2015.

[134] L. I. Smith, “A tutorial on principal components analysis,” *Cornell University, USA*, vol. 51, no. 52, p. 65, Feb. 2002.

[135] T. Chen, K. Qian, A. Mutanen, B. Schuller, P. Järventausta, and W. Su, “Classification of electricity customer groups towards individualized price scheme design,” in *Proc. of North American Power Symposium*, Morgantown, WV, Sep. 2017.

[136] G. Le Ray, P. Pinson, and E. M. Larsen, “Data-driven demand response characterization and quantification,” in *Proc. of IEEE Manchester PowerTech*, Manchester, Jun. 2017.

[137] S. Ikeda and H. Nishi, “Sparse-coding-based household clustering for demand response services,” in *Proc. of IEEE 25th International Symposium on Industrial Electronics (ISIE)*, Santa Clara, CA, Jun. 2016.

[138] C.-N. Yu, P. Mirowski, and T. K. Ho, “A sparse coding approach to household electricity demand forecasting in smart grids,” *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 738–748, Mar. 2017.

[139] Irish Social Science Data Archive. (2012), “CER smart metering project,” [Online]. Available: http://www.ucd.ie/issda/data/commissionforenergyregulationcer/.

[140] P. Senin, “Dynamic time warping algorithm review,” Information and Computer Science Department, University of Hawaii at Manoa Honolulu, Tech. Rep., Dec. 2008.

[141] I. Gurrutxaga, I. Albisua, O. Arbelaitz, J. I. Martín, J. Muguerza, J. M. Pérez, and I. Perona, "SEP/COP: An efficient method to find the best partition in hierarchical clustering based on a new cluster validity index," *Pattern Recognition*, vol. 43, no. 10, pp. 3364–3373, Apr. 2010.

[142] A. Sarda-Espinosa, "Time series clustering along with optimizations for the dynamic time warping distance," *Comprehensive R Archive Network (CRAN)*, Aug. 2017.

[143] P. U. Herath, V. Fusco, M. N. Cáceres, G. K. Venayagamoorthy, S. Squartini, F. Piazza, and J. M. Corchado, "Computational intelligence-based demand response management in a microgrid," *IEEE Transactions on Industry Applications*, vol. 55, no. 1, pp. 732–740, Jan. 2019.

[144] J. Kwac, J. I. Kim, and R. Rajagopal, "Efficient customer selection process for various DR objectives," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1501–1508, Mar. 2019.

[145] T. K. Wijaya, M. Vasirani, and K. Aberer, "When bias matters: An economic assessment of demand response baselines for residential customers," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1755–1763, Jul. 2014.

[146] "PJM empirical analysis of demand response baseline methods," KEMA, Inc, Tech. Rep., 2011.

[147] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang, "Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings," *Applied Energy*, vol. 195, pp. 659–670, Mar. 2017.

[148] S. Park, S. Ryu, Y. Choi, J. Kim, and H. Kim, "Data-driven baseline estimation of residential buildings for demand response," *Energies*, vol. 8, no. 9, pp. 10 239–10 259, 2015.

[149] K. Li, B. Wang, Z. Wang, F. Wang, Z. Mi, and Z. Zhen, "A baseline load estimation approach for residential customer based on load pattern clustering," *Energy Procedia*, vol. 142, pp. 2042–2049, Aug. 2017.

[150] L. Hatton, P. Charpentier, and E. Matzner-Løber, "Statistical estimation of the residential baseline," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1752–1759, May 2016.

[151] Y. Weng and R. Rajagopal, "Probabilistic baseline estimation via gaussian process," in *Proc. of IEEE Power & Energy Society General Meeting*. Denver, CO: IEEE, Jul. 2015.

[152] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network," in *Proc. of 51st Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, Mar. 2017.

[153] 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, Jan. 2019.

[154] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi, "Long short-term memory networks for accurate state-of-charge estimation of li-ion batteries," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6730–6739, Aug. 2018.

[155] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," *arXiv preprint arXiv:1506.00019*, 2015.

[156] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[157] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 9–48.

[158] 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.

[159] J. Jazaeri, T. Alpcan, R. Gordon, M. Brandao, T. Hoban, and C. Seeling, "Baseline methodologies for small scale residential demand response," in *Proc. of IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia)*. IEEE, Nov. 2016, pp. 747–752.

[160] T. Song, Y. Li, and X.-P. Zhang, "A cluster-based baseline load calculation approach for individual industrial and commercial customer," *Energies*, vol. 12, no. 64, Dec. 2019.

[161] M. Sun, Y. Wang, F. Teng, Y. Ye, G. Strbac, and C. Kang, "Clustering-based residential baseline estimation: A probabilistic perspective," *IEEE Transactions on Smart Grid*, 2019.

[162] Y. Weng, J. Yu, and R. Rajagopal, "Clustering-based residential baseline estimation: A probabilistic perspective," *Electrical Power and Energy Systems*, vol. 100, pp. 508–516, Mar. 2018.

[163] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and information systems*, vol. 7, no. 3, pp. 358–386, 2005.

[164] T. M. Rath and R. Manmatha, "Lower-bounding of dynamic time warping distances for multivariate time series," *University of Massachusetts Amherst Technical Report MM*, vol. 40, 2002.

[165] F. Wang, K. Li, C. Liu, Z. Mi, M. Shafie-khah, and J. P. Catalão, "Synchronous pattern matching principle based residential demand response baseline estimation: Mechanism analysis and approach description," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6972–6985, Nov. 2018.