An Informed Long-term Forecasting Method for Electrical Distribution Network Operators

AKPERI, BRIAN,TEMISAN

An Informed Long-term Forecasting Method for Electrical Distribution Network Operators

Brian Temisan Akperi

A Thesis presented for the degree of
Doctor of Philosophy

School of Engineering and Computing Sciences
University of Durham
England

March 2017

Dedicated to

Alero Akperi and Toritse Akperi

An Informed Long-term Forecasting Method for Electrical Distribution Network Operators

Brian Akperi

Submitted for the degree of Doctor of Philosophy
March 2017

Abstract

Northern Powergrid (NPG) is an electrical distribution network operator in the UK servicing Yorkshire and the Northeast of England. Currently they produce long-term eight year forecasts for each substation on the network with an emphasis on an annual maximum demand (MD) figure. The current method used by NPG is thought to oversimplify the problem and does not give enough insight into changes in substation demand. In order to inform their current forecast, the novel CL-ANFIS method uses a combination of machine learning techniques for both forecasting and general insight to the drivers of demand. Also introduced here are novel techniques for determination of MD at NPG and methods for handling load transfer periods.

In order to address a problem of this size, a twofold approach is taken. One is to address the drivers of demand such as weather, economic or demographic data sets through the use of statistics and machine learning techniques. The other is to address the long-term forecasting problem with a transparent technique that can aid in explaining the drivers of demand on any given substation. Techniques used include cluster analysis on demographic data sets in addition to ANFIS as a forecasting method. The results of the novel CL-ANFIS method are compared against the current NPG forecast and show how more insight into substation demand profiles can drive the decision-making process. This is done through a combination of using a tailored customer database for NPG and leveraging the information provided by the membership functions of ANFIS.

Publications

Conference Papers

1. **B. Akperi** and P. Matthews. Analysis of Clustering Techniques on Load Profiles for Electrical Distribution. 2014 IEEE International Conference on Power System Technology (POWERCON), 2014a.

2. **B. Akperi** and P. Matthews. Analysis of Customer Profiles on an Electrical Distribution Network. 2014 49th International Universities Power Engineering Conference (UPEC), 2014b.

Declaration

The work in this thesis is based on research carried out at Durham University, the School of Engineering and Computing Sciences, England. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Acknowledgements

I would like to thank the following people who have helped me achieve the results presented in this thesis:

- My supervisor Dr. Peter Matthews for allowing me to work on an interesting project and providing his support and guidance.
- The sponsor Northern Powergrid for putting forth this interesting research project and in particular Alan Creighton and Suninder Deagon for their help.
- Dr. Douglas Halliday and all the members of the multidisciplinary CDT in energy.
- My good friend Guanmei (Mona) Wang who gave me the crazy idea of doing a PhD in the first place.
- Riccardo Iacopetta Shirres for doing some good work in implementing methods presented here for NPG use.
- The support of my mother Eveline Pellenkoft and father Mejuya Akperi whom I love dearly.
- Alero and Toritse for giving me the motivation to carry on.

Acronyms

ACF	Auto Correlation Function
ACS	Average Cold Spell Correction
ANFIS	Adaptive Neuro-based approach to Fuzzy Inference Systems
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BEIS	Department for Business, Energy and Industrial Strategy
BIS	Department for Business, Innovation and Skills
CHP	Combined Heat and Power
CL-ANFIS	Customer-Led ANFIS
CLNR	Customer-Led Network Revolution
CRI	Composite Risk Index
DECC	Department of Energy and Climate Change
DLE	Distribution Load Estimate
DNO	Distribution Network Operator
DT	Decision Tree
EHV	Extra High Voltage
EMU	European Monetary Union
FIS	Fuzzy Inference System
GA	Genetic Algorithm

GENEFER	Genetic Neural Fuzzy Explorer
GSP	Grid Supply Point
HW	Holt Winters exponential smoothing
KNN	K Nearest Neighbour
LSE	Least Squares Estimate
LV	Low Voltage
MAE	Mean Absolute Error
MAP	Maximum a Posteriori estimation
MAPE	Mean Absolute Percentage Error
MD	Maximum Demand
MIDAS	Met Office Integrated Data Archive System
MSE	Mean Squared Error
NB	Naïve Bayes
NEM	National Electricity Market (of Australia)
NG	National Grid
NPG	Northern Powergrid
OAC	Output Area Classification
Ofgem	Office of Gas and Electricity Markets
ONS	Office of National Statistics
PCA	Principal Component Analysis
PDF	Probability Density Function
RBFNN	Radial Basis Function Neural Network
RIPPER	Repated Incremental Pruning to Produce Error Reduction
RMS	Root Mean Square
RMSE	Root Mean Squared Error
RQ	Research Question
SCADA	Supervisory Control and Data Acquisition

SVM Support Vector Machine

TAIEX Taiwan Capitalization Weighted Stock Index

TPC Taiwan Power Company

WPD Western Power Distribution

Nomenclature

ACC	Accuracy
$AR(p)$	Autoregressive model of order p
C	Matrix of cluster centres for fuzzy C-means (Section 3.2.7)
CV_k	K-fold cross validation error
FN	False Negative
FP	False Positive
FPR	False Positive Rate
F_t	Holt Winters time series (Section 6.2.3)
H_0	Null hypothesis
I	Current
$I(\cdot)$	Indicator function
K	Number of K closest neighbours (Section 3.2.2)
K	Number of K clusters (Section 3.2.6)
M	Non-adjusted MD (Section 2.3.2)
$MA(q)$	Moving average model of order q
MCC	Matthew's Correlation Coefficient
M_{acs}	ACS MD (Section 2.3.2)
$O^1_{A,i}$	Membership function for a fuzzy set A_i
P	Real Power
$P(\cdot)$	Probability of event
Q	Reactive Power
$RSS(\cdot,\cdot)$	Residual sum of squares
R^2	Coefficient of determination

S	Apparent Power
S	Xie-Beni index (Section 6.2.1)
$S(A, B)$	Similarity measure between fuzzy sets A and B (Section 6.2.2)
T	Output layer terms (Section 3.2.7)
T	Dunn index (Section 6.2.1)
TN	True Negative
TP	True Positive
TPR	True Positive Rate
T_{-1}	Noon temperature on day before MD day (Section 2.3.2)
T_{-2}	Noon temperature two days before MD day (Section 2.3.2)
T_0	Noon temperature on day of MD (Section 2.3.2)
T_c	Effective Temperature (Section 2.3.2)
V	Voltage
$V_j)$	Function space
X	Input layer terms (Section 3.2.7)
Z	Hidden layer terms (Section 3.2.7)
$\bar{R}$	Davies-Bouldin index (Section 4.3.1)
χ^2	Chi-squared test of independence
$\gamma(\cdot, \cdot)$	Covariance of two time series (Section 3.2.7)
μ	Mean of distribution
$\mu_{A_i}(x)$	Membership function for a fuzzy set A_i
ϕ	Haar father wavelet
ψ	Haar mother wavelet
$\rho(\cdot, \cdot)$	Autocorrelation function
σ	Standard deviation
$\sigma(x)$	Sigmoid function
$\mathbf{err}_{train}$	Training error
$\mathbf{var}(\cdot)$	Variance
a	Width of generalised bell membership function (Section 6.6)
a_i	Actual MD values

b_t	Holt Winters linear trend component
c	Centre of generalised bell membership function (Section 6.6)
c_t	Holt Winters seasonal correction factor
$d(\cdot,\cdot)$	Distance metric
f_i	Forecasted MD values
k	Kurtosis value (Section 5.1.1)
n	Sample size
p	p value (Section 4.1)
p_i	Probability for event i
$s(i)$	Silhouette value (Section 7.1)
$s.e.$	Standard error
s_t	Holt Winters base signal (Section 6.2.3)
t	t score (Section 4.1)
x_t	Time series

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Northern Powergrid (NPG) is an electrical distribution network operator (DNO) which services many types of customers throughout Yorkshire and the North East of England. Electrical demand data is collected using SCADA (Supervisory Control and Data Acquisition) facilities at frequent intervals of every 30 minutes. This information is summarised in the DLEs (Distribution Load Estimates) documentation created by NPG but there remains underlying uncertainties which are not captured. Some of the uncertainties involved are included in the automated judgment of determining the annual maximum demand (MD). Even once the data is collected, its interpretation also raises uncertainties. For example, the impact of new demand or generation connections and when they would have an impact on the system. In the future, it will be important to not only have an accurate forecast of demand but also an informed forecast.

This thesis will create an informed forecast methodology using a combination of statistical and machine learning techniques to enhance the forecasting methodology by NPG.

1.1 Challenges Facing Northern Powergrid

NPG has provided the challenges and uncertainties they have with the SCADA data. Their primary difficulties are in establishing the MD figure which is used as the basis for demand forecasts in their current method.

1.1.1 Uncertainties With the Raw SCADA Data

Firstly, there are uncertainties regarding issues with the raw SCADA data including the following:

- The selection of MD for input on the DLEs is a key uncertainty. This assessment is done through an automated method but involves a lot of engineering judgment as well. This is examined in greater detail in Chapter 2.
- These are also temporary demand transfers where load is shifted from one substation to another in order to manage faults and network development. This is also examined in greater detail in Chapter 5.
- In the demand profile there are natural erratic demand peaks and generation troughs. Sometimes these peaks are ignored in the selection of MD but it is a subjective area where engineering judgment needs to be applied.
- There is natural variation in the demand itself with a variable accuracy of $\pm 3\%$.
- SCADA equipment can produce erroneous data which can cause spikes or missing data for various time durations. If there are faults with the SCADA equipment, this can also lead to erroneous data.

1.1.2 Uncertainty in Interpreting the Raw Data

Regarding the interpretation of the data, new information can play a large role in modifying the forecast.

- New demand connections could of course lead to increased network demand. Although the capacity and date of energisation are known for a new customer, the rate at which they draw additional demand is unknown. This could potentially take years before they cause effect on the forecast.
- New generation connections could cause a decrease in net demand. Similarly to new demand connections, it can be unknown at times when the reduction would materialise in the demand profile.

- There may be reduced demand or disconnections from existing customers. This would lead to a decrease in demand but it is not identifiable whether such demand reduction or disconnection is temporary or permanent.

- In the past Average Cold Spell Correction (ACS) was used to normalise network demand for temperature. The aim of ACS is to account for variations in demand caused by temperatures in the winter period. However, NPG does not currently use ACS as there were concerns that it introduced errors. Further details are given in Chapter 2.

- The diversity factor is the ratio of the MDs of the supply point substation against the aggregated MDs of the primary substations. It is possible for the MDs at the satellite substations to increase while the demand at the source substation decreases due to the reduced co-incidence of the demand profile peaks at the satellite substations.

- The power factor varies throughout the year but the one recorded in the DLEs is taken to be the power factor at the time the MD occurred.

1.1.3 Future Uncertainties

This thesis is aimed at understanding the historical demand. However, NPG has outlined some areas in which some uncertainties will need to be understood in order to be able to incorporate them into the forecasting process.

- Energy prices will drive energy usage. For example, consumers may take greater advantage of energy reduction initiatives and use equipment which is more efficient.

- The DNO may introduce time of use tariffs to reduce peak demands. There will be a need to understand how effective these tariffs are at reducing demand.

- The introduction of smart metering is also expected to have an impact on demand measurement and forecasting methodologies. More data should be available from customers which will move the methodologies away from a planning based approach towards a real time operational approach. Although there will still need to be a governance structure in place for reinforcement on the network.

Some of these uncertainties will be addressed in the proposed forecasting method such as when dealing with missing data from SCADA equipment. Other issues such as uptake of demand from a new customer or the impact of smart grids will not be considered as they are outside the scope of this work.

1.2 Current Challenges in Long-Term Demand Forecasting

The need to understand the historic demand data for NPG and other DNOs is influenced by the quickly changing landscape in energy usage and how energy demand data will be collected.

In particular, the Department of Energy and Climate Change (DECC) has detailed future energy pathways in the 2050 Pathways Analysis (Department of Energy and Climate Change, 2010). At the time of this writing, note that DECC was dissolved in July 2016 and was merged with the Department for Business, Innovation and Skills (BIS) to form the Department for Business, Energy and Industrial Strategy (BEIS). This analysis was done as a result of climate change and an initiative to reduce greenhouse gas emissions by at least 80% by 2050 in comparison to 1990 levels. In Fig. 1.1, the breakdown of pathways is detailed in terms of the supply and demand from each sector.

On the left of the bar chart is the reference pathway labelled “R”. This represents a scenario where there is no attempt to decarbonise and no new energy technologies are deployed. It is clear to see that the diversity of energy supply sources is non-existent in comparison to the pathways A-F. These all represent scenarios with various attitudes and uptake of certain energy technologies. For example, pathways “A” represents a substantial cross-sector effort to decarbonise while pathway “D” represents a scenario where there is very little uptake of renewable energy. For the

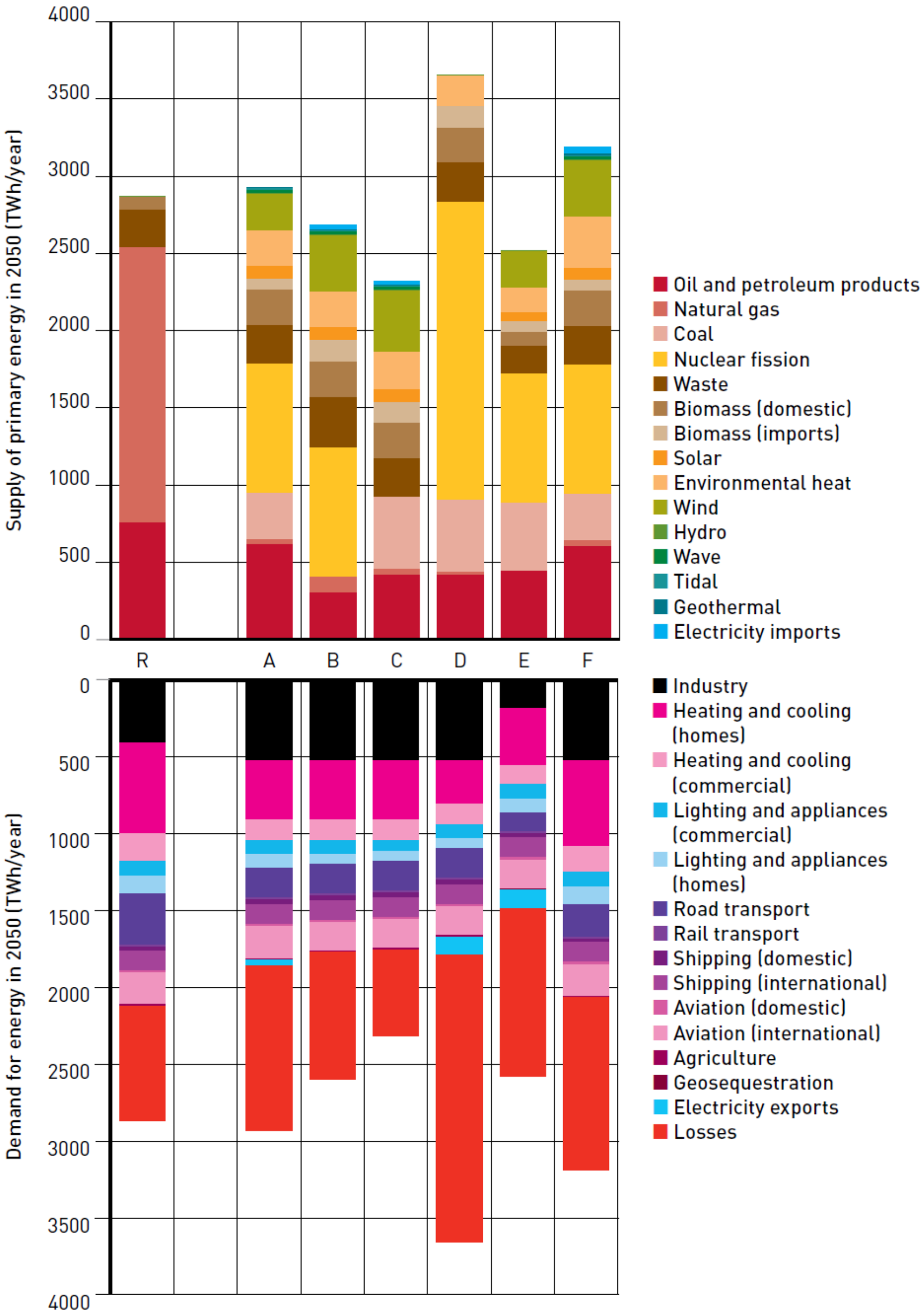


Figure 1.1: Projected Energy Demand and Supply in 2050 by DECC (Department of Energy and Climate Change, 2010)

base case of 2010 when this study was conducted, supply was comparable to reference pathway except with some proportion of nuclear energy. The demand in 2010 was also similar to the reference pathway but with a reduced demand for heating and cooling (200 TWh/year) (Department of Energy and Climate Change, 2010). The losses given in the supply bar chart are relative to the scenarios they are in. For example, pathway “D” is the low renewables scenario resulting in an increased uptake of nuclear energy which also leads to higher energy losses. Conversely, pathway “C” represents a scenario with little nuclear uptake resulting in much lower energy losses.

As there are many uncertainties about the type of technologies that will be used in the future, a scenario based approach like this is suitable for such a forward thinking analysis. There are many financial, political and socio-economical factors that contribute towards these pathways and considering each of them is complex.

So in general, understanding the drivers of national demand is extremely important because of the implications it will have on generation and emissions. NPG is one of the DNOs in the UK which covers the Yorkshire and Northeast regions of England as seen in Fig. 1.2.

For the DNOs, there are a variety of reasons for understanding historic demand and future forecasts. They have the requirement of meeting the contractual firm capacity at substations but they also are exploring other avenues of innovation. In order to address future energy problems, NPG had the Customer-Led Network Revolution (CLNR) project (completed in 2014) (Northern Powergrid, 2016) which sought to find better alternatives to dealing with new demands on the network rather than the traditional approach of reinforcing the network. They did this through a smart grid technology trial for each of their customer sectors (domestic, small and medium-sized enterprises, industrial & commercial and distributed generation). They concluded that domestic customers can be flexible with their network usage and that they contribute less to network peak demand than originally assumed. For industrial and commercial customers, they concluded that demand side response is a reliable option in order to address network constraints.

Through this project, NPG have been able to develop tools that benefit other DNOs such as (Northern Powergrid, 2016):

- Network Planning Design and Decision Support tool
- New policy and technical recommendations

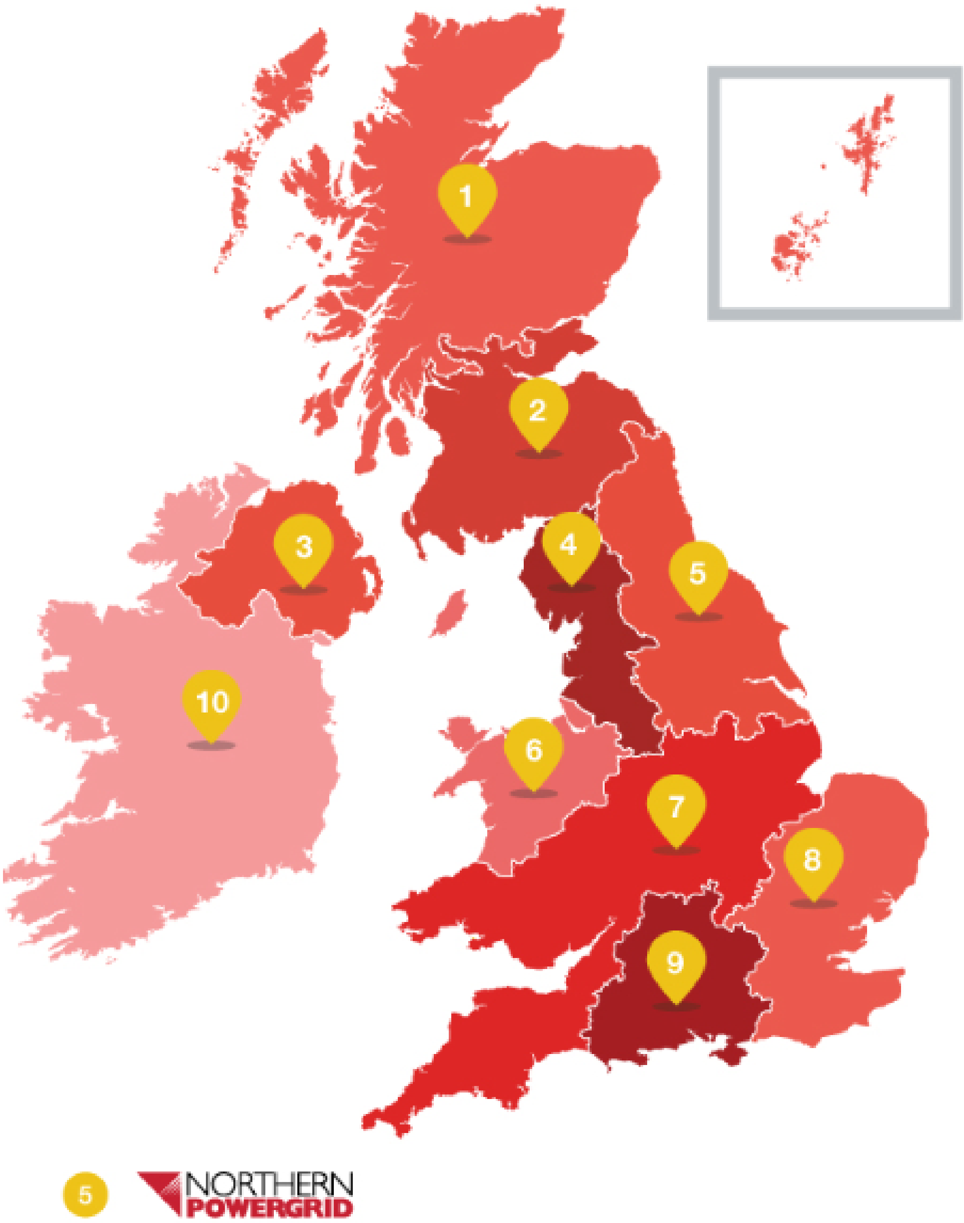


Figure 1.2: Map of the DNO Areas in the UK with NPG operating in Area 5 (Energy Networks Association, 2014)

- Lessons learned reports
- "How to" guidelines for equipment and training

The changing landscape in energy supply and demand is driving the research done in this thesis and in other related projects (National Grid, 2013; Northern

Powergrid, 2016; Western Power Distribution, 2013). In order to better meet the energy demands of the future, it is clear that a substantial amount of research must be undertaken to meet those needs.

1.3 Research Questions

Some key research questions can be developed based on the challenges NPG are facing and the current state of research in the electrical distribution forecasting area. The following key research questions (RQ) are laid out here which will be addressed in future areas of the thesis with a final discussion in the concluding Chapter 8.

RQ1: How can statistical and machine learning methods be used to analyse internal factors such as demand variation?

Justification: This is directly related to the challenges facing NPG currently including the uncertainties within the raw SCADA data in Section 1.1.1.

RQ2: How can statistical and machine learning methods be used to analyse external factors such as temperature and customer type?

Justification: This is one of the key drivers for NPG in terms of gaining a better understanding of their network. This has already influenced projects such as the CLNR and external factors are a natural extension for this work as well.

RQ3: How can more transparency be gained about current and future trends after analysing internal and external factors?

Justification: In particular, NPG requires methods where a qualitative explanation can be derived from the results rather than a pure black box technique.

RQ4: How can the forecast and insights gained from statistical analysis be used to enhance the business as usual for NPG?

Justification: In particular, NPG needs to be able to practically use the method in such a way that it can be implemented into their current architecture.

RQ5: How can an evaluation of a new method be conducted, given the multiple influencing factors which affect the load?

Justification: There needs to be a level of evaluation to justify the use of the new method by NPG and it must be conducted on every main component of the method.

1.4 Research Contributions

The purpose of the project is to not only develop a method which is capable of forecasting demand but also to develop a greater understanding of the underlying substation demand trends from data that is routinely collected from SCADA and other data systems. This is done with a desire to form a more informed and robust view of the required load related investment. Currently, NPG has a system which handles data directly from SCADA systems without any supplementary analysis input into the forecast. If there is a fault or unexpected load on the system, it is dealt with on a case-by-case basis. New connections and generation are considered but underlying drivers of electrical demand such as weather, economic conditions, and demographic information are not considered. Therefore, the goal is to have a more robust understanding of the most influencing factors on demand and using this information to supplement and enhance a forecasting tool.

This has been addressed in the original contribution CL-ANFIS (Customer-Led ANFIS). With this method, NPG have a more robust method where the generated forecast is supported by customer segmentation using machine learning techniques. This allows for a greater understanding of demand on substation-by-substation basis. They can implement these methods through a combination of the code produced in MATLAB and Excel spreadsheets such that it can run in a Windows environment. As of this writing, the methodology will be delivered to NPG for trial use.

In addition, a novel approach to NPG's MD selection method is presented in Chapter 5. This presents an automated method which removes the necessity for NPG engineers judgement in choosing a MD for the current year. In addition to the potential time saving and reduction in bias, this method also forms the basis for testing in Chapter 7. Chapter 5 also explores an approach to handling load transfers which can have an unwanted effect on the determination of maximum demand and forecasting in general.

1.5 Structure of Thesis

In this chapter, the motivation for the current research has been introduced and the contributions of the thesis have been outlined. The structure of the following chapters are as follows.

Chapter 2 (RQ1, RQ2) will explicitly describe the current NPG practice in their determination of MD, forecasting and historic research they have performed. In particular it will examine their documentation which describes their methods and preliminary research NPG has conducted. This will give further motivations for this research and background for the problem.

Chapter 3 (RQ1, RQ2, RQ3) is a more general literature review discussing the current techniques used in data mining and time series analysis with applications in multiple industries. The approach here is to present both the most commonly used methods in time series and machine learning analytics as well as direct applications of those tools. In this way, a wide variety of areas are covered for use in this thesis and future work.

Chapter 4 (RQ2) explores additional data sets that could be used in conjunction with the informed demand forecasting method. This includes temperature data from the Met Office, demographic information from publicly available sources and in-house demographics from NPG.

Chapter 5 (RQ1) details some of the methods used in cleaning the original SCADA data set and the automated MD selection method. This process is central prior to the use of ANFIS as any missing data or erratic spikes can have an unwanted effect on the forecast. It is also important to have an automated method of choosing MD such that a comparison against current NPG methods can be made.

Chapter 6 (RQ3, RQ4, RQ5) details the CL-ANFIS methodology with initial results following forecasting on available data. The bulk of the proposed method is reported here and in addition, there is an analysis of the customer types by their proposed quantitative labels.

Chapter 7 (RQ5) deploys further evaluation techniques to check the algorithm for robustness and compares the method to current NPG practices. The comparison to the NPG forecast is done in a similar way using MDs so that a direct comparison of forecasting aspect of CL-ANFIS can be made.

Chapter 8 concludes the thesis by discussing the findings and proposes further work and other avenues that could be explored.

Chapter 2

The Current Northern Powergrid Method

In order to address the forecasting problem, this chapter will explore the current methods used by NPG as well as older research they have conducted on their current procedures. The forecasting problem is a result of both underlying uncertainties such as the collection of raw load data and data interpretations. For example, new load and generation connections can cause changes in demand and thus a change in the forecast. In Section 2.1, a brief overview of NPG and its role as a distribution company is given. Section 2.2 explains some of the motivations for the creation of a Load Related Investment Plan by NPG. Section 2.3 explains the NPG method in greater detail including some of their own research on the effectiveness of their method. Section 2.4 concludes with the findings of the NPG method.

2.1 The Role of Northern Powergrid in Electricity Distribution

Energy distribution in Great Britain is conducted by one of the 14 DNOs which are owned by six companies. One of these companies and the sponsor of this research is NPG. The transmission of electricity is conducted solely by the National Grid. National Grid and the DNOs are subject to the rules and regulations set by industry regulator OFGEM. In addition to transmission and distribution, there are also the supply companies of electricity which pay the National Grid and DNOs for use of their assets to deliver energy to individual customers.

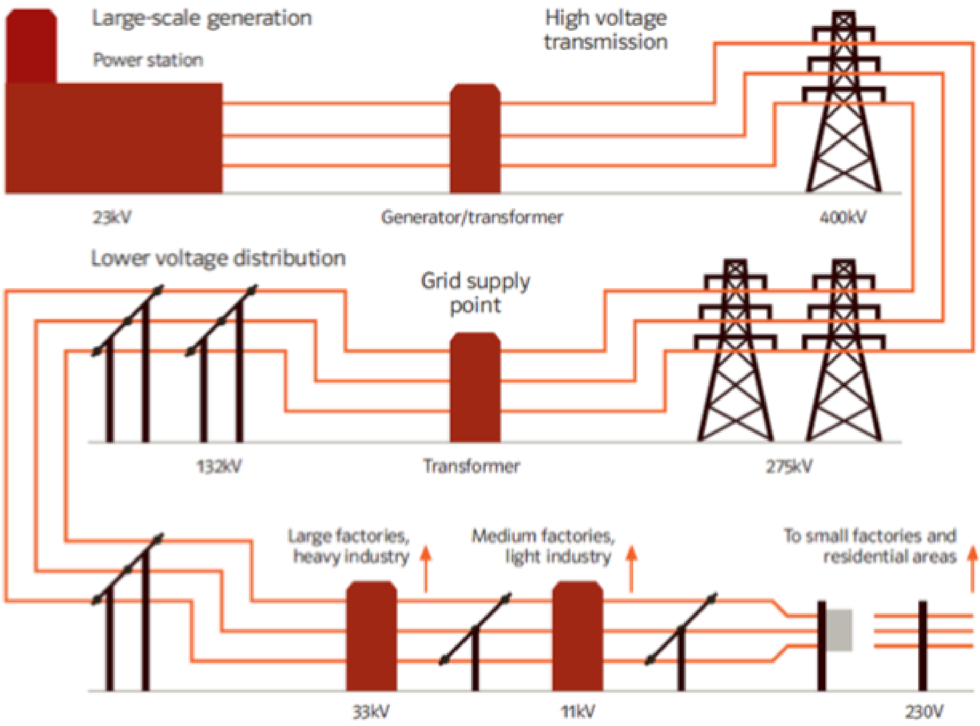


Figure 2.1: UK Electrical Distribution Network Diagram (EDW Technology, 2017)

For the purposes of this research, the focus will be on energy demand as viewed by DNOs. Fig. 2.1 shows a diagram of the UK distribution network. The National Grid transmits electricity at the 400kV and 275kV levels. These are then stepped down to the 132 kV sub-transmission level which is the first level operated and owned by the DNOs. DNOs own sites known as Grid Supply Points (GSPs) which are connected to the National Grid (Northern Powergrid, 2015). These GSPs typically operate at this 132 kV level and form the highest layer of the total DNO network.

The 132 kV sub-transmission level is then stepped down to 66/33 kV levels which are classified as extra high voltage (EHV). These are often large industrial customers which will have their own transformer.

The primary substations which are the main focus of this research operate at medium voltage levels of 20/11/6.6 kV. The most common are the 11 kV substations which are reached by stepping down from the 132 kV or EHV levels. Although distribution to individual residential users at the 240 V level is outside the scope of this research, the types of customer the primary substations serve is key in gaining a deeper understanding of load demand.

2.1.1 Definitions and Basic Overview of Northern Powergrid Method

NPG collects demand data from a SCADA (Supervisory Control and Data Acquisition) system which is collected in a data historian called PI (CE Electric UK, 2009; OSIsoft, 2017). This is done for all the substations they operate and is recorded in voltage-amperes (VA). This is due to measuring the apparent power $|S|$ on an alternating current power system as $|S| = |V||I| = \sqrt{P^2 + Q^2}$ where P is the real power and Q is the reactive power (IEEE, 2000). The complex power S in phasor form as a function of the apparent power $|S|$ is given as $S = |S|\angle\psi = P + jQ$ where ψ is the angle of difference between voltage and current. Although not an SI unit, apparent power is typically used as a power rating as it is simply the product of RMS (root-mean-square) voltage and RMS current (IEEE, 2000). Each substation will have a firm capacity associated to them which is the contractually agreed upon amount of load NPG must be able to supply to the customer. This figure is supplemented by a MD and a load forecast of the MD. The MD is the maximum recorded load at a substation subject to certain conditions (see Section 2.3.1) which is forecasted on over an eight year period. If the agreed upon firm capacity D is exceeded for that substation during the load forecast period, then some work on that part of the network must occur (see Section 2.2). That is, for any given substation with MD d_0 in the current DLE year and forecast MDs d_i for $i = 1, 2, ..., 8$, then $d_i < D$ must be satisfied for all $i = 0, 1, ..., 8$. Associated with this is the concept of network risk which can be defined as the frequency of interruptions to a customer at a site (Kaplan, 1990). The composite risk index specific to NPG is discussed in Section 2.2.

2.2 Load Related Drivers

Before detailing the DNO method and past experiments, it is important to consider NPG's drivers related in developing their Load Related Investment Plan. Namely, the distribution network must have the capacity to support customer load. There are general requirements such as the Engineering Recommendation P2/6 (Scott, 2007) and the Electricity, Safety, Quality and Continuity Regulations 2002 (Wilson, 2002). These requirements help ensure that electrical distribution equipment is operable and meeting the purpose it was created for. NPG also has their own internal review for assessing network risk called the Composite Risk Index. The CRI places emphasis on substations which are operating over firm capacity. Substations which are over firm capacity are said to be the highest risk but there is also a consideration of the degree to which it is over firm capacity and the length of time over firm capacity. In addition there is a consideration of historic load demand. In particular, MD figures are used to determine if a rapid rate of growth indicates a high risk substation. Creighton (2011) also notes some areas where the CRI can be improved. Namely that there should be a consideration of the geographic distribution with a high CRI index. This would be useful in the instance of an outage where a load transfer may not be able to take place due to the unavailability of adjacent substations. Also there should be a consideration of the sensitivity of the current factors to determine if they should be weighted equally in the risk index. The key areas of interest for CRI factors according to Creighton (2011) are the derivation of MD, firm capacity, time over firm capacity and load forecasts.

The selection of MD is based on certain criteria and is not the absolute maximum. More stringent criteria for the selection of MD will increase the MD and subsequently increase the risk associated with the substation.

The selection of the firm capacity value is key in determining substation reinforcement. The firm capacity is dependent upon the rating of the equipment supplying the substation. However, in the cases of interconnected rings, the circuit may be the limiting factor so there needs to be a better understanding of the firm capacity taking into account interconnected ring circuits for related substations.

The number of consecutive half hour periods over firm capacity can also be a factor in the CRI. Namely, the cyclic ratings of transformers can be affected. One way to deal with the over firm capacity substations is to perform short-term load transfers with low voltage transfer capacity for those substations that exceed the firm capacity by a significant amount.

Creighton identifies load forecasts as a factor for change in network risk. Therefore, this project will have a direct impact on how risk is currently assessed. Note that Creighton considers a scenario based forecast over a 20 year period as one possible approach. For the purposes of the CRI, a broad approach such as this may be preferable although the DLEs consider an 8 year forecast.

Finally in Creighton (2011), there is a consideration of the intervention methods when reinforcement is required. Some of the traditional methods include:

- transformer replacements,
- change of network topology, and
- permanent/temporary transfer of load.

In the future, there will be a greater use of distributed generation in order to maintain a constant system performance. This is due to the increased uptake of renewables and technologies such as combined heat and power (CHP). In addition, smart techniques such as active network management can incorporate customer response to reduce demand as well as managing network assets (CE Electric UK, 2009). After understanding some of the motivations, a more detailed look at NPG's load estimate process can be discussed.

2.3 Northern Powergrid Methodology

The primary reference for this section will be CE Electric UK (2009) as it is the main source which explicity describes the NPG forecasting method. The load estimate process is seen as the first stage of the integrated "end to end planning" process (CE Electric UK, 2009). The overall process takes input load data, produces observations and an 8 year ahead MD forecast in the DLEs. A first analysis of the data will identify any issues such as substations operating over firm capacity and solutions are proposed.

The DLEs and the overall process helps in compliance with certain license requirements. Every DNO must publish a long-term development statement under condition 25 of the distribution license (Gas and Electricity Markets Authority, 2017). In addition, the Grid Code (National Grid Electricity Transmission plc, 2017) requires the DNO to produce demand data at specific times for all grid supply points (GSPs) and demand forecasts for the next 7 years.

In section 2.3.1, the determination of MD is discussed. Section 2.3.2 examines how ACS was implemented in the past. Section 2.3.3 dicusses underlying growth rates in the NPG forecast process. Section 2.3.4 discusses adding large load changes due to new connections. Section 2.3.5 explains the documentation known as the DLE. Section 2.3.6 detail the problems considering by NPG in the load estimate process.

2.3.1 Determination of Maximum Demand

The current NPG method is based around the determination and forecasting of the MD figure. The MD figure is defined as (CE Electric UK, 2005):

> The highest value of demand recorded in any half hour at the metered point of connection with the NPG network under normal system operation and configuration conditions.

After the half hourly demand is extracted from the PI system, the MD is determined based on the load duration curve. An example for a particular substation is given in Fig. 2.2. The load duration curve shows the number of half hourly periods $C(i)$ which fall above a particular load value.

Two variables are considered when choosing the MD. Let y be the absolute maximum, $r \in [0, 1]$ be a factor scaling the absolute MD to a new MD value, and $h \in [ry, y]$ be a half hour demand period exceeding the percentage load ry. Then

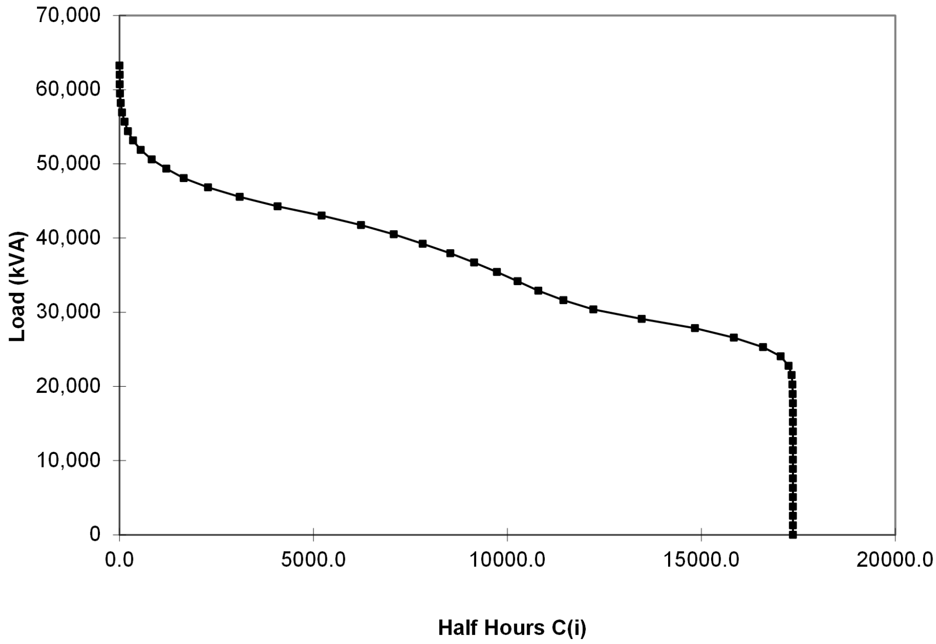
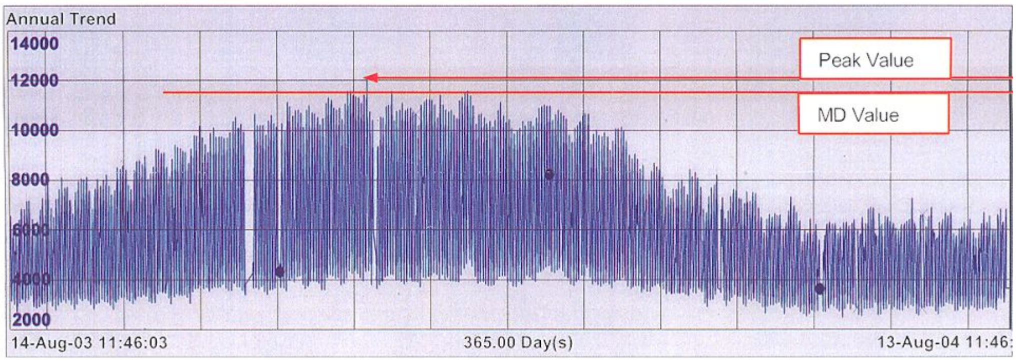


Figure 2.2: Sample Load Duration Curve

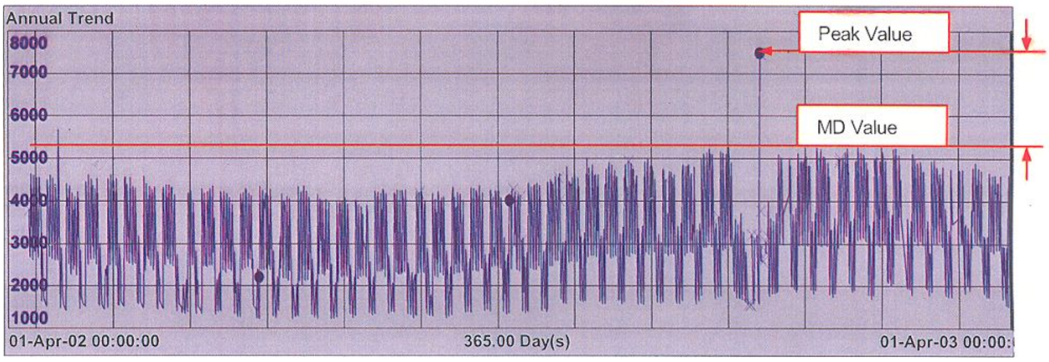
the MD is:

$$\min\{ry \text{ when 15 total } h \text{ values occur}, \quad ry \text{ when 15 consecutive } h \text{ values occur}\} \tag{2.3.1}$$

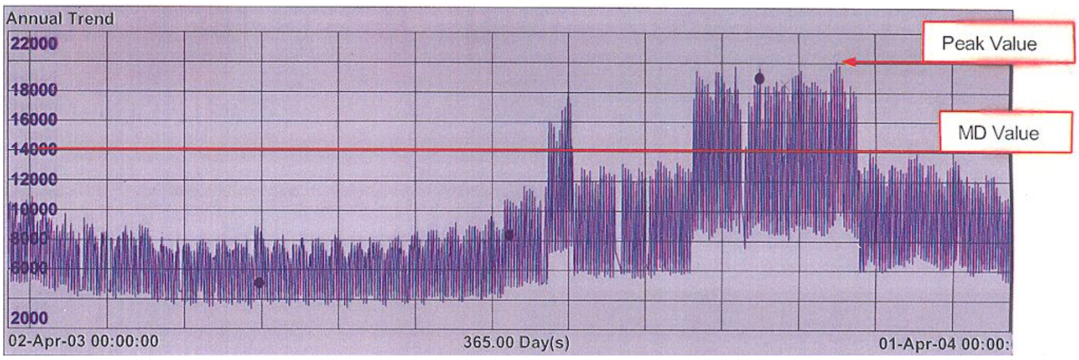
This is done to attain a reasonable MD value according to NPG based on other studies they have completed. The value of 15 was determined to be the most fit for purpose in internal studies at NPG. If the profile does not feature any discrepancies as in Fig. 2.3a then the MD value will be close to the actual peak. However, if there are any large erroneous spikes such in Fig. 2.3b, they are not considered in the MD determination. An exception to these cases occurs in Fig. 2.3c when there is a temporary load transfer for a period of a few weeks. This load transfer is also disregarded in the MD determination. In this instance, the load transfer does not accurately reflect the typical business as usual for that substation so any future planning should not take this into account. Once the MD is determined, the forecasting process begins.



(a) Case I - Stable Profile



(b) Case II - Mostly Stable Profile with Erroneous Spikes



(c) Case III - Profile with a Load Transfer

Figure 2.3: Choice of MD in Various Scenarios (CE Electric UK, 2009)

2.3.2 Average Cold Spell Correction

ACS is an attempt to use temperature data to account for demand variation during the winter. ACS was used at NPG in the past but is no longer currently used. However, it is still important to understand how ACS was implemented in order to observe any impact temperature data had on the NPG method.

The temperature data is obtained from the Met Office at climate data logger stations. Therefore temperature data is not local to individual substations. The base winter temperature for ACS is $-1.5°C$. Maximum demand is then increased or decreased dependent on the temperature difference away from $-1.5°C$. The ACS correction equation is (CE Electric UK, 2009):

$$M_{acs} = M + (0.7M(T_e + 1.5)) \tag{2.3.2}$$

$$T_e = 0.43T_0 + 0.33T_{-1} + 0.24T_{-2} \tag{2.3.3}$$

where M_{acs} is the ACS MD, M is the non-adjusted MD, T_e is the effective temperature, T_0 is the noon temperature on the day of MD, T_{-1} is the noon temperature on the day before MD and T_{-2} is the noon temperature two days before MD. As the temperature data is not local to each substation, there can be some variation in local temperatures versus the actual Met Office temperature data. This will result in variance of $\pm 0.7\%$ per $°C$ difference up to maximum of 7.5%. Although there is no documentation for the coefficients values used here CE Electric UK (2009), based on the values of the coefficients for T_e we can see that more recent days are weighted more heavily.

There have been instances in the DLEs where ACS was applied when the MD occurred in the summer even though it should only be applied for winter MDs. This resulted in a maximal 7.5% increased MD change for 106 substations in 2004. (CE Electric UK, 2009) theorises that MDs in summer months are a result of changes in energy usage such as:

- domestic use of gas for heat rather than electricity,
- reduction in initiatives to promote electricity usage through tariffs, and
- rise of air conditioning usage in commercial buildings.

It is noted that this change in energy usage, climate change and the overall impact of temperature on load data will have an impact on the uncertainty of ACS. Further work is discussed in Chapter 4 to determine the impact of temperature data on NPG substations.

2.3.3 Load Forecasting for Underlying Growth

After the data has been extracted from PI and stored, growth rates are applied to the MD to produce the forecast. In the past, growth percentage forecasts were produced by using the following 5 sets of data (CE Electric UK, 2009):

1. Actual historic data
2. Actual historic consumption data
3. Cambridge Econometrics Industry Forecasts (Electricity, Gas, Water and Manufacturing) (Cambridge Econometrics, 2017)
4. Cambridge Econometrics Regional Forecasts (North East and Yorkshire and Humberside) (Cambridge Econometrics, 2017)
5. NG (National Grid) Seven Year Statement

This was found to be more effective at a high grid supply point level rather than at a primary substation level because of the way in which unit forecast data was split into five baskets of High, Low, Urban, Rural and EHV. The definitions with their associated load growth were originally the following (CE Electric UK, 2009):

- High - Substations situated in development areas receiving large numbers of enquiries and in significant large load blocks. 5% growth per annum
- Low - Substations situated in smaller planned development areas such as retail parks etc. 2% growth per annum
- Urban - Substations situated geographically in urban/suburban surroundings not subject to any specific local development. 0.5% growth per annum
- Rural - Substations situated geographically in rural surroundings not subject to any specific local development. 0.5% growth per annum
- EHV - EHV Customer owned substations, not affected by local development. 0% growth per annum

However, the use of these five metrics “is not particularly scientific or documented” (CE Electric UK, 2009) and the “overall percentage growths are determined by using simple judgement without any calculations as such carried out.” (CE Electric UK, 2009) The documentation seems to be highly focused on gathering accurate forecasting results with little consideration for how external metrics data could impact overall trends.

Table 2.1: DLE Growth Rates for Substation Categories

YEAR	HIGH	LOW	URBAN	RURAL	EHV
13/14	2.00	1.00	0.50	0.50	0.00

In the DLEs, the updated table of underlying growth percentage is presented as in Table 2.1. Although in actuality, the rates in Table 2.1 are not currently used. According to the code of practice (CE Electric UK, 2005), the overall demand growth varies little year to year and is usually in the order of 1%. Currently NPG only uses two underlying growth percentages: 0% for extra high voltage customers and 0.5% for all other primaries. While this choice is of small significance in the immediate 1-2 year future, these choices of underlying growth are more inaccurate for the 8 year forecasting horizon.

EA Technology (EA Technology, 2017) carried out an exercise to determine how effective NPG’s forecasting method was in comparison to just using a naïve method of having the forecast for a year ahead be the actual value from the past year (CE Electric UK, 2009). In this assessment, only one year was tested despite a seven year ahead forecast being provided. This was due to the idea that the accuracy of the one year ahead forecast would affect future forecast years causing compounding errors. EA Technology used the previous years’ forecasted MD to make a comparison to the actual recorded MD value. The mean absolute percentage error (MAPE) was used as a uniform metric which would allow for comparison of a time series with small values against one with large values. It is given as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{f_i - a_i}{a_i}\right| \tag{2.3.4}$$

where n is the total number of forecasts, f_i are the forecasted MD values and a_i are the actual MD values. As a benchmark, a naïve forecast was used which assumes the forecasted MD is the same as the actual MD from the previous year. The initial results on the sample substations are given in Table 2.2 and Table 2.3.

Table 2.2: MAPE for One Year Ahead Forecasts (CE Electric UK, 2009)

Base Year	00/01		01/02		02/03	
Forecast Year	01/02		02/03		03/04	
	DLE	Naïve	DLE	Naïve	DLE	Naïve
Northeast					8.49%	6.53%
Yorkshire	8.56%	7.94%	8.98%	8.58%		

Table 2.3: MAPE for Two Year Ahead Forecasts (CE Electric UK, 2009)

Base Year	00/01		01/02	
Forecast Year	02/03		03/04	
	DLE	Naïve	DLE	Naïve
Yorkshire	12.62%	12.75%	11.33%	10.99%

These results show that either the naïve approach performed better or there was a negligible difference between the methods. It is also noted that the usage of MAPE lead to discrepancies in terms of what is an acceptable error. The examples given in CE Electric UK (2009) are as follows:

Example 1	Example 2
Actual Demand = 3 MVA	Actual Demand = 30 MVA
Forecast Demand = 3.7 MVA	Forecast Demand = 37 MVA
Percentage Error = 23.33%	Percentage Error = 23.33%

Even though both examples have the same percentage error, only example 1 is considered an acceptable error while example 2 is not. This is due to the absolute difference in Example 2 being much larger than that of Example 1.

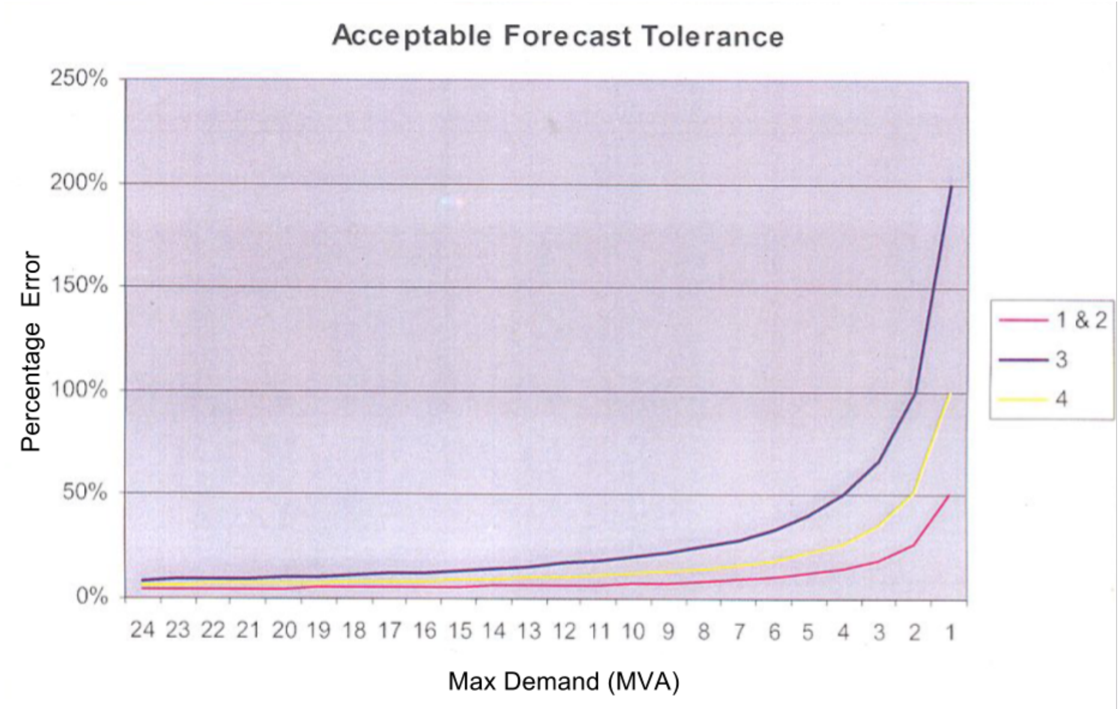


Figure 2.4: Acceptable Forecast Tolerance Opinions

Fig. 2.4 (CE Electric UK, 2009) shows acceptable forecasting errors dependant on the MD based on four engineers' opinions. Although there is some difference of opinion, it is clear that significantly higher forecast errors are acceptable at low MD values.

2.3.4 Forecasting Large Load Changes

In addition to the underlying forecast growth, the forecast demand is supplemented by large load changes of 1 MVA or above. Large loads are predicted to connect within the first year of the forecast. The information regarding new connections is not easy to establish so there are no hard rules where a new large load is certain to go ahead. There are many problems with the current information received from the connections process. NPG applies the load changes directly to the substation demand forecasts for the year they are due to come in.

An informal process of asking design engineers to detail any envisaged load transfers is carried out when forecasting large load transfers. In the process of determining MD, load transfers are often considered manually rather than using any defined ruleset. A data cleansing process of shifting load transfers should be used so they do not affect the forecasting tool. A historical look at the frequency of load transfers on particular substations could be investigated.

The generation data in the company is difficult to follow (CE Electric UK, 2009) because the demand used in the DLEs is the net result of the gross load connected to a substation. The incorporation of generation is treated similarly to the addition of new loads, the main difference being that only the net demand load is recorded. Since new generation could have an affect on forecasted load, they should be accounted for in some way but similar to new loads, only connections made in the short term are considered in the DLEs.

2.3.5 Distribution Load Estimate and Firm Capacity

The final stage involves recording the forecasted information in a user-friendly spreadsheet. The substations are grouped by their supply point substation. Each primary substation has the previous years actual demand along with a forecast for the next 8 years. Also recorded is the power factor and any connected generation at the primary substation. The supply point forecasts are created by aggregating all of the primary substations and applying a diversity factor calculated from the historical year. At the bottom of the sheet, existing generation is recorded along with any future large load or generation connections. The new large load changes are added directly to the 8 year forecasting period, usually within the first two years.

Alongside each substation there is also a rating of the firm capacity. The firm capacity is typically defined as being the "amount of load that can be secured following the loss of the largest single item in that system." (CE Electric UK, 2009) If load exceeds this firm capacity rating, then there is deemed to be a need for reinforcement at that particular substation. Most substations will have at least two transformers so defining the firm capacity to be one of those transformers is acceptable. For single transformer primary substations, firm capacity is routinely set at 8 MVA and design engineers confirm this is a known problem as this does not always accurately represent the firm capacity of the substation (CE Electric UK, 2009).

2.3.6 Overall Load Estimate Process Problems

In the concluding sections of CE Electric UK (2009), some of the problems with the overall load estimate process were highlighted. With regards to the determination of MD values, some of the issues were discussed. Issues with the raw data are mentioned such as missing data and spikes, which can have a direct effect on MD determination. Regardless, it is stated that 35% to 50% of substations must be manually checked in the load estimate process every year. However, the aim is to always have some level of engineering judgment rather than a fully automated process. The engineering judgment is affected by the number of spikes in the profile and their duration. In their own studies, NPG found there is a correlation between the number of spikes beyond the perceived MD envelope and engineering judgment. This has led to the most recent definition of MD (CE Electric UK, 2009):

> The highest value of demand recorded in any half hour under normal network feeding arrangements and taking into account of the impact of short duration high load values over a 12 month period.

This correlates to the new method in Eqn. 2.3.1. Also discussed by CE Electric UK (2009) were the underlying growth rates. Econometric growth data is not available for site specific areas on a substation by substation basis. Growth rates for larger areas are available but they will mask regional trends. The issue that although overall system MD does not change greatly (in the order of 1-2%), primary substation MD can vary greatly up to as much as $\pm 30\%$. Therefore, the focus should be on demand growth in the locality of each primary substation.

One suggested method would be to use new load enquiries as a basis for load growth. The relationship between new applications and demand growth could then be seen but this would require some new IT development in the current process. Another method would be to use substation growth categories as seen in Table 2.1. This would help to differentiate between substations with a flat load growth and those that are likelier to see change in load. Ultimately, these growth rates were not used but the idea of categorising substations in this manner is seen throughout the literature. The growth rates were not used because they did not accurately capture the true variation at each substation.

2.4 Conclusion

This chapter has focused on documenting current and historical NPG practices for their overall forecasting method. In relation to the overall research questions for this thesis, it specifically relates to RQ1 and RQ2. That is, both internal and external factors have been considered by NPG and there are specific methods associated with both types. For example, their MD selection method as a consideration for an internal factor and Average Cold Spell correction for temperature data.

However, the purpose of this research is to not only develop a method which is capable of forecasting demand load but also to develop a greater understanding of the underlying substation demand trends from SCADA systems, with a view to forming a more informed and robust view of the required load related investment. Currently, NPG uses the PI system which handles data directly from SCADA systems without any supplementary analysis input into the forecast. If there is a fault or unexpected load on the system, it is dealt with on a case by case basis. New connections and generation are considered but underlying drivers of electrical demand such as weather, economic conditions, and demographic information are not considered. Therefore, the goal is to have a more robust understanding of the most influencing factors on demand and using this information to supplement and enhance a forecasting tool.

Chapter 3

Techniques for Load Analysis

This initial literature review is conducted in the aim of discovering the current techniques and methods used in forecasting and trend analysis. This is done in accordance to satisfying both academic and commercial responsibilities with the goal of exploring techniques currently available and building upon them. This literature review sets out to follow a systematic approach suggested by Fink (2005) that can be reproduced in the future. In Section 3.1, the literature search method is detailed including the research questions. In Section 3.2, general techniques in machine learning will be explored. Section 3.3 will go over some examples of studies used in various industries and Section 3.4 will conclude and summarise the findings.

3.1 General Techniques in Machine Learning

Regardless of sector, much of the literature in data analysis uses common techniques in data mining and machine learning. Therefore it is important to introduce these ideas before looking at more specific examples. In this section, both supervised and unsupervised learning methods are introduced.

3.1.1 Supervised Learning

Supervised learning involves finding a function to describe a dataset with some input/output pair (x_i, y) where x_i are input features and y is the output feature to be learned. The main purpose of supervised learning will be to check if there is some clear relationship between datasets from NPG and any external data sources. For example, supervised learning can be used to establish if there is a connection between demographic data and substation demand.

Decision Trees

The decision tree is a data mining algorithm which gives a graphical representation of data classification. A data set is characterised by its attributes and class. Attributes are labels used to describe the characteristics of the data while the class is a special attribute that is used to identify the data point. The method operates by choosing the most appropriate attribute at each node of the tree based on various metrics. One such metric is information gain. Information gain is a measurement calculated from entropy which is a measure of uncertainty of a random variable. That is,

$$\text{entropy}(p_1, ..., p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{3.1.1}$$

where p_i is the probability of the event occurring (Hall et al., 2011). Entropy calculations often use base 2 by convention and are measured in bits. It can intuitively be thought of a measurement of uncertainty because a high probability suggests low entropy and a low probability suggests high entropy. Along each branch, the algorithm will choose the attribute with the lowest entropy because it will result in the most correct classifications.

Decision trees provide a good intuitive visual representation of the data while also giving conditional statements along the branches. They are relatively quick to produce and do not require any domain knowledge or parameter setting. They are able to handle multidimensional data and are currently used in many industries including financial analysis, molecular biology, and medicine (Han et al., 2011).

Decision trees also have certain limitations. If proper precautions are not taken, they are prone to having high variance and can overfit to the training data. Small changes to the data or initial node can cause different trees to be constructed. In order to avoid this, methods such as tree pruning can be used. Tree pruning will cause the decision tree to be less accurate on the training data but more accurate on any unseen test data. Also, since decision trees handle one attribute at a time, this can lead to more inaccuracies. Two attributes that are almost equal in their discriminatory ability can lead to two different decision trees depending on which is used as the first node. However, taking the better attribute initially may not necessarily lead to a decision tree which is overall more accurate (Anderson, 2013).

Some popular applications of decision tree in the literature include:

- Friedl and Brodley (1997) used decision tree classifiers to classify remote sensing data sets for land cover classes such as forests, tundra and grasslands. It was found to be a useful tool for the data analysts as they are intuitive and useful for examining hierarchical relationships between the input data. The decision tree was also insensitive to noise and no assumptions are needed about the distribution of the data.

- Vlahou et al. (2003) used decision tree classification for diagnosis of ovarian cancer based on protein mass spectrum profiles. It was found to be 80% accurate when using five protein peaks although they state further preprocessing of the data could improve accuracy.

- Yu et al. (2010) used a decision tree method to estimate building energy performance indices. The results showed a 92% accuracy on the test data and was able to rank the significant factors of building energy usage. They also noted the interpretability of the method over regression techniques and neural networks.

K-Nearest Neighbours

The K-nearest neighbour classification algorithm is a simple method which classifies a data point based on the majority class of the closest K neighbours. That is, let K be a positive integer and let $\mathcal{N}_0$ be the K points which are closest to an observation x_0. Then, KNN estimates the conditional probability for a class j in Y as the fraction of points in $\mathcal{N}_0$ whose response values equal j (James et al., 2014).

$$P(Y = j|X = x_0) = \frac{1}{K} \sum_{i \in \mathcal{N}_0} I(y_i = j) \tag{3.1.2}$$

Here, $I(y_i = j)$ is the indicator function which is 1 when $y_i = j$ and 0 otherwise. KNN will assign x_0 to the class with the largest probability. Optimising for the best choice of K involves tuning the decision boundary to perform well on training data but also well on the testing data so as to avoid overfitting. A low choice of K could potentially lead to overfitting while a high choice of K could lead to a poor classifier.

Some examples of the KNN classifier being used in the literature include:

- Zhang et al. (2006) used a hybrid SVM-KNN classifier for image recognition. KNN is effective at dealing with the multiclass nature of image recognition. A disadvantage is that KNN suffers from high variance in the bias-variance tradeoff. Thus this hybrid approach using another method support vector machines (SVM) (Andrew Ng, 2017) was taken and Zhang et al. (2006) found that it dealt well with this multiclass problem.

- Li et al. (2001) used a hybrid GA-KNN approach to identify genes which correlate to tumor presence. They tested this hybrid approach using colon cancer and leukemia data. Li et al. (2001) found that the combination of genetic algorithm (GA) (Mathworks, 2017) and KNN was capable of choosing predictive genes amongst noisy data.

Classification Rules

Classification rules attempt to classify data points by logical conditions. The type of classification rule algorithm described here is RIPPER (repeated incremental pruning to produce error reduction) (Hall et al., 2011). Initially, the data set is split into two sets, a growing set and a pruning set in a 2:1 ratio. The growing set is used to build an error free rule set. This is done by examining one class at a time and using a covering algorithm to build perfect rules for each class. The problem with using this algorithm alone is that it might overfit the data by creating rules that only match the training set but do not deal well with unseen data. Therefore, the pruning set is used to decide if rules can be cut back from the growing set.

Although decision trees and classification rules are similar, one advantage of the

RIPPER classification rules is that once the first rule classifies a certain portion of the data, it is disregarded for the remainder of the algorithm so there are no rules with overlapping data points. This makes rules succinct and clear which may not be the case for decision trees that all stem from a single node.

This method cannot be used as a forecaster of energy since it is a supervised learning method but it can be used as a classifier for other data sources. The rules are easily interpretable and offer more flexibility in comparison to decision trees.

A few applications of classification rules in the literature include:

- Cohen (1996) uses variations of classification rule algorithms including RIPPER to filter and file e-mails. It was found that only a few hundred training data points were needed for a steep learning curve and acceptable error rate. That is, the introduction of a few data points reduced the error rate quickly.

- Basu et al. (1998) used an implementation of the RIPPER classification algorithm on a data set of movie recommendations. A priority for them was to use all the available information for more accurate and interpretable results which RIPPER provided.

Bayesian Probability

Naïve Bayes is a statistical modelling technique used for classification that is governed by Bayes rule of conditional probability: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ for two events A and B. The naïvety in the name is a result of the assumption that all events are independent of one another. The independence assumption and Bayes rule leads to (scikit-learn developers, 2016):

$$P(y|x_1, ..., x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1, ..., x_n)} \tag{3.1.3}$$

Since $P(x_1, ..., x_n)$ is a constant, the classification rule becomes

$$P(y|x_1, ..., x_n) \propto P(y)\prod_{i=1}^{n} P(x_i|y) \Rightarrow \tag{3.1.4}$$

$$\hat{y} = \arg\max_{y} P(y)\prod_{i=1}^{n} P(x_i|y) \tag{3.1.5}$$

Maximum a posteriori (MAP) estimation can then be used to determine $P(y)$ and $P(x_i|y)$.

Although there is an assumption of independence of events, the classifier can perform well on various training sets. Because of the independence assumption, each feature distribution can be modelled as a one dimensional distribution so it can handle high dimensional data sets well (Zhang, 2004).

Some limitations of naïve Bayes classification include the treatment of continuous features and binning them for classification purposes. Also in the event of class imbalance, the resulting probabilities can become skewed (Rennie et al., 2003).

Some applications of Bayesian methods include:

- López et al. (2009) used a combination of principal component analysis (PCA) (See Section 4.3.3 for further details) and Bayesian classifiers to diagnose Alzheimer's disease from tomography images of brains. They found that using this methodology resulted in 88.6% and 98.3% accuracy for two sets of imaging methods. It showed a significant improvement of over using the entire set of features before PCA.

- Yerima et al. (2013) used a Bayesian classification method to detect malware on Android devices. They classified applications as benign or suspicious using a set of features where defining characteristics of malware appears. They found that by training their Bayesian classifier with both Android malware samples and benign applications, their method outperformed popular signature-based antivirus software.

3.1.2 Unsupervised Learning

Unsupervised learning attempts to find some pattern in a set of inputs x_i in the absence of an output feature. A popular set of unsupervised learning techniques are clustering algorithms. By using clustering, data can be segmented in such a way that data points belonging to that segment have similar features. Because of the exploratory nature of this unsupervised method, clustering has been used in a wide array of industries such as document clustering (Andrews and Fox, 2007) or in medical imaging (Masulli and Schenone, 1999). An overview of some of the clustering methods in Akperi and Matthews (2014a) are presented in the following subsections. Also described here are the concept of association rules as another type of unsupervised learning.

K-Means Clustering

K-means clustering (MacQueen, 1967) is one of the most popular clustering techniques used across various disciplines on a wide variety of data. The basic steps of the algorithm are as follows:

1. Choose number of K clusters.
2. Assign data points to a cluster centre based on a distance metric.
3. Calculate the mean of each cluster group which becomes the new centre.
4. Repeat 2-3 until all data points are assigned to the same cluster.

Although this is a simple method that is widely used, there are classical problems associated with this method. One of the disadvantages is that the initial random choice of cluster centres can often cause very different clusters to form. A popular technique to address this as suggested in Xu and Wunsch (2005) and Hall et al. (2011) is to run the algorithm several times and choose the solution with the lowest sum of squared distance between the data and cluster centroids. Furthermore, there is also the issue of sensitivity to outliers as all points are forced into clusters. This will be considered when analysing the shapes of each cluster. Finally there is always the issue of the choice of number of K clusters which can be addressed with validity metrics such as Dunn's index (Dunn, 1974) or Davies-Bouldin criteria (Davies and Bouldin, 1979).

Hierarchical Clustering

The hierarchical clustering method is based on a tree structure known as the dendrogram (Xu and Wunsch, 2005). This can be done in a top down approach known as the divisive method which starts at a single cluster and performs binary splits until all clusters only have one member. However, this method is computationally intensive and not often used. More commonly used is the agglomerative method which is a bottom up approach built starting from single member clusters and combining clusters until there is only one cluster at the root node. The algorithm is as follows:

1. Start with N clusters where N is the number of data points.
2. Combine clusters based on a linkage method starting from the clusters which are closest together.
3. Add the newly formed cluster to the distance matrix.
4. Repeat 2-3 until there is only one cluster containing all elements.

The choice of linkage method for hierarchical clustering will also have to be considered. In Everitt et al. (2001) some of the most common linkage methods are summarised. The average linkage method is seen as being one of the most robust methods and is the average distance between all pairs of data points where one comes from each group. This is in contrast to simpler methods known as single linkage (nearest neighbour) and complete linkage (furthest neighbour) where distance is simply calculated by the nearest or furthest points in each group. An additional linkage criteria known as Ward's method was also considered by Chicco et al. (2006) but it was found that average linkage was better at rejecting dissimilar load profiles whilst Ward attempted to find groups of the same size. A choice of number of clusters must be made here by choosing a threshold horizontal division on the dendrogram.

Fuzzy C-means Clustering

In this type of clustering, data points do not have to belong to a single cluster but instead have degrees of membership in $[0, 1]$ that denote the extent to which a point is similar to that cluster centre. Otherwise, the procedure is quite similar to K-means with the following steps (Dunn, 1974):

1. Choose number of K clusters and initialise random centre points.

2. Update the membership matrix U by $u_{ij} = \left(\left(\sum_{v=1}^{K} \frac{d(x_i,c_j)}{d(x_i,c_v)}\right)^{\frac{1}{m-1}}\right)^{-1}$, where $u_{ij} \in U$ is the fuzzy membership matrix, $d(\cdot,\cdot)$ is a chosen distance metric such as Euclidean distance, $x_i \in X$ is the matrix of load profiles, c_j is the cluster centre and $m > 1$ is the fuzzification parameter.

3. The matrix of cluster centres $C = (c_i)$ is then updated
$c_i = \left(\sum_{j=1}^{N}(u_{ij})^m x_j\right)\left(\sum_{j=1}^{N}(u_{ij})^m\right)$, $i = 1, ..., K$

4. Repeat 2-3 until the matrix of centres stabilises.

The disadvantages of using fuzzy C-means are similar to using K-means in that there is no definitive method to identify the initial partitions and that the method is sensitive to outliers. The uniqueness of fuzzy membership can offer more insight than crisp clustering because it can help to show the uniqueness (or non-uniqueness) of the cluster centres and the similarity of load profiles to the centres.

Association Rules

Association rules are logical rulesets based on attribute values of a dataset and can give information about trends in that dataset. Association rules are similar to classification rules in that they both use attribute values to give a ruleset about data. However, association rules are more general because they do not necessarily focus on one class at a time and the algorithm is an unsupervised learning method. Some terminology used in association rule discussion is the notion of *item sets* which are attribute-value pairs that have a prespecified minimum *coverage*. The coverage (also known as support) is the number of instances the association rules predict correctly. The coverage as a ratio to the total number of instances is known as the *confidence*.

The Apriori algorithm is an implementation of association rules and (Hall et al., 2011) is useful for finding rules on large datasets. First, all one item sets with a user specified minimum support are generated and from these, the two item sets are generated and so forth. This is done using the fact that the union of constituent one item sets must be of minimum coverage in order for the two item set to be of the minimum coverage. Similarly, the union of the three two item sets must all be of minimum coverage for the three item set to be of minimum coverage. The candidate $(n + 1)$ item sets are checked by passing through a hash table to check if the constituent n item-sets are in the table. Then a check is made to count coverage of the $(n + 1)$ item set and if it meets the minimum coverage it stays in the table. After the item sets are generated, the ruleset has to be created. This is done by generating rules from each item set and checking if the rule meets the prespecified confidence value and then adding it to the ruleset.

Association rules can give interesting information about the patterns in a data set. They do not necessarily focus on the classes and instead consider the patterns in the attributes in generality. The user defined parameters such as coverage and confidence can give great control over the output. The error given by the confidence is user specified so if general rules with a high support are sufficient then a lower confidence value can be used and vice versa.

The disadvantage of this algorithm is that it is not a classifier and error metrics are user specified so there is little analysis that can be carried out. They may give an idea about the important relationships in the data but are not useful enough for classification purposes.

Association rules could help to analyse external data sets in relation to energy load data. However, since they do not offer any classification ability, it would be difficult to use this algorithm in conjunction with others.

In the literature, some applications of association rules are the following:

- Creighton and Hanash (2003) used association rules to extract knowledge from gene expression data. The results showed some associations which can be explained biologically. Other non-random patterns found could lead to new hypotheses and Creighton and Hanash (2003) state that association rules will be used in future analysis of biological systems.

- Brijs et al. (1999) created a model for product selection in a market basket analysis based on the frequency of item sets. By using data from a convenience store, Brijs et al. (1999) was able to identify the cross-selling effects of pairs of goods bought. This could lead to further profitability for a store through the use of product placement.

3.1.3 Methods for Time Series and Forecasting

The data mining techniques examined so far are applicable to exploring data sets via unsupervised methods or for classification purposes through supervised methods. This may be of use for external data sets where time is not a factor but for continuous data which may be nonlinear, other techniques will be considered. Thus, the next logical area to explore which will contribute to demand forecasting will be areas in data mining and machine learning used to forecast time series. Some of the most popular techniques in literature include traditional time series methods such as the Autoregressive Moving Average (ARMA) and more recent advancements in machine learning including neural networks. Some preliminary discussion is done here to explore the theory behind these techniques.

Autoregressive Moving Average

The autoregressive (AR) models are based on being able to predict a time series x_t from its past p values $x_{t-1}, x_{t-2}, x_{t-p}$.

The autoregressive model of order p, $AR(p)$ is defined as (Shumway and Stoffer, 2010):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t \tag{3.1.6}$$

where $\phi_1, ..., \phi_p$ are constants and w_t is white noise. They are so-called autoregressive because they use past values of the series to create the model.

An alternative to the autoregressive model is the moving average model of order q defined as (Shumway and Stoffer, 2010):

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + ... + \theta_q w_{t-q} \tag{3.1.7}$$

where there are q lags in the moving average and $\theta_1, ..., \theta_q$ are coefficient parameters. The $MA(q)$ model would be used instead of the $AR(p)$ model by considering the autocorrelation function (ACF) (Shumway and Stoffer, 2010):

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} \tag{3.1.8}$$

where $\gamma(s,t) = \text{cov}(x_s, x_t)$ of two time series x_s, x_t.

The ACF for the $AR(p)$ will tail off gradually while the ACF for the $MA(q)$ will cut off after q time lags.

In many cases since the $AR(p)$ model is a function of past time values and the $MA(q)$ model is a function of lagged error terms, they are combined to form the autoregressive moving average model. The autoregressive moving average model $ARMA(p,q)$ of order p,q is a combination of the $AR(p)$ and $MA(q)$ models and is defined as (Shumway and Stoffer, 2010):

$$x_t = \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + ... + \theta_q w_{t-q} \tag{3.1.9}$$

Some applications of the ARMA family of techniques include:

- Cuaresma et al. (2004) use the ARMA family of models to forecast electricity spot prices. They found that by modelling the data hour-by-hour, the forecasting ability of the linear univariate time series greatly increased. If the price spikes are taken into consideration, then the forecasting can be even further improved. They noted that the models were able to reproduce in-sample dynamics better than out-of-sample forecasts. There was a trade-off in the simplicity of the linear model assumption and the computational complexity of generating the model.

- Torres et al. (2005) used ARMA models to forecast wind speeds up to 10 hours in advance. After the transformation and standardisation of the hourly wind speeds, they found that the use of ARMA models significantly improved the forecast accuracy in comparison to the persistence models which assumed a linear 0% growth. They found that for low wind speed less than 11 m/s, the root mean squared error was low and not affected by the length of the forecast. The largest errors were due to the variance in high wind speeds.

Artificial Neural Networks

Neural networks can be applied to numeric data to predict nonlinear behaviour. The concept of neural networks begins with the perceptron. In a basic two class case where the dataset is linearly separable, the perceptron learning algorithm works as follows (Friedman et al., 2001). The weight vector for the separating hyperplane $w_0a_0 + w_1a_1 + ... + w_ka_k = 0$ is set to zero. Then until all instances are classified correctly, for each instance in the training set, the attributes are added to the weight vector if they belong to the first class ($w_0a_0 + w_1a_1 + ... + w_ka_k > 0$) or subtracted if they belong to the second class ($w_0a_0 + w_1a_1 + ... + w_ka_k < 0$). This will move the hyperplane in the correct direction until all training instances are classified correctly, assuming they are linearly separable.

We can extend the idea of this simple algorithm to multilayer perceptrons, one of the first examples of neural networks. In general, neural networks consist of three layers. The *input layer* consists of attribute values each of which represent a node in the network. The layer above this is what is known as a *hidden layer* because the nodes are not directly observed. These represent the perceptrons mentioned earlier. Then finally there is the *output layer* which is represented as a linear combination of the hidden layer units.

The algorithm for learning the weights is called *backpropagation*. It uses the optimization algorithm *gradient descent*. Gradient descent operates by modifying weights based on the strength of each contribution to the final prediction. For classification, using the normal perceptron rule we have a step function that outputs either a 0 or 1 depending on what class the instance belongs to. Using gradient descent, derivatives are taken and since such a step function is non differentiable an alternative *sigmoid* function $\sigma(x) = \frac{1}{1+e^{-x}}$ is used. The error function $E = (y - \sigma(x))^2$ is minimised in the algorithm where y is the class label and $\sigma(x)$ is the sigmoid function. In gradient descent, the derivative of the error function with respect to the weights are taken. Then each of these derivatives are multiplied by a constant called the *learning rate* and is then subtracted from the current value of the weight. The algorithm stops when the change in parameter becomes too small. Note that this algorithm may only find a local minimum.

The neural network can be described by the following equations (Friedman et al., 2001):

$$Z_m = \sigma(\alpha_{0m} + \alpha_m^T X), \quad m = 1, ..., M, \tag{3.1.10}$$

$$T_k = \beta_{0k} + \beta_k^T Z, \quad k = 1, ..., K, \tag{3.1.11}$$

$$f_k(X) = g_k(T), \quad k = 1, ...K \tag{3.1.12}$$

where $Z = (Z_1, Z_2, ..., Z_m)$, and $T = (T_1, T_2, ..., T_k)$. The terms α_{0m} and β_{0k} are bias units which acts as the intercept of the perceptron.

The Z terms are the hidden layer which act as the perceptrons. The X terms form the input layer consisting of the attributes and the function $g_k(T)$ gives the output based on a linear combination of the input layer. In regression, usually $k = 1$ and $g(T)$ is just the identity. In classification, the *softmax* function $g_k(T) = \frac{e^{T_k}}{\sum_{l=1}^{K} e^{T_l}}$ is used.

Neural networks can be extended further with multiple hidden layers and various activation functions apart from the sigmoid function mentioned here. Neural networks have a wide application in prediction and are also used for classification problems. Some applications in the time series context include:

- Quan et al. (2014) used neural networks to construct prediction intervals for the short-term load forecasting of wind power. The purpose of the prediction interval construction is to quantify the uncertainty of point forecasts by giving associated probabilities and confidence intervals to points. They found that using their method produced higher quality prediction intervals than ARIMA, exponential smoothing and naïve models.

- Jammazi and Aloui (2012) used neural networks to perform oil price prediction. They used a hybrid model combining a wavelet decomposition and a multilayer back propagation neural network. By investigating the architecture of the neural network and different types of activation functions, they created robust tests of their method. They found that their hybrid method outperformed the conventional back propagation neural network.

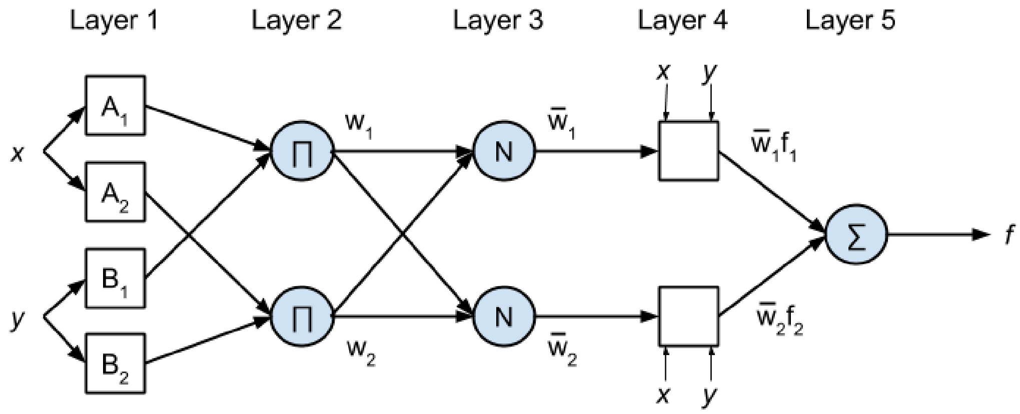


Figure 3.1: ANFIS Architecture

Adaptive Network based approach to Fuzzy Inference Systems

The adaptive network based approach to fuzzy inference system (ANFIS) developed by Jang (1993) also has the ability to model and predict non-linear data in a similar methodology to neural networks. A similar network structure is used to artificial neural networks with the ability to understand data through membership functions and fuzzy rules. The fuzzy inference system relies on the concept of fuzzy logic and fuzzy sets. A fuzzy set is a pair (U, m) where U is a set and $m : U \to [0, 1]$. The function m is called a *membership function* which shows the degree to which an element belongs to U. This gives rise to the fuzzy if-then rules of the form: If x is in A_1 and y is in B_1 then $f_1 = p_1x + q_1y + r_1$. These are known as Takagi-Sugeno type rules (Takagi and Sugeno, 1985) where A_1 and B_1 are fuzzy sets and f_1 is a linear combination of the input variables using learned parameters $\{p_1, q_1, r_1\}$.

The architecture of ANFIS is in five main layers as shown in Fig. 3.1 (Jang, 1993). For illustrative purposes, the two input ANFIS is described here.

1. In the first layer, there is a node function $O^1_{A,i} = \mu_{A_i}(x)$ where x is the input to node i and A_i is the fuzzy set which can be described by qualitative labels (e.g. small, medium, large). This $O_{A,i}$ is called the membership function of A_i. There are different choices for the type of membership function but they are normally bell-shaped and lie in $[0, 1]$. A common generalised bell membership function is given by:

$$f(x, a, b, c) = \frac{1}{1 + |\frac{x-c}{a}|^{2b}} \tag{3.1.13}$$

The parameters $\{a, b, c\}$ are referred to as the premise parameters.

2. In the second layer, the product of these membership functions are obtained. That is, $w_i = \mu_{A_i}(x) \times \mu_{B_i}(y)\ i = 1, 2$.

3. In the third layer, the weights are normalised over the total sum of all the weights.

4. In the fourth layer, the inputs are combined linearly and multiplied by their associated normalised weights. That is

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i), i = 1, 2 \tag{3.1.14}$$

The parameters $\{p_i, q_i, r_i\}$ are referred to as the consequent parameters.

5. The final layer is a summation of these final values $\sum \bar{w}_i f_i$.

The hybrid learning rule in Jang (1993) and detailed in Sumathi and Paneerselvam (2010) is a combination of least squares estimate (LSE) and gradient descent. In the forward pass, the premise parameters are kept constant whilst the consequent parameters are calculated using LSE. The error measure for each training pair is computed in the forward pass. In the backwards pass, gradient descent is used as in artificial networks.

Some examples of ANFIS used in the time series literature include:

- Yun et al. (2008) used a combination of the radial basis function neural network (RBFNN) for forecasting electric demand and then using ANFIS to adjust the load forecasting based on changes to the real-time pricing. They showed that when adding ANFIS to the RBFNN to consider pricing, the forecasting error improved.

- Chang et al. (2011) used a hybrid AR and ANFIS model in order to forecast stock prices on the Taiwan stock exchange, TAIEX. The volatility in prices of the TAIEX is considered by the AR model and then refined by ANFIS which uses the fuzzy if-then rules for qualitative knowledge. They found that this combined approach outperformed other recent stock price prediction models.

3.1.4 Conclusion on General Techniques in Machine Learning

As part of this literature review we have considered a variety of methods including supervised, unsupervised and regressive techniques. The applications of the supervised and unsupervised techniques should prove useful in analysing external data sources such as demographics and customer information. In those cases, patterns need to be found that can be related to NPG substations. In terms of the methods for forecasting, ANFIS will be carried forward in later parts of the thesis as it is able to fulfil both RQ1 and RQ3 through transparency in its membership functions. This will be explained further in later parts of this thesis.

3.2 Industry Case Studies

In order to understand how the techniques discussed in Section 3.2 are applied, a few case studies will be discussed here. The industries of finance, weather and electrical distribution are chosen as they are some of the most relevant to the research objectives as stated in RQ2. Finance and weather are often seen as drivers of the change in electrical demand so analysis of techniques in those areas can be analogous to the techniques used in electrical distribution.

3.2.1 Finance

The financial literature in relation to time series forecasting is mainly concerned with the use of machine learning techniques such as neural networks and genetic algorithms. Artificial intelligence methods are seen as a way of improving and in some cases replacing a domain experts decision making. Zhang and Zhou (2004) describe a need for an automated approach to handle large amounts of incoming financial data to assist in decision making. They categorise data mining into four types: association rule mining, classification, clustering analysis and time series mining. Then they give an overview of existing data mining techniques such as neural networks, genetic algorithms, statistical inference (cluster analysis, factor analysis, time series analysis), rule induction and data visualization. The main note here is that depending on the data and the required output for the user, a different technique should be used.

Financial time series are described as random, noisy, non-linear and non-stationary. There is widespread use of neural networks. However, hundreds of new algorithms have been created in recent years to deal with time series data. Zhang and Zhou (2004) then go on to discuss some financial applications such as prediction of the stock market. They found that neural networks were better predictors than regression models. For example, taking lagged time inputs for the neural network at $x(t), x(t-1), ..., x(t-k)$ can be used to predict $x(t+1)$.

The key challenges listed by Zhang and Zhou (2004) include the choice of data mining methods, their parameters and the integration of multiple data mining techniques. Although the primary objective is to obtain a forecast of maximum demand for NPG, any time series forecasting technique including neural networks can potentially be applied to the half hourly load data.

Case Study of Linear Econometric Model Against Non-Linear Computational Model Using Fuzzy Neural Network

Kooths et al. (2004) state that inflation forecasts are highly significant for central banks and policy makers. Many forecasting methods have been applied to the area such as univariate or multiple regression models and artificial intelligence systems. Kooths et al. (2004) seeks to find whether the predictive power of dynamic single equation linear models outperform the computational approach using genetic fuzzy rules while taking into account the wide variety of economic inflation theories. The data set used is a quarterly time series of the inflation rate in the European Monetary Union (EMU) from the first quarter of 1980 to the final quarter of 2000 accounting for 80 total observations. It is noted that historical circumstances are not taken into account and since the data is artificially constructed, it may be biased. The EMU inflation rate defined here is the quarter-to-quarter change of an artificially constructed consumer price index using individual weights for different countries. At the time of this study, there was not sufficiently long time series available for the EMU so artificially constructed data was the most promising line of work. This paper is more focused on developing a technical discussion from this data set.

The econometric forecasting approach uses a autoregressive distributed-lag approach of the form:

$$y_t = \alpha_0 + \sum_{i=1}^{q_0} \alpha_{1,i} y_{t-i} + \sum_{i=0}^{q_1} \alpha_{2,i} x_{1,t-i} + \cdots + \sum_{i=0}^{q_n} \alpha_{n,i} x_{n,t-i} + \epsilon_t \tag{3.2.15}$$

where y is the dependent variable regressed on q_0 lags of its own and n explanatory variables x are regressed on q_1 to q_n lags. It was found that all of the single error correction term models performed well and had a significant R^2 value. However, there were also some statistical diagnostic tests that were carried out that suggested some non-linearity that can be better explained by a computational approach.

The genetic-neural fuzzy rule based method is used by Kooths et al. (2004) as a more computational approach. Unlike the econometric approach, no theoretical model is needed to link the inputs and outputs. Prior knowledge is used to pre-select the variables given as inputs but not the relationship between the inputs and outputs. GENEFER (GEnetic NEural Fuzzy explorER) first fuzzifies all the variables, then sets up a rule base by using an evolutionary algorithm. This works by combining fuzzy sets of variables that have the highest degree of membership for the crisp (unaltered) values. The rules are of the form:

> IF Input 1 is medium AND Input 2 is very large AND Input 3 is low
> THEN output is very low.

These rules with linguistic labels allow the system to be interpretable by the end user. It was found that the in-sample tests for forecasting had comparable results to the econometric approach. In-sample tests could be misleading so out-of-sample forecasts were computed as well. Measures such as root mean squared error (RMSE) and mean squared error (MSE) were used as well as a 2x2 confusion matrix to determine the ability of the model to correctly predict change of a variable in a positive or negative direction, regardless of magnitude.

A total of 19 models were tested in out-of-sample forecasts which included a mixture of econometric and computational models. For one-step ahead forecasts, only GENEFER models passed the χ^2 test for the confusion matrix suggesting that they can predict turning points more reliably. For the four-step ahead forecasts, GENEFER models on average had better results than the econometric approach indicating that they are more accurate for longer term forecasts. Similarly to the one-step ahead forecasts, some GENEFER models passed the χ^2 test based on the confusion matrix while none of the econometric models did. However, the best model was actually a GENEFER model using an error correcting term derived from econometric modelling, suggesting that a combination of the two techniques might be the best method moving forward.

Of particular interest to NPG and this research is the ability to provide interpretable results for the end user. The linguistic ruleset that a method like GENEFER is able to provide would enhance knowledge outside of the forecasting accuracy. It is also noted that one particular method does not need to be adhered to but a mixture of methods suitable for the data set can be used.

3.2.2 Weather

Much of the well cited academic literature in the weather area focuses on natural disasters and the impact of climate change rather than everyday weather observations. This sort of focused research is to be expected and the ability to model these circumstances might draw some parallels to load forecasting. Here is such a case study that considers model uncertainty in relation to climate data.

Case Study for Climate Uncertainty

Hawkins and Sutton (2009) open by saying that decision makers in a wide variety of organizations are seeking quantitative climate predictions. The paper asks what the scope is for narrowing the uncertainty of these predictions by further investigations in climate science. Uncertainty in climate predictions arise from three factors: internal variability of the climate system, model uncertainty and scenario uncertainty such as greenhouse gas emissions. The impact of these uncertainty factors varies with the forecast time scale. In the long term, many decades into the future, scenario uncertainty and model uncertainty dominate. In the short term (within the next two decades), model uncertainty and internal variability are the most important factors.

Fifteen different models predicting the global annual mean surface air temperature were used subject to three different scenarios from the Intergovernmental Panel on Climate Change regarding greenhouse gas emissions (Hawkins and Sutton, 2009). This gives a total of 45 predictions. Each prediction was fit using a fourth order polynomial with ordinary least squares.

The raw predictions X for each model m, scenario s and year t can be written as (Hawkins and Sutton, 2009):

$$X_{m,s,t} = x_{m,s,t} + i_{m,s} + \epsilon_{m,s,t} \tag{3.2.16}$$

where the reference temperature is denoted by i, the smooth fit is represented by x, and the residual (internal variability) is ϵ.

Each model is also weighted by W_m which represents the ability to predict the global mean warming in 2000 from the means of 1971-2000. The internal variability for each model is defined as the residuals from the fits (Hawkins and Sutton, 2009):

$$V = \sum_m W_m \text{var}_{s,t}(\epsilon_{m,s,t}) \tag{3.2.17}$$

The model uncertainty is estimated from the weighted variance:

$$M(t) = \frac{1}{N_S} \sum_S \text{var}_m^W(x_{m,s,t}) \tag{3.2.18}$$

where N_S is the number of scenarios. Finally, the scenario uncertainty is the variance of the weighted means of the three scenarios:

$$S(t) = \text{var}_s(\sum_m W_m x_{m,s,t}) \tag{3.2.19}$$

By separating the model uncertainty out in this way, it is assumed that the total variance is the sum of the three parts. It is now possible to see what the most important factors are regarding uncertainty and how much influence researchers have to control them.

There is scope for climate science to reduce model uncertainty and internal variability and this is important in regards to adaptation to a changing climate. For example, Hawkins and Sutton (2009) supposes if there is great uncertainty, building a sea wall needs tolerance for more extreme events. Therefore, reducing uncertainty is of great economic value.

3.2.3 Electrical Distribution

The literature related to electrical distribution is more purely academic as it is difficult to find the practices of specific distribution companies. However, the literature that can be examined includes sponsored research by electrical distribution companies as well as public innovation projects. By considering these types of research projects, some insight into their methodology can be discerned.

Knowledge-Based Short-Term Load Forecasting

Ho et al. (1990) looks at short term load forecasting using a knowledge-based expert system. This project was supported by Taiwan Power Company (TPC) in 1990. The paper states that traditional statistical techniques (exponential smoothing, multiple regression, Box-Jenkins, etc.) are good at short-term load forecasting on normal days. The paper also gives some insight into the method used at Taiwan Power Company, although it is possible that it has changed in recent years. They use a basic linear regression model for short term load forecasting where the 24 hourly loads for a day of the week (e.g. Mondays) are retrieved from the past seven weeks with the exception of holidays and other special days. Then they use a simple model $L(i) = L_{AV}(i)(1+K\Delta T(i))$ where $L(i)$ is the forecasted load for hour i, $L_{AV}(i)$ is the average load for hour i, K is a proportionality constant and $\Delta T(i)$ is the difference between the forecasted daily high/low temperature and the average daily high/low temperature. The low temperature is used instead of the high temperature during the early hours of the day. Using this short term load model, TPC forecasts 24 hours ahead with a mean absolute error of less than 3.5% on "normal" days. These "normal" days exclude holidays and extreme weather such as tropical storms.

As a short term load model, it is possible to see a forecast with such a low error being achieved. However, such a method could not be extrapolated to a longer period of time because those small errors would continue to compound on each other over months and years. On the other hand, because figures are reliable over this short term period, it could be possible to look into combining a short-term forecast with a long-term forecast.

The new proposed system to deal with the non-normal days is done by first classifying load profiles into day types. Some of the day types include weekdays (excluding holidays), Saturdays, Sundays, holidays and Chinese New Year. The index used to judge how similar two load profiles are to each other is done by an index, $D = [\sum_{i=1}^{24}(L_1(i) - L_2(i))^2/24]^{\frac{1}{2}}$. After the load profiles are classified into a type, the hourly loads must be forecasted. Since the proposed method normalises hourly loads, the daily peak and trough have to be forecasted. Again this is done with a linear regression model that relies on outdoor temperature as the dependent variable. When looking at the findings, it was clear that using human experts or the knowledge based expert system had comparable results. Both were better than a traditional Box-Jenkins statistical model.

Load Research and Load Estimation in Electricity Distribution

The output of Seppälä (1996) is a result of the Finnish load research project started in 1983 and attempts to address two problems which are the estimation of hourly customer load and distribution load estimation. Effectively, the purpose of this paper has similar intentions to the purpose of the NPG project although it leaves out some detail.

The Finnish load research project uses "simple form load models" since "most mathematical load models developed for forecasting purposes are so far too complicated to be directly applied to studies of distribution networks." This simple form load mode for the hourly load $\bar{P}(t)$ as a function of annual energy W_a is

$$\bar{P}(t) = L_c(m(t), d(t), h(t)) \cdot W_a \tag{3.2.20}$$

where $m(t)$, $d(t)$, and $h(t)$ are classifying functions for the month, day type, and hour at time t and L_c is an estimated parameter of the average of the hourly load consumption divided by the annual consumption of the customer.

The Finnish load research project classified customers itself and divided customers into types based on purpose and based on heating. The customers included residential, agriculture, industry and service. The heating types included direct electric heating, partly storage electric heating, full storage electric heating, dual heating and heat pumps.

The research also went through some of the historical problems the project faced which included temperature standardisation. Because heating is taken into account as a classification for customers, its correlation to outdoor temperature is also considered and this is modelled by a linear increase. Load management on electric storage heating leads to results which are randomly distributed. The most problematic area seemed to be linking the researched load models with the utility's customer data. According to Seppälä (1996), "For each customer, a load model is selected with special linking rules." This is verified by looking at the sum of the utility's total load and the sum of the load models. However, if the linking rules are not accurate and do not correctly associate a customer with the correct load model, it can lead to inaccurate results.

The author then goes on to discuss modelling customer load by statistical distribution functions, in particular the normal and lognormal distributions. Seppälä (1996) then considers a binomial process of individual electric appliances being turned on or off in a home and models a skew distribution based on a randomization of these appliances being on or off. This simulates the customer's load variation. On a domestic customer scale, this could prove to be useful but when taking into account industrial type customers it would be difficult to monitor and manage their systems. It also makes too many assumptions on what is actually a part of a customer's use. This sort of approach might be more useful when looking at smart metering research because the required data would be more readily available.

In the last section, Seppälä discusses distribution load estimation. Seppälä uses a weighted least squares estimation method to minimise the error between a customer's hourly annual energy and the hourly load estimate using linear algebra.

In the model itself, Finnish utilities do seem to take into account multiple customer types which seem to imply greater insight. This detail is not covered in the paper however. The load model discussed is kept as simple as possible based on multiplying annual energy by a factor based on time.

Western Power Distribution LV Network Templates

Western Power Distribution (2013) (WPD) has done a study to develop LV (Low Voltage) network templates by giving a view of power flows and voltage at a given substation to address the impact of low carbon technologies and distributed generation. They have done this in a three step process, a summary of which is given here.

The data used in this research consists of two sets: fixed data and monitored data. The fixed data includes information on 948 substation profiles including network topology characteristics and customer mixes. The monitored data for 824 substations are the recordings made at 10 minute intervals for voltage, current and real power delivered (kW). This is the data used later in the clustering stage. The monitored data then undergoes a sense checking process for data errors. For voltage on LV substations, this involves a check of a pregiven tolerance band of -6% to +10%. An upper limit of 2kW for demand can be established for domestic households but no lower limit due to large variation. Using this upper limit of demand, some sense checking on the current can be completed. For commercial and industrial customers, there needs to be engineering judgment involved. There are also issues related to the resolution of the current data. For the power delivered, a check with the power factors are made so that real power delivered is close to real power calculated.

At the data analysis stage, the goal is to find daily patterns from the monitored data through the use of clustering. Two classic clustering methods are considered: k-means and hierarchical. The agglomerative hierarchical method is chosen because it assumes no prior knowledge about a choice of clusters. In order to determine the criteria that merges two groups into a single cluster, different normalisation considerations were made but the chosen method was that the data is normalised by the maximum value of real power delivered for each day. This is further broken down into groups concerning just weekdays and weekends and different seasons. For substations where monitored data was not available, a classification method using multinomial logistic regression was used where fixed data attributes such as customer class numbers and transformer ratings assign substations to the previously found clusters.

To develop the LV network template, 10 clusters were chosen based on the decreasing benefit of increasing the number of clusters by looking at the group sum of squares. Then based on the customer profiles WPD has, each template can be given a qualitative description such as domestic dominance, or industrial and commercial dominance. WPD state that this project has developed templates that can estimate substation load profiles with 80% accuracy. This equips network planners with more information to make more informed decisions. The paper also states that they are implementing this methodology as a tool at WPD.

Scenario Based

For a more scenario-based approach, the National Grid (NG) have produced the UK Future Energy Scenarios document (National Grid, 2013). Similar to the Pathways to 2050 document produced by the Department of Energy and Climate Change (2013), they envisage scenarios based on governmental policy, consumer behaviour and uptake of new technologies. Hence, they detail two different scenarios: Gone Green and Slow Progression.

Gone Green is a scenario where all future environmental targets are reached, namely 15% of total energy from renewables by 2020, greenhouse gas emissions meeting carbon budgets by 2027 and 80% reduction in greenhouse gas emissions by 2050. Slow progression is, as its name states, a scenario where slower progress is made toward environmental goals. National Grid makes clear that these scenarios are not forecasts and do no assign any likelihood of these occurring.

Underpinning both of these scenarios are what NG refers to as axioms. These are statements assumed to be true and are constructed by members of NG and key stakeholders. For example in the Slow Progression scenario, it is assumed that the UK 2020 renewables target is missed and pressure for UK carbon targets to be abandoned grows. Conversely, in the Gone Green scenario, all targets are met and there is no change to EU and UK policy.

Interestingly, when looking at the total electricity demand for the UK, the Gone Green scenario depicts higher demand mainly due to favourable economic conditions for industrial consumers. This suggests that the economy is a key driver in these scenarios so it is understandable that NG receives so much input from its stakeholders when constructing their axioms.

3.3 Conclusion

There are a wide variety of techniques and methods discussed in data mining and the different application sectors but this initial review gives a starting point moving forward. The research questions addressed in this chapter was focused mainly on RQ1 and RQ2. That is, the choice of statistical and machine learning techniques on analysing internal and external factors must be considered in a literature review. In addition, RQ3 is also considered as using such techniques could create more transparency in the current NPG forecasting method.

From the financial literature, it was found that there is a need for automated tools to assist in the decision making process. Zhang and Zhou (2004) discussed many techniques used in artificial intelligence that could also have an impact on load forecasting. Kooths et al. (2004) found that using a fuzzy rule algorithm could provide results with interpretable rules. Appropriate methods need to be selected but there are a range of tools out there according to the financial literature.

For the weather sector, the literature was focused on trends and the impact of climate change. Hawkins and Sutton (2009) were concerned with looking at uncertainty in climate models and the impact research can have to minimise this uncertainty. They were mainly concerned with the financial implications but also for short-term forecasting, climate research can have an impact by reducing internal variability and model uncertainty.

The electrical distribution papers suggested that companies prefer using simple models and then supplementing this with specialist knowledge. For example, Ho et al. (1990) look at supplementing Taiwan Power Company's short term load model with day type definitions. This approach of specialist models provides superior results to a purely statistical model. Seppälä (1996) commented that mathematical load models are too complicated for forecasting but considering the time when this paper was written, there might be more powerful tools that can give more insight. The knowledge for this model comes from associating customer types with a load model. Although that specific method was not discussed, it does imply that these Finnish utilities also use a knowledge based approach rather than a purely statistical one. Other DNOs have innovation projects as well such as Western Power Distribution (2013) where they have classified customers using clustering techniques.

Across all sectors, the common theme was finding an appropriate model for the data using any specialist knowledge where appropriate. Then after finding the appropriate model, the uncertainties and errors associated with it should be analysed to see if any refinements can be made.

3.3.1 Supporting the Need for New Research

The fact that there is such a wide variety of literature in these areas suggests that there is no consensus method when forecasting or analysing trends in data. Currently, NPG does not supplement their forecasts with much extra information such as customer types or economic conditions. They rely on historical load information and engineering judgment. New research is required to find the best method for forecasting and analysing load profiles and determining a systematic way of taking outside knowledge into account. The literature review has given a good starting point of techniques to consider.

In particular, electrical distribution seemed to be focused on having as simple a model as possible. According to the financial literature, the benefits of artificial intelligence come not only from a more accurate forecast but a more informed view. More research is needed in this particular area so that intelligent data mining solutions can be deployed across the energy sector. There is a great focus on forecasting accuracy and not enough on the interpretation of the forecast and the reasoning behind it. This should be achievable with a data mining approach supplemented by knowledge from outside influencing factors.

Chapter 4

Data Mining for Informed Substation Trends

Before detailing the forecasting algorithm, it is important to consider the impact of external data sets on the forecasting method. In particular, there will be an analysis of temperature data, economic and customer based demographic data as these areas are traditionally seen as having the largest impact on electrical demand. Section 4.1 will consider temperature data from the Met Office and compare it to NPG demand using linear regression. Section 4.2 will consider the Office of National Statistic's Output Area Classifications using supervised learning algorithms. Section 4.3 analyses a demographic customer base specific to NPG and Section 4.4 will summarise the findings.

4.1 Temperature and Weather Considerations

As part of the research driver to understand the underlying demand trends, external data sources are considered to investigate their impact on demand. Weather data and specifically temperature data is commonly used in demand load research to help explain underlying trends. The common winter peak seen in many load profiles suggests that temperature affects demand in some cases but not every load profile follows the typical winter peak trend.

The temperature data by MIDAS (Met Office Integrated Data Archive System)

(NCAS British Atmospheric Data Centre, 2013) is available on an hourly scale. Thus it would be inappropriate to use classification rule or other data mining classifiers which do not deal well with large sets of numeric data. Instead, simple linear regression will be used to determine if there is correlation between temperature and load.

The model for simple linear regression is given by the linear function: $y = \beta_0 + \beta_1 x$. The model attempts to find the intercept β_0 and slope β_1 that best approximates the data. Of course, as this linear function is only a model, there will be errors between the plotted data and the estimated linear function known as *residuals*. Thus the model of the data points y_i against x_i is $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where ϵ_i are the residuals for each x_i against y_i.

The method used for obtaining the parameters is known as ordinary least squares. The goal of this is to minimise the sum of the squared residuals (Weisberg, 2005)

$$RSS(\beta_0, \beta_1) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \tag{4.1.1}$$

It can be proven using calculus that the least squares estimates are given by $\beta_1 = r_{xy} \frac{SYY}{SXX}^2$ and $\beta_0 = \bar{y} - \beta_1 \bar{x}$ where r_{xy} is the sample correlation coefficient, $SYY = \sum (y_i - \bar{y})^2$ and $SXX = \sum (x_i - \bar{x})^2$.

After the linear regression line is found, some diagnostic tests will be used to determine if there is a correlation and how strong that correlation is. There are various tests that can be deployed to check whether there is a linear correlation in the data. Here, the student's t-distribution is used as the load data is a sampling of the entire normally distributed data set due to the mismatch in data set size between the load data and temperature data. Using the student's t-distribution, it is possible to reject or accept the null hypothesis of a correlation. The null hypothesis for the slope of the regression line is given by:

$$H_0 : \beta_1 = 0 \tag{4.1.2}$$

If there is a correlation, this hypothesis would be rejected as the slope is significantly different from 0. The t score is calculated as $\frac{\beta_1}{s.e.(\beta_1)}$ where the standard error is $s.e. = \sqrt{\frac{\sigma^2}{SXX}}$ where σ^2 is the estimated variance of the sample obtained from the residuals (Weisberg, 2005). Then using the t score and the degrees of freedom of the sample $n-2$ where n is the number of observations, the student's t-distribution can be used to obtain a p value which gives a measure of the credibility of the null hypothesis. A low p value rejects the null hypothesis while a high p value does not reject it.

The coefficient of determination R^2 is a measure of the degree of fit and is defined as $R^2 = 1 - \frac{RSS}{SYY}$ where $RSS = SYY - \frac{(SXY)^2}{SXX}$.

4.1.1 Experiment

The experiment presented here determines if there is a correlation between air temperature provided by MIDAS and electric demand provided by NPG. This is done by using linear regression and calculating the p value of the regression by way of the t-test and the coefficient of determination R^2.

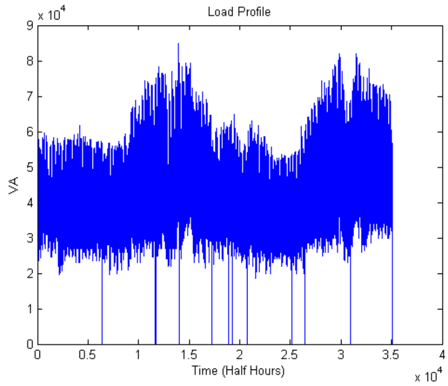
Due to the sporadic availability of the hourly weather data, a few case weather stations are chosen by hand to experiment on against corresponding load. A selection of four substations from Yorkshire and the North East are used.

Instead of using all the hourly load values, a selection of the specific hour noon are used. This is done to avoid the seasonal effects on load. For example, an air temperature of 10°C could occur both during winter and summer at different times of day so there would be a drastic difference in load values if the entire yearly load profile is considered. Thus the data will consist of one air temperature and load reading per day at noon over a period of 2 years from April 2011-March 2013.

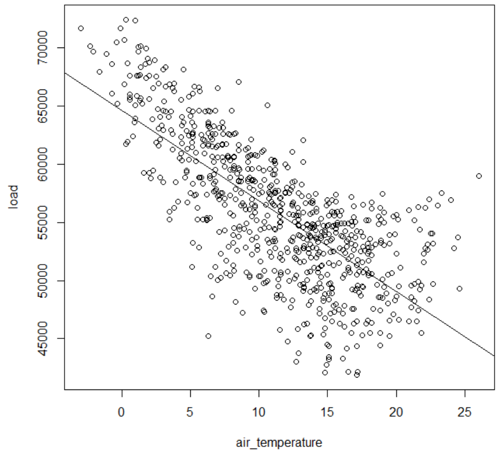
4.1.2 Potter House - North East

Table 4.1: Summary Information of Potter House Temperature Regression

	Estimate	Std. Error	t score	p value
Intercept	64625.31	342.89	188.47	$<$ 2e-16
air_temperature	-778.29	27.07	-28.75	$<$ 2e-16



(a) Potter House Load Profile



(b) Potter House Load Against Air Temperature, $R^2 = 0.5317$

Figure 4.1: Potter House Substation Investigation, April 2011 - March 2013

The first load profile presented here is the Potter House substation from the North East region. The load profile in Fig. 4.1a consists of some domestic type customers as it displays traditional winter peaks over two years. However, the time periods outside the winter peaks seem to be more flat. The plot of load against air temperature with the regression line is given in Fig. 4.1b.

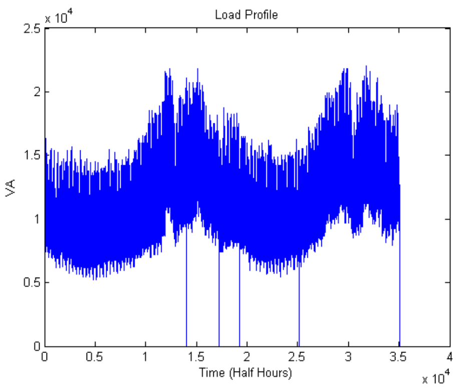
The low p values in Table 4.1 suggest that the correlation is significantly different from zero so the null hypothesis can be rejected. The R^2 value of 0.5317 is in the middle of its possible range indicating that there is a correlation present but there is some deviation away from the regression line. In particular, the higher temperatures seem to be more scattered which corresponds to the original profile in that the load outside the winter peak were more constant. However, the correlation is strong enough in a case such as this that the negative correlation between load and temperature can be established.

4.1.3 Scarborough 11kV - North East

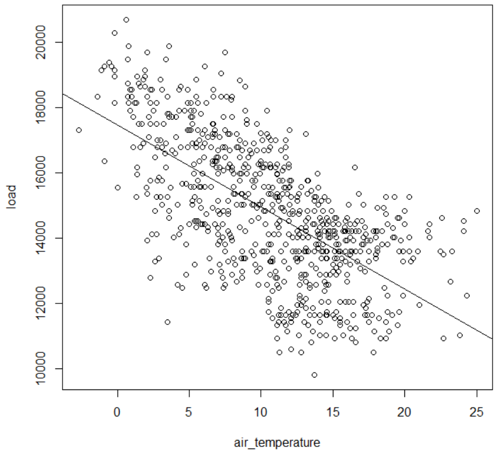
The next load profile tested in the North East region is the Scarborough 11kV substation. This load profile in Fig. 4.2a also displays the traditional winter peak

Table 4.2: Summary Information of Scarborough Temperature Regression

	Estimate	Std. Error	t score	p value
Intercept	17470.90	132.22	132.14	$< 2e\text{-}16$
air_temperature	-251.62	11.15	-22.57	$< 2e\text{-}16$



(a) Scarborough 11kV Load Profile



(b) Scarborough 11kV Load Against Air Temperature, $R^2 = 0.4112$

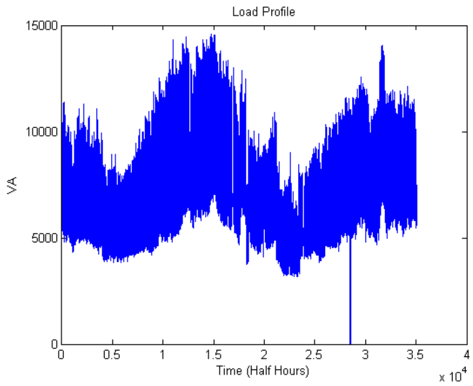
Figure 4.2: Scarborough 11kV Substation Investigation, April 2011 - March 2013

so some correlation between load and temperature is expected. The plot of the load regression line is given in Fig. 4.2b.

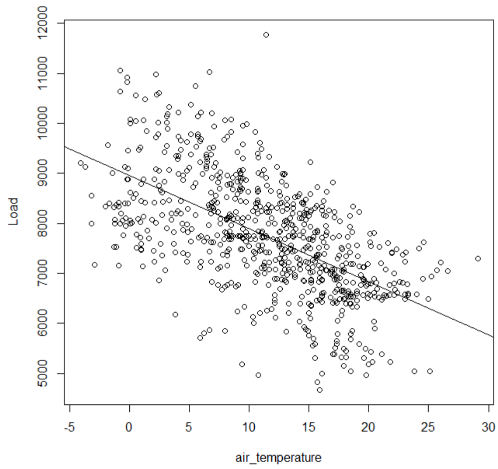
Although the Scarborough load profile displays more traditional domestic characteristics than the Potter House load profile, the R^2 value is actually lower. The relative size of the demand loads between Potter House and Scarborough does have an effect on the R^2 value. This is apparent when considering the t value of air_temperature which is a measure of the number of standard deviations away from 0. In this case it is lower than in Potter House. The summary information in Table 4.2 shows that the null hypothesis can be rejected due to the small p values.

Table 4.3: Summary Information of Snaithing Park Road Temperature Regression

	Estimate	Std. Error	t score	p value
Intercept	8958.88	71.55	125.22	$<$ 2e-16
air_temperature	-106.53	5.45	-19.56	$<$ 2e-16



(a) Snaithing Park Road Load Profile



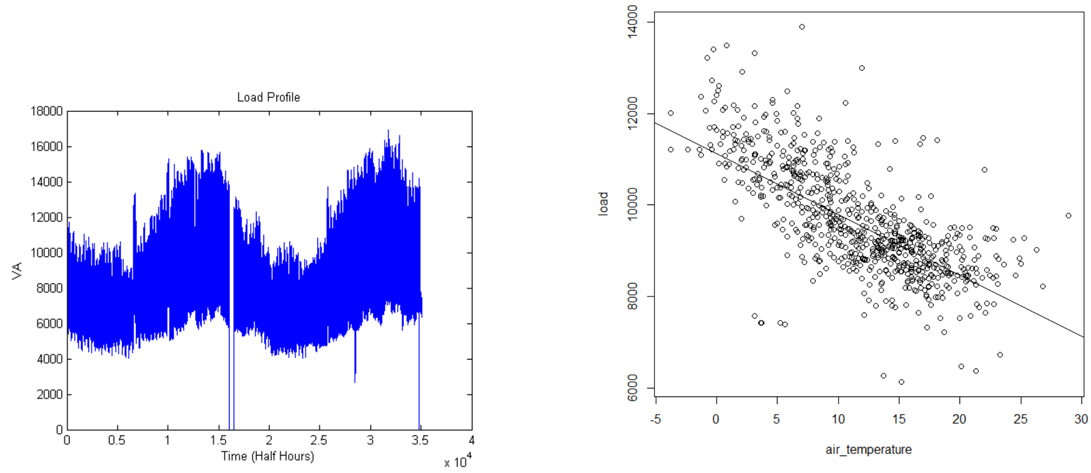
(b) Snaithing Park Road Load Against Air Temperature, $R^2 = 0.3442$

Figure 4.3: Snaithing Park Road Substation Investigation, April 2011 - March 2013

4.1.4 Snaithing Park Road - Yorkshire

The next load profile tested is the Snaithing Park Road substation from the Yorkshire region. The load profile in Fig. 4.3a shows distinct traditional winter peak and summer lows. Therefore the negative correlation between air temperature and load is expected to be strong. Fig. 4.3b shows the linear regression line between temperature and load for the Snaithing Park Road substation.

The low p values in Table 4.3 suggest that there is a correlation in this data set. The R^2 value is not particularly high and this is evident visually in the graph as data points are quite spread out from the regression line. This is surprising since the shape of the load profile suggests distinct peaks and troughs during warmer and cooler temperatures. Although there is a general pattern, there is a lot of inconsistency during this time period. It is possible that during less busy times of



(a) Durkar Low Lane Load Profile

(b) Durkar Low Lane Load Against Air Temperature, $R^2 = 0.2928$

Figure 4.4: Durkar Low Lane Substation Investigation, April 2011 - March 2013

the day there is a stronger correlation between load and temperature, but in order for the data to be meaningful, these correlations should be present at all times of the day.

4.1.5 Durkar Low Lane - Yorkshire

Table 4.4: Summary Information of Durkar Low Lane Temperature Regression

	Estimate	Std. Error	t score	p value
Intercept	11121.45	99.53	111.7	$< 2e\text{-}16$
air_temperature	-133.16	7.65	-17.4	$< 2e\text{-}16$

The next load profile tested is the Durkar Low Lane substation from the Yorkshire region. This load profile in Fig. 4.4a also shows a traditional winter peak and summer lows. As seen in the previous cases, this does not necessarily imply that the correlation will be strong. Fig. 4.4b shows the linear regression line between temperature and load for the Durkar Low Lane substation.

The low p values in Table 4.4 suggest that there is also a correlation in this data set. However, as with the Snaithing Park substation, the R^2 value is not particularly high. There is certainly a general correlation but these results suggest that using air temperature on its own is not enough to show a strong correlation.

4.1.6 Comments on Data Quality

This investigation involved finding hourly air temperature data at Met Office weather stations near selected NPG substations. Unfortunately, hourly air temperature data is not available from all stations recorded in the MIDAS data set when filtered by postcode. Therefore, it is difficult to do a meaningful investigation of all primary substations in NPG's network. Daily maximum and minimum temperatures are more readily available but do not offer the same level of detail as hourly data.

Compounding on this is the variance of customer types on any given substation. If any two customers react differently to changes in temperature than this will be masked at the distribution level. Typically, analysis of temperature and weather conditions can be more explicit when considering individual customers.

4.1.7 Conclusions on Air Temperature Data

It is shown here that there exists a negative correlation between air temperature and demand load in traditional demand load profiles with a winter peak. This follows intuition and is expected. However, the availability of temperature data provided by MIDAS and proximity of substations to weather stations leads to many missing and inconclusive results. This study shows a selection of four substations where in each case the substation is close to a MIDAS weather station and they all have some mix of domestic customers leading to traditional winter peaks. There would be no significant correlation for substations that do not display the traditional winter peak. This will vary on a substation-by-substation basis but using air temperature alone would provide minimal impact on the forecasting accuracy of a new method. By not normalising the demand, there was some variance in the value of R^2 but here a more explicit view of the demand-temperature relationship was needed so that NPG can see the effect of temperature change in terms of voltage amperes. The data could still be used to help contribute to a rule based output but due to the weakness of the correlation, this would most likely not be very useful. This corroborates with NPG's conclusions from their previous studies where they stopped using ACS because it introduced errors in their forecasting process in some cases.

For future work, an overall picture of other weather metrics such as rainfall and visibility should be combined with air temperature in a multivariable analysis against demand load. Air temperature is only one aspect of the weather and so far it has proven to be too simplistic to be used on its own. The main issue is the availability of the data and the scale at which it is used. Weather considerations would be more informative on an individual customer basis.

4.2 Output Area Classification

Economic information such as household income has been used in the literature to determine load patterns. One such method for categorising customer types is the Office of National Statistics' (ONS) Output Area Classifications (OAC). The output area is defined by the ONS as the lowest geographical area where census estimates are released (Bond and Insalaco, 2007). The output areas have similar sized populations and are built from aggregations of postcodes. Additionally, they were also designed to group together households that are socially similar and this led to the creation of the OAC. The overall methodology is given in Fig. 4.5.

A brief summary of the OAC method is given here. The OAC method begins with the output areas from the 2011 UK census. Then initial census variables are selected from the 2011 Census Key Statistics and Quick Statistic tables (Office for National Statistics, 2015). The initial total was 167 variables across various socioeconomic and demographic areas which are consistent throughout the UK. The data then had to be prepared, transformed and standardised which includes:

- conversion into percentages or index scores
- applying log, box-cox and inverse hyperbolic sine transformations in order to reduce skew in the data
- standardising the data such that variables are on the same scale

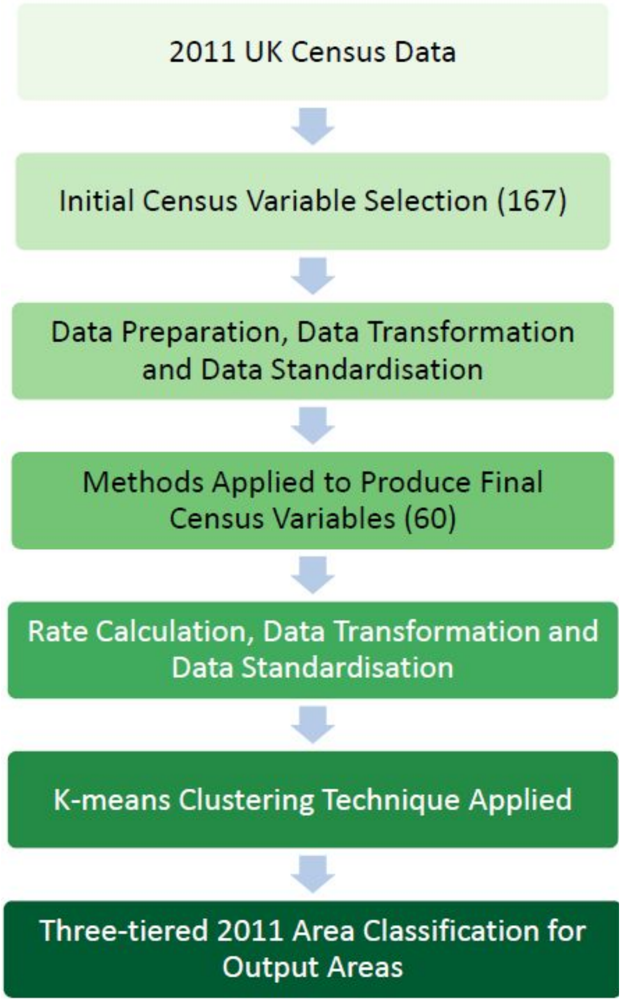


Figure 4.5: OAC Methodology (Office for National Statistics, 2015)

After applying these techniques, the final census variables are selected which reduces the 167 variables to 60 final variables. Then the k-means clustering method is used to partition the data set. The algorithm was applied iteratively in order to produce a hierarchy of clusters consisting of 8 supergroups, 26 groups and 76 subgroups.

For the purposes of this investigation, there will be a focus on the eight supergroups to avoid the high dimensionality in the groups and subgroups. The eight major supergroups are given the following labels (Bond and Insalaco, 2007):

1. Rural Residents

2. Cosmopolitans

3. Ethnicity Central
4. Multicultural Metropolitans
5. Urbanites
6. Suburbanites
7. Constrained City Dwellers
8. Hard-Pressed Living

By considering general labels given by the ONS, it would be ideal to classify NPG substations in such a way that is supported by public demographic information. This would provide the greater insight which NPG needs to make more informed decisions about network reinforcement.

4.2.1 Experiment with OAC

The half hourly demand data available from NPG will be used to find potential patterns between demand and the OAC supergroups. The data set consists of 270 substations from the Yorkshire region where an OAC supergroup can be associated with the substation for the period April 2010- March 2011. The data set itself will contain the maximum demand figures for each substation but a few other metrics will be included:

1. Total number of customers: This figure is provided by NPG so there is a customer figure associated to each substation in this analysis.
2. 2011 Maximum Demand (MD): This is the MD as calculate by NPG from their DLEs. This figure is central to all discussions of the NPG load forecasting process so is included here.
3. MD:Customer Ratio: In order to mitigate the effect of number of customers on MD, a simple ratio is provided as an additional feature. If a substation has a larger number of industrial customers, this can more greatly increase the MD without increasing the total number of customers as greatly.

4. Annual Mean Demand: In addition to the MD, some metrics for the entire annual demand profile are calculated as features. The mean here is the mean of the entire half hourly annual load profile.

5. Annual Median Demand: The median here is the median of the entire half hourly annual load profile.

6. Demand Standard Deviation: The standard deviation of the annual profile will give information about the spread of the load.

7. Demand Skewness: The skewness of the annual profile will provide a measure of the asymmetry of the load.

8. Demand Kurtosis: The kurtosis of the annual profile will provide a measure of the thickness of the tails in the distribution of the load.

For items 4-8, both the original figures and normalised figures will be tested in order to determine if a normalisation process makes any similarities between substations more apparent. In Chapter 3, some popular machine learning techniques were detailed for classification. In order to provide some insight, the decision tree and naïve Bayes classifiers will be used. The decision tree is selected for its interpretability and feature selection capabilities as stated in Section 3.2.1. The naïve Bayes classifier is used for its simplistic modelling capabilities but good prediction capabilities as outlined in Section 3.2.4. Within the MATLAB implementation of these algorithms, there is room for various optimization processes such as decision tree pruning. These will be explored in addition to testing the viability of the model through k-fold cross validation. The error E in the following sections is simply the classification error

$$E = \frac{1}{n}\sum_{i} I(y_i \neq \hat{y}_i) \tag{4.2.3}$$

where n is the number of substations used, y_i is the actual OAC group and $\hat{y}_i$ is the predicted OAC group. $I(y_i \neq \hat{y}_i)$ is the indicator function that is 1 when $y_i \neq \hat{y}_i$ and 0 otherwise. The cross validation error is defined as (Hastie and Tibshirani, 2009):

$$CV(\lambda) = \frac{1}{K}\sum_{k=1}^{K} E_k(\lambda) \tag{4.2.4}$$

where K is the number of equal-sized data proportions known as folds. The model is trained K times on $K - 1$ folds with their respective λ parameters and tested on the remaining fold. This error is averaged and the result is known as the k-fold cross validation error.

4.2.2 Decision Tree Classification

Regarding the overall errors in Table 4.5, it is clear that the decision tree did not perform well in classifying substations by their OAC supergroup. There was a marginal improvement when normalised data was used but not a significant one. The 10-fold cross validation used showed that the algorithm would not be able to handle new information well if additional substations required classification.

Table 4.5: Overall Decision Tree Errors

	Training Error	Cross Validation Error
Original	0.356	0.774
Normalised	0.307	0.730

Table 4.6: Decision Tree Training Misclassifications

	Original Data			Normalised		
	Misclassed	Total	Error	Misclassed	Total	Error
Rural	8	49	0.16	9	49	0.18
Cosmopolitans	9	20	0.45	3	20	0.15
Ethnicity Central	2	7	0.29	6	7	0.86
Multicultural	18	36	0.50	4	36	0.11
Urbanites	10	51	0.20	16	51	0.31
Suburbanites	20	30	0.67	17	30	0.57
Constrained City	12	24	0.50	9	24	0.38
Hard-Pressed	17	53	0.32	19	53	0.36

Table 4.6 details the training classification errors broken down by the OAC supergroups. When comparing the two tables, there does not seem to be any consistency

regarding errors in the supergroups. The one exception being the Rural Residents group which consistently had a low error for both the original and normalised data. However, there are many instances such as with the Cosmopolitans group that had a higher error using the original dataset than the normalised dataset. Regardless, most of the errors are unacceptably high and cannot be used in any type of engineering judgment or understanding.

Further optimisations such as decision tree pruning were attempted but only offered a minor improvement on the cross validation error. This supports the conclusion that there is no relationship between the features provided and the OAC supergroup.

4.2.3 Naïve Bayes Classification

In Table 4.7, the overall errors of the naïve Bayes classification method are given. As with the decision tree, it is clear that both the training error and 10-fold cross validation error are too high to be useful for engineers at NPG. The difference between the original dataset and the normalised data is minimal.

Table 4.7: Overall Naïve Bayes

	Training Error	Cross Validation Error
Original	0.552	0.709
Normalised	0.570	0.683

Table 4.8 shows the classification errors of the OAC supergroups using Bayes classification. There are only minor differences between the original data and normalised data but the resulting errors are high across all groups. The Urbanites supergroup had the lowest classification errors across both the original features and normalised features almost all of the other supergroups had a classification error of more than 0.5.

There does not seem to be any relationship between the features provided on demand and the OAC supergroups based on these high errors. Any optimisation techniques produced a cross validation error of more than 0.5 so any further analysis

Table 4.8: Naïve Bayes Training Misclassifcations

	Original Data			Normalised		
	Misclassed	Total	Error	Misclassed	Total	Error
Rural	17	49	0.35	20	49	0.41
Cosmopolitans	10	20	0.50	13	20	0.65
Ethnicity Central	4	7	0.57	4	7	0.57
Multicultural	25	36	0.69	24	36	0.67
Urbanites	15	51	0.29	13	51	0.25
Suburbanites	26	30	0.87	27	30	0.90
Constrained City	11	24	0.46	13	24	0.54
Hard-Pressed	41	53	0.77	40	53	0.75

of this particular data set will not be completed.

4.2.4 Conclusions on OAC Classification

By using two popular classification algorithms, it was determined that the 2011 OAC supergroups did not provide accurate groupings based on NPG load data. This is most likely due to the areas which the substations serve being too large for the OAC supergroup to capture accurately. The postcodes used for the substations were in a specific area but the substations serve postcode areas that have different OAC supergroups. Therefore more specific information is required which is examined in the following subsection.

4.3 Customer Information by NPG

One of the key problems with examining external data sources at the distribution network level is obtaining data specific to the regions the substation is situated. However, NPG has commissioned for some customer data that can be specifically be traced to individual substations. An investigation from Akperi and Matthews (2014b) is presented here analysing this customer information.

4.3.1 Introduction and Related Work for Customer Information

Customer data has been collected by the sponsor NPG for their primary substations. The initial categorisation is by postal sector which can then be associated to a primary substation in that area. The two major headings are domestic houses and commercial buildings both of which are further split into subcategories. There are 20 domestic house types and 15 commercial building types for a total of 35 subcategories. This high dimensionality makes it difficult to understand the impact of individual contributions. Thus a combined approach of principal component analysis with input from national demand trend using clustering has been used to determine load profile types worth greater investigation.

This demographic information that specifies the type of customer associated to each substation is usually not readily available so analysis of this type of data is not often seen. There are several studies that perform customer classification based mostly on load patterns without a preliminary demographic study (Chicco et al., 2005; Pitt and Kirschen, 1999). This section aims to determine whether an in-house classification of customers is useful at the distribution level.

National demand data is available from the National Grid on a half hourly scale (National Grid, 2014). This represents the prototypical load profile for the country accounting for national trends in weather and economic conditions. Substations in a distribution network on a local scale are not expected to follow this pattern but deviation from this trend can logically be attributed to the individual customer make-up of the substation.

A detailed customer breakdown on the distribution network is available that would normally would not be available in a general study. The need for this type of information stems from the SCADA data being too noisy to offer engineers detailed insight. There is a financial incentive to understand the data such as tariff determination for energy suppliers but also a more general understanding is needed for reliability analysis. In the literature, data mining techniques such as clustering and classification algorithms are popular as in Pitt and Kirschen (1999). Prevalent in both Chicco et al. (2005) and Pitt and Kirschen (1999) is the concept of obtaining representative sets of load profiles based on customer and meteorological conditions. Work done in Figueiredo et al. (2005) presents a framework that also uses a data mining approach but with additional indices for classification based on the time of day. These data mining methods are all able to offer customer categorisation methods for SCADA data with no need for additional information. This study will be unique in the synthesizing of in-house customer classifications with a novel trend identification method.

4.3.2 Data Set Description

First, the customer breakdown data is commissioned by NPG which associates to each NPG substation, the number of customers in 35 different categories, 20 of which are domestic houses and 15 of which are commercial building types.

The domestic housing types are given a label 1-20 as seen in Table 4.9 based on five descriptors: fuel, location, size, tenure and age. The fuel descriptor indicates whether or not the house has a mains gas connection or only uses electricity. The location descriptor shows if the house is either in a rural or urban area. The size descriptor shows if the house is large or small where a small house is generally a flat. The tenure descriptor indicates whether the house is social (council) housing or not. The age descriptor of the house is given by pre-selected time periods. New buildings (sparsely populated) are defined as those built within the last year of the data compilation, a recent house is one built post 1980, old houses from 1920-1979 and houses pre-1920 as very old.

The commercial customer types only have their descriptors available which are:

1. Business at Home
2. Shops & Other Retail Outlets
3. Sports, Leisure, Entertainment, Holiday Activities
4. Unknown
5. Warehouses and Wholesalers

Table 4.9: Domestic Housing Definitions

	Fuel	Area	Size	Tenure	Age
1	Gas	Urban	Large	Other	Old
2	Gas	Rural	Large	Other	Old
3	Gas	Urban	Large	Social	Old
4	Gas	Urban	Large	Other	Very old
5	Gas	Urban	Large	Other	Recent
6	Gas	Rural	Large	Other	Recent
7	Gas	Rural	Large	Social	Old
8	Gas	Rural	Large	Other	Very old
9	Elec	Rural	Large	Other	Old
10	Elec	Rural	Large	Other	Very old
11	Elec	Urban	Large	Other	Old
12	Gas	Urban	Small	Other	Old
13	Gas	Urban	Small	Social	Old
14	Gas	Urban	Small	Other	Very old
15	Elec	Urban	Small	Other	Old
16	Gas	Urban	Large	Social	Recent
17	Gas	N/A	Large	N/A	New build
18	Gas	N/A	Small	N/A	New build
19	Elec	N/A	Large	N/A	New build
20	Elec	N/A	Small	N/A	New build

6. Office and Administration
7. Head Office
8. Other
9. Schools & Educational Establishments
10. Transport
11. Workshops & Repair Centres
12. Factories & Manufacturing
13. Hospital & Medical Establishment
14. Places of Worship and
15. Police, Fire, Ambulance, Courts, Prisons, Civil Defence, Libraries

.

Second, the half hourly demand data for 513 substations is available from a period of April 2010 - March 2013. Some substations cannot be attributed to the customer database because of missing information so this is reduced to 436 substations in total.

Third, the half hourly demand data is publicly available from the National Grid from April 2010 - March 2013 which represents the prototypical load profile for the country.

4.3.3 PCA Overview

The aim of principal component analysis (PCA) is to analyse multivariate data by way of reducing the dimensionality of the data set. This is done by transforming the original space into a new set of uncorrelated variables in decreasing order of importance. The first principal component (PC) retains most of the variance in the original data set, the second PC retains the second most variance and so on.

The principal component space can also be rotated to help with interpretation. After picking a lower dimensional space from PCA, the space can be rotated so that the loadings fall closer to the principal component axes. In general, the principal component space can be rotated either by an orthogonal transformation or an oblique one Jolliffe (1986). Suppose the original matrix of loadings is given by A_m, a $35 \times m$ matrix in this investigation where m is a chosen number of dimensions. For an orthogonal transformation, a rotation matrix T is found such that a new matrix $B_m = A_m T$ of loadings optimises a simplicity criteria Jolliffe (1986).

Analysing relationships between variables in the original space can be difficult just by examining a covariance matrix and similarly the principal component scores of the transformed data. Therefore it is useful to have a graphical representation of the data being examined. The purpose of the biplot is threefold. One is to plot the PC scores in a 2D (or 3D) space so that relationships between individual substations can be deduced in the new space. Second is to determine where the original variables lie in the new PC space and their relation to each other. Finally, the simultaneous position of both the data point PC scores and original variables represented by vectors can be considered. The direction of the vectors can give an indication of a point having a high or low PC score in that variable.

4.3.4 K-Means Clustering Overview

K-means clustering is known as an unsupervised learning algorithm because its objective is not to predict a classification but rather to find patterns in data without labels. After specifying the number of desired clusters, points are chosen at random to be cluster centres. All points are assigned to their closest centre by a distance metric, typically Euclidean distance. Then the means of all points in each cluster are taken and these become the new centres. The algorithm stops when all points are assigned to the same cluster in consecutive rounds (Hall et al., 2011).

The initialisation of points chosen as cluster centres can prove to be problematic as they can often result in different clusters. In order to obtain a good solution, the clustering algorithm has been run 100 times and the solution with the lowest total squared distance between the substation points and the centroids will be chosen (Hall et al., 2011).

There is evidence to support that when using K-means clustering with the Euclidean distance metric that using the first few principal components yields accurate clusters (Yeung and Ruzzo, 2001). In order to determine the number of clusters, the Davies-Bouldin (DB) criterion will be used which is a ratio of within cluster and between cluster differences. Suppose S_i and S_j are dispersion measures which are the average distances between each point in the clusters and their respective centroids. $M_{i,j}$ is the Euclidean distance between the ith and jth clusters. Then the DB index $\bar{R}$ is

$$\bar{R} = \frac{1}{N}\sum_{i=1}^{N} max_{i\neq j}\{R_{i,j}\},\ R_{i,j} = \frac{S_i + S_j}{M_{i,j}}. \tag{4.3.5}$$

The best clustering solution has the smallest DB index (Davies and Bouldin, 1979).

4.3.5 Comparing Monthly Growths

In order to compare the monthly growths of the demand with the National Grid demand, all of the substation demands and national demand from April 2010 - March 2013 are normalised. A comparison can be achieved by finding the difference between the monthly changes and then calculating the variance to show the spread of errors. The change in variance from the lowest error substation to the highest is shown in Fig. 4.6.

A decision regarding the choice of variance value where load profiles are said to more closely follow the national trend is nontrivial. The point to be selected as a threshold value needs to select a subset of the data that not only includes outliers but also includes those points which start to become more sparsely spread out. In Section 4.3.7, several threshold values are looked at but for now, 0.01 is selected as a threshold value because variance points become more spread out and it leaves out about a quarter of the data as being the furthest away from the national trend.

4.3.6 Usage of Principal Component Analysis

Principal component analysis was applied to the customer profiles of the 436 substations in the distribution network. A scree plot of the variance explained by the first

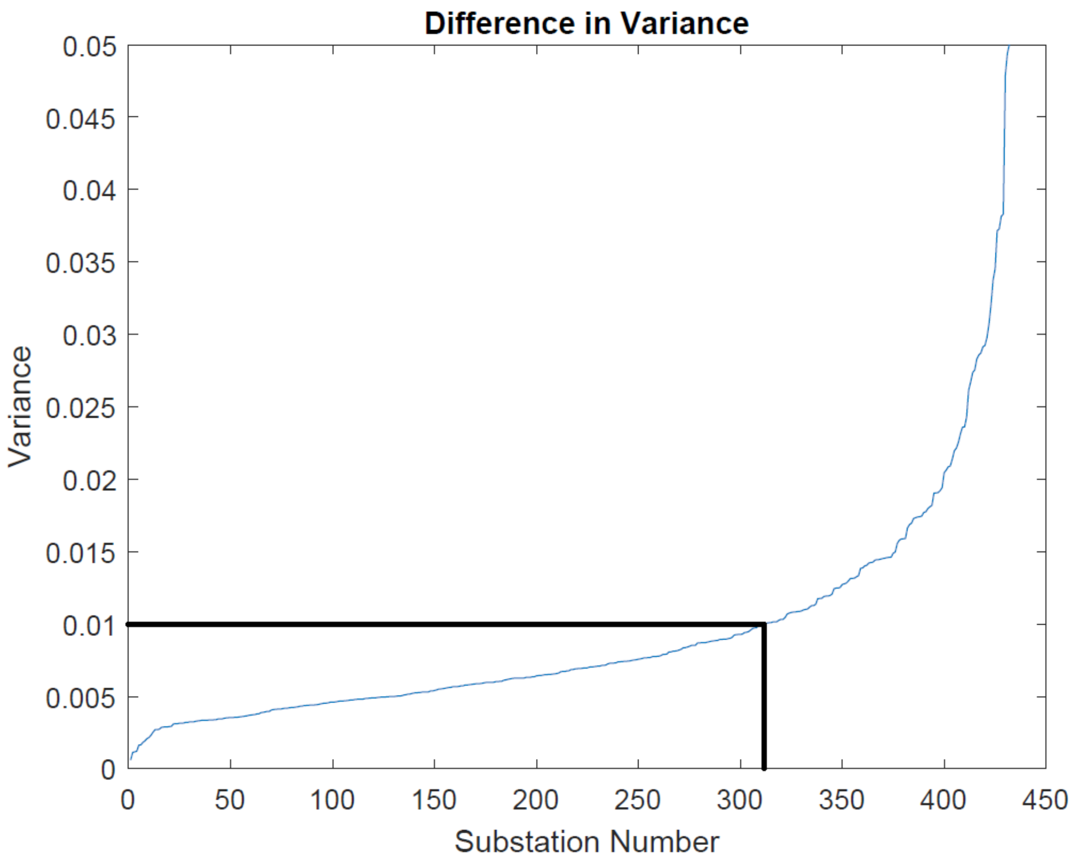


Figure 4.6: Difference in Error Variance between Substation Trend and National Trend

10 principal components is given in Fig. 4.7. A popular method for determining the number of principal components that should be used is looking at where the scree plot starts to level out. For the purposes of visualising the PC space, the first three PCs are selected and this is also justified by the variance explained in the scree plot. The first three PCs explain about 71% of the variance in the original space.

This data was also plotted in Matlab using the biplot function in Fig. 4.8 along with the scores for each of the 436 substations. By convention, the biplot function makes the element with the largest absolute value in each column of the factor loadings matrix a positive value by changing the sign on the entire axes. In this case, PC2 was mirrored. This does not change the meaning of the plot. Also, the scores of the actual substations are scaled down so that they will fit the plot.

In Fig. 4.8, it is possible to see a visual representation of the first two principal

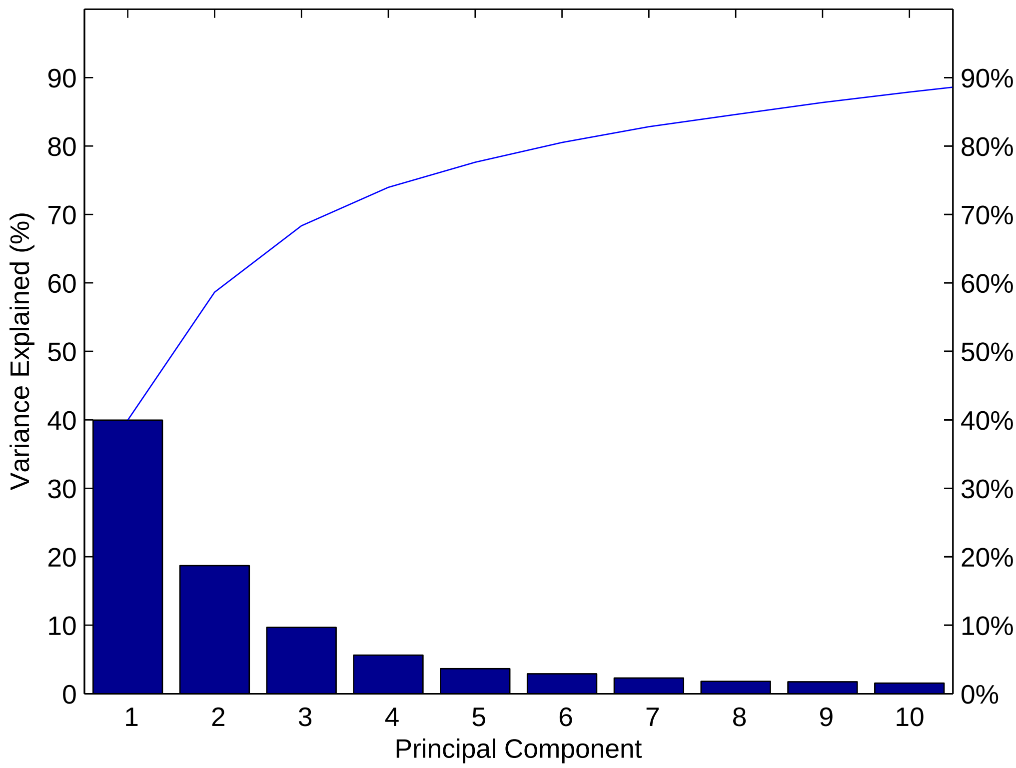


Figure 4.7: Scree Plot of Variance Explained by Principal Components

components (from the 3D space). From here, it is possible to see the correlation between certain categories of customers. In the figure, the customer labels are R for domestic rural housing, U for domestic urban housing, C for commercial customers and N for new domestic buildings. Almost all of the commercial customer classifications are grouped together as well as certain types of domestic customers where the main divide seems to be whether the customer is urban or rural. The scores of the substations themselves seem to be more concentrated towards the left of the plot which suggests there is a greater influence from domestic customers, in particular those in urban environments.

Afterwards, K-means clustering is applied on the three dimensional PC scores and by using the DB criterion, seven clusters are found to be optimal. Fig. 4.8 shows the clusters in the PC space.

Now using a variance of 0.01 as a threshold value, the substations which fall below and above this can be separated on the same PC space as in Fig. 4.9. This choice

Table 4.10: Substations that Matched Trend by Cluster

	Matches Trend	Does not match trend
1 - Yellow	31	5
2 - Magenta	31	7
3 - Cyan	11	0
4 - Red	70	59
5 - Green	40	23
6 - Blue	110	25
7 - Black	19	5

is based on the elbow method by considering Fig. 4.6 and looking for a significant change in variance difference. Substations which deviate greatly are in purple and those which do not are in green. By comparing this to the initial K-means clustering of the space, Table 4.10 shows the number of substations in each cluster that fall

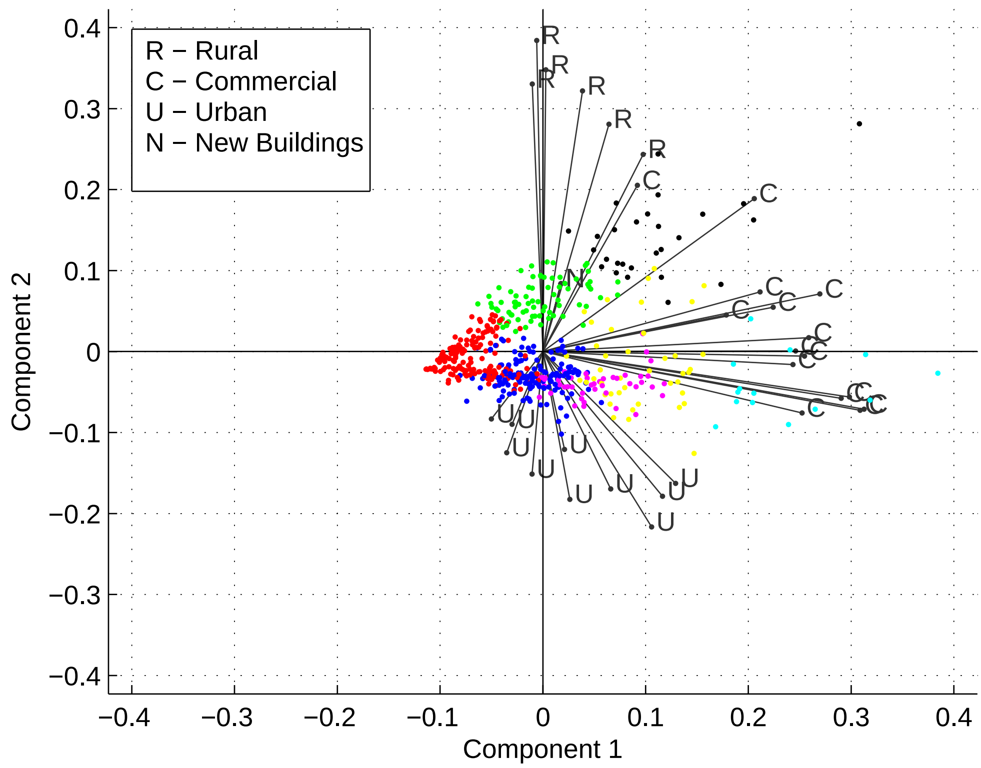


Figure 4.8: K-Means Clustering of PC Space

above or below this threshold value. The clear anomaly in this table is the red cluster which is concentrated in the left of the plot. These substations are the ones which most strongly deviate from the commercial variable vectors on the right of the biplot. Another interesting observation is that the substations in the green and blue cluster containing the next highest proportion of substations that deviate from the national trend are still distinct from the red cluster. This suggests that the spread of the urban variables may also be significant. In particular, domestic groups 1, 3 and 5 in the bottom left are all large urban houses built no earlier than 1920 whereas the other urban customers occupy flats or have no mains gas connection. Although this is not enough evidence to dismiss the need for investigating the latter group, it does suggest the former group of customers is worth further investigation.

As an illustration of the power of PCA, Fig. 4.10 shows the difference between a load profile with a high PC1 score and a load profile with a low PC1 score. The load profile in Case I not only has a higher demand but is more orderly and levelled.

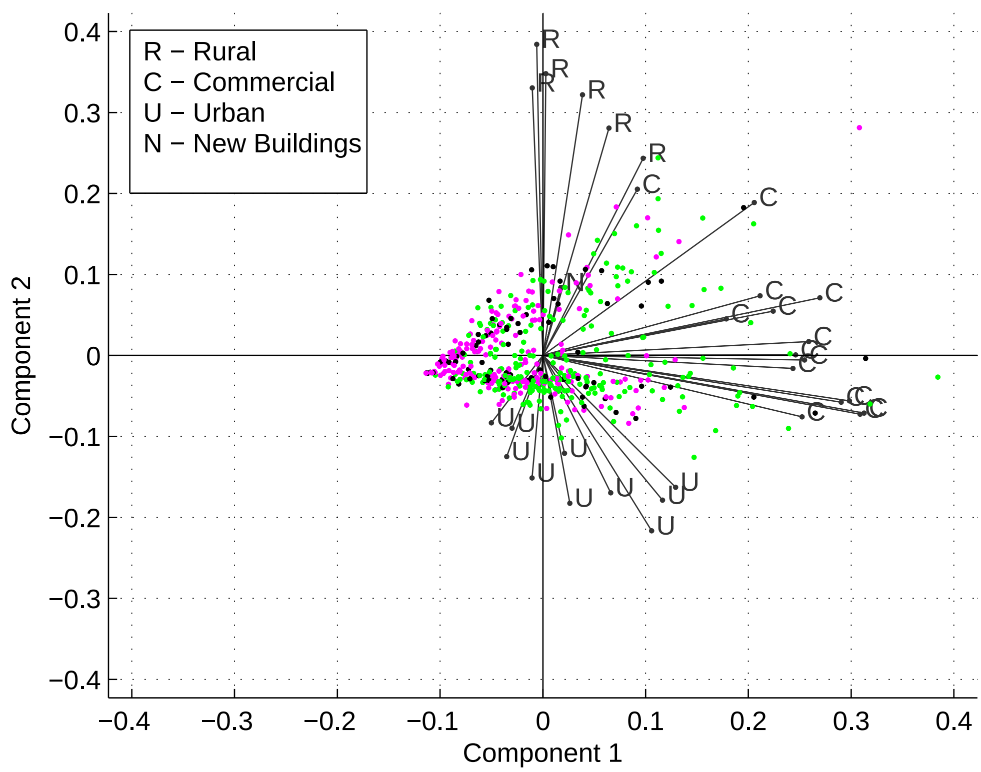
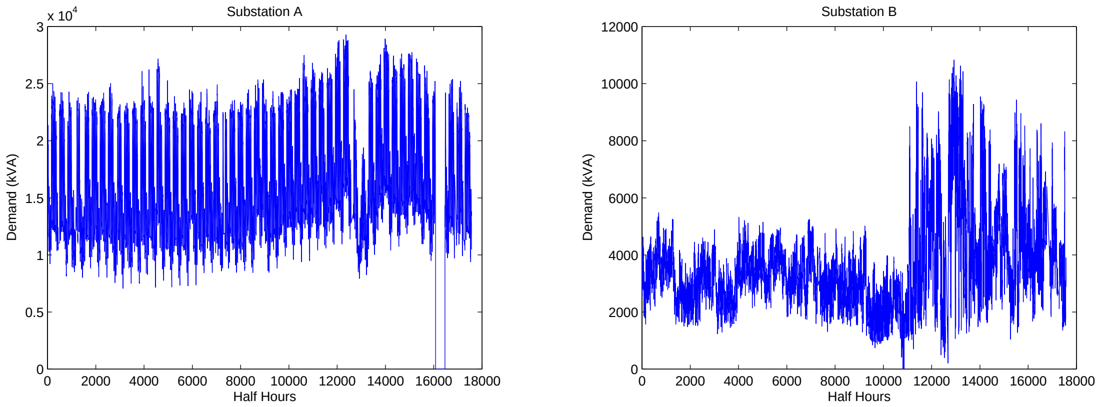


Figure 4.9: Substations which Deviate from National Trend



(a) Case I - High PC1 Score

(b) Case II - Low PC1 Score

Figure 4.10: Difference between Load Profiles in Principal Component 1

This is due to the customer mixture on this substation having more industrial type customers. The load profile in Case II is erratic by comparison so it is unsurprising that it would not follow a national trend as closely. More household domestic customers and only a few industrial customers are using load from this substation which leads to a more erratic profile.

4.3.7 Adjustment of threshold value and error metrics

As the choice of threshold parameter is somewhat subjective, there needs to be an allowance for it to be adjusted by distribution network engineers. Using the cluster analysis, it is also possible to assign error metrics to these groupings. In addition to the threshold, there needs to be allowance for adjustment of the proportion (P) of substations in a cluster that define whether it follows the trend or not. Let the clusters which have a proportion of substations greater or equal to P that do not match the national trend be a prediction of a substation not following the national trend. Then by using a confusion matrix, the engineer can gain an understanding of the accuracy of the cluster groupings dependent on the choice of threshold value.

The confusion matrix shown in Table 4.11 where the labels are defined as follows. The labels "MT" and "DNMT" represent "matches trend" and "does not match trend" respectively. The true positives (TP) are the substations that are correctly predicted as following the national trend. The true negatives (TN) are the

Table 4.11: Confusion Matrix

		Predicted	
		MT	DNMT
Actual	MT	TP	FN
	DNMT	FP	TN

substations that are correctly predicted as not following the trend. A false positive (FP) is a substation predicted as following the trend when it is not and a false negative (FN) is a substation that is predicted to not follow the trend when it does. Therefore the overall accuracy (ACC) is defined as Hall et al. (2011):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.3.6}$$

In addition, the true positive rate (TPR) in Eq. 4.3.7 identifies the ratio of substations that match the trend that are correctly identified. Similarly, the false positive rate (FPR) in Eq. 4.3.8 identifies the ratio of substations that do not follow the trend that are correctly identified.

$$TPR = \frac{TP}{TP + FN} \tag{4.3.7}$$

$$FPR = \frac{FP}{FP + TN} \tag{4.3.8}$$

Lastly, a measure called Matthew's correlation coefficient (MCC) in Eq. 4.3.9 is used to account for the biases present in the other metrics by using true positives and negatives and false positives and negatives. MCC ranges from 0 to 1 and a higher value denotes better results. It is analogous to Pearson's product moment correlation coefficient as a measure of fit between the predicted and actual cases in the binary confusion matrix Powers (2007).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{4.3.9}$$

Table 4.12: Error Metrics for 30% Proportion [NaN is undefined]

P=30%				
	ACC	TPR	FPR	MCC
T = 0.0035	0.892202	0	0	NaN
T = 0.0046	0.775229	0	0	NaN
T = 0.0054	0.669725	0.053691	0.010453	0.130771
T = 0.0064	0.56422	0.050251	0.004219	0.146207
T = 0.0076	0.522936	0.231076	0.081081	0.198556
T = 0.0093	0.65367	0.654485	0.348148	0.285301
T = 0.0127	0.733945	0.773639	0.425287	0.305018
T = 0.0204	0.915138	1	1	NaN
T = 0.0928	0.997706	1	1	NaN

Table 4.13: Error Metrics 50% Proportion [NaN is undefined]

P=50%				
	ACC	TPR	FPR	MCC
T = 0.0035	0.892202	0	0	NaN
T = 0.0046	0.777523	0.061224	0.014793	0.123593
T = 0.0054	0.674312	0.187919	0.073171	0.172317
T = 0.0064	0.600917	0.336683	0.177215	0.183437
T = 0.0076	0.610092	0.772908	0.610811	0.175523
T = 0.0093	0.690367	1	1	NaN
T = 0.0127	0.800459	1	1	NaN
T = 0.0204	0.915138	1	1	NaN
T = 0.0928	0.997706	1	1	NaN

In Tables 4.12-4.14, a few different threshold values are recorded and which roughly correspond to a value taken every 50 substations along with their corresponding ACC, TPR, FPR and MCC. Note that "NaN" means the value is undefined for MCC in cases where $TPR = 0$ or $FPR = 0$. It is evident that in all cases, ACC is a poor measure of how well this method performs, especially in cases where

Table 4.14: Error Metrics 70% Proportion [NaN is undefined]

P=70%				
	ACC	TPR	FPR	MCC
T = 0.0035	0.892202	0	0	NaN
T = 0.0046	0.754587	0.204082	0.085799	0.156329
T = 0.0054	0.538991	0.644295	0.515679	0.122881
T = 0.0064	0.456422	1	1	NaN
T = 0.0076	0.575688	1	1	NaN
T = 0.0093	0.690367	1	1	NaN
T = 0.0127	0.800459	1	1	NaN
T = 0.0204	0.915138	1	1	NaN
T = 0.0928	0.997706	1	1	NaN

the selected threshold is too high or low. Similarly, TPR and FPR can be misleading because of the nature of the method. In the cases where TP, FP, FN and TN are all nonzero, MCC is the most unbiased metric of the performance of this method.

Ultimately, it is left to the judgment of the network engineers as to how much error is acceptable for the national trend to be used as a proxy for a substation trend. Even the best recorded MCC in Table 4.12 of 0.305 is still relatively low but it still highlights the most problematic substations. Therefore, depending on the requirements of error acceptance, this methodology can be used to determine which substations follow the national trend or which substations are the most problematic or a combination of both.

4.3.8 Conclusions on NPG Customer Information

By using PCA, domestic urban, domestic rural and commercial customers can be successfully grouped in the principal component space which follows intuition. By using a biplot, it is possible to achieve a visualisation about the relationship between the PC scores of the substation and the customer variables. Additionally, through the use of clustering and external data, substations with certain PC scores can be identified to be more problematic than others. This provides a basis for looking at the properties of certain customers that deviate the most from the national trend. Of course since this data is unique to the sponsor, it relies on the accurate assessment of customers connected to the distribution network in their respective postal sectors. Therefore this work can be furthered by using more publicly available demographic databases. Furthermore, the shape of load profiles should also be investigated as it pertains to the distinction between customer groups in this investigation.

4.4 Conclusion

When considering supplementary information to a forecasting model, it should be determined if that information is fit for purpose. In this chapter, RQ2 is the main research question which is brought to a close here. That is, external factors and their potential impact on demand variation were considered here. Due to the fact that the demand at the substation level is an aggregation of hundreds or thousands of customers, it can be difficult to use publicly available databases that capture all of the variance within that range. Many of the studies on electrical distribution forecasting methods focus on forecasting for individual customers or have individual customers available.

It is not necessarily the case that databases such as MIDAS or the OAC supergroups are not significant or useful in the discussion of electrical demand but rather for NPG substations at the distribution level, they are not fit for purpose. The data set which NPG provided allowed for a similar discussion to what the OAC supergroups attempted to do but was much more specific to individual substations. Therefore, the NPG customer information will adopted in the CL-ANFIS method detailed in Chapter 6.

Chapter 5

Methods for Analysing Substation Demand Data

The raw data received from NPG needs to be analysed for extreme outliers, missing data and load transfer scenarios. In Section 5.1, a simple novel technique for producing a maximum demand figure for NPG is introduced which is comparable to the NPG MD selection method. This is a novel method to take into account any outliers or load transfers which would require some engineering judgement from NPG. Such an automated method will be needed when doing a comparison between the CL-ANFIS forecast and NPG's DLE forecast. This is a key deliverable for enhancing the NPG forecasting method. Also introduced is a novel method for the detection of load transfers which are periods of time when demand is temporarily transferred from one substation to another. Section 5.2 uses the outlier and load transfer detection methods to determine the base load profile. The base load profile is the assumed annual profile after outlier and any other anomalous periods of demand are adjusted.

5.1 Consideration of Maximum Demand and Outlier Detection

NPG must determine maximum demand (MD) values for their substations in order to document them in their DLEs. The half hourly substation demand data is stored in a system called PI. Then the maximum demand is determined by using an algorithm based on the load duration curve. The methodology for this is detailed in Section 2.3.1. In order to make unbiased comparisons to the NPG method, a comparable MD selection method is introduced here as well as the use of an outlier detection technique appropriate for time series data.

5.1.1 Automated Maximum Demand Selection

After obtaining load profiles from their SCADA systems, a comparison between the current MD and the MD of the previous year is made. If the difference exceeds a certain threshold value, then further inspection by an engineer is required. NPG has documented a few different scenarios (CE Electric UK, 2009) on how the maximum demand would be selected depending on the stability of the profile. Namely the engineer would judge whether some demand peaks are clearly outliers or load changes and ignore them. However, the choice is not always clear and two engineers may not necessarily agree on the same peak. Thus there is an interpretation error of up to 10% (CE Electric UK, 2009) on all maximum demands which need to be inspected. NPG has recognised that there is a need for assessment guidelines to be developed for the determination of maximum demand.

Here a novel automated process is introduced to determine maximum demand. The algorithm is as follows:

1. Input the substation half hourly load profile.

2. Find and store the maximum demands for each day in the load profile.

3. Calculate the kurtosis of the entire load profile.

4. If the kurtosis of the load profile is less than 3 then:

   (a) Find and eliminate outliers from the initial smoothed MD curve using wavelet based detection.
   (b) Smooth this curve using a moving average filter.
   (c) Find the maximum value of the smoothed curve and output as the maximum demand.

5. If the kurtosis of the load profile is greater or equal to 3 then:
   (a) Flag substation as having a distribution with fat tails.
   (b) Find and eliminate outliers from the initial smoothed MD curve using wavelet based detection.
   (c) Smooth this curve using a moving average filter.
   (d) Find the maximum value of the smoothed curve and output as the maximum demand.

Firstly, the daily maximum demands are stored. Since the maximum values are of the most concern, taking daily maximums drastically reduces the amount of data points the rest of the algorithm must deal with while still capturing the demand trend. Larger time windows could distort the data by extending outliers for a longer period of time. A shorter time window than daily demand would give more data but not for any noticeable benefit whilst increasing running time of this and future methods.

Then the kurtosis is of the entire demand profile is calculated and is given by the equation:

$$k = \frac{E(X-\mu)^4}{\sigma^4} \tag{5.1.1}$$

where μ is the mean of a distribution X and σ is the standard deviation of X (Mathworks, 2013a). Kurtosis is a measure of how outlier prone a distribution is. The kurtosis of the normal distribution is 3. Thus the algorithm takes two separate routes depending on whether the kurtosis indicates it is more or less outlier prone than a normal distribution.

Ivezić et al. (2014) gives an illustration of distributions with positive $k-3>0$ and negative $k-3<0$ kurtosis values. A high kurtosis value is expected if there

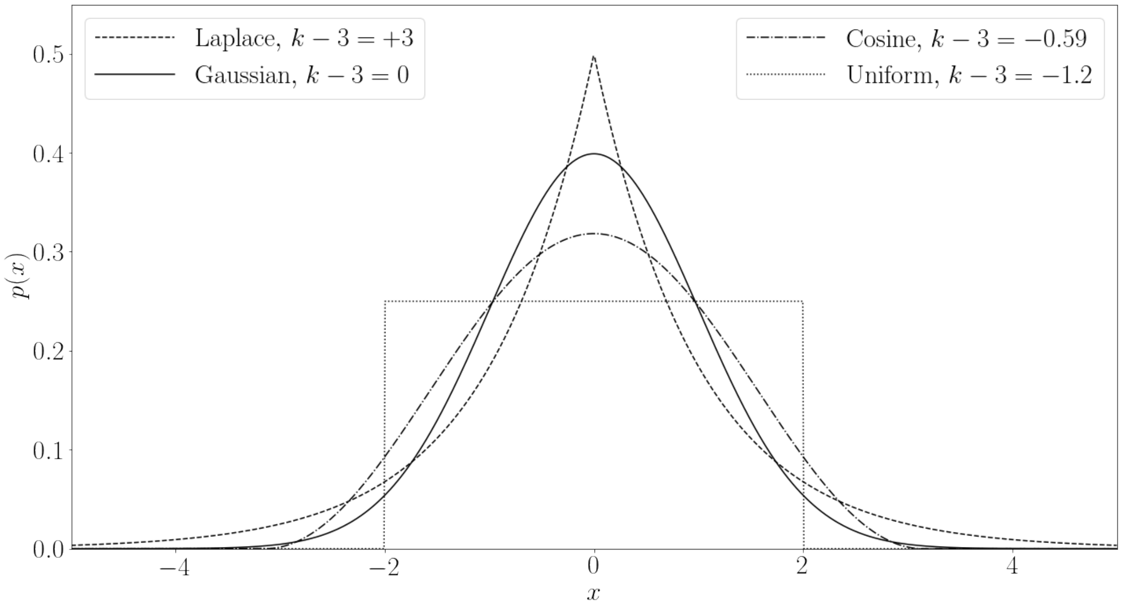


Figure 5.1: Example of Distributions with Positive and Negative Kurtosis Values (Ivezić et al., 2014)

is a large concentration of probability in the tails. A low kurtosis value is expected if the distribution has small tails. Here the kurtosis is used as a flag in order to indicate how outlier prone the substation demand data is. The outlier detection itself is done using a wavelet method which is discussed in Section 5.1.2.

In both instances, the curve will be smoothed using a moving average filter. This smooths the data by replacing each point with the average of its neighbouring points. This is given by the equation:

$$y_s(i) = \frac{1}{2N+1}(y(i+N) + y(i+N-1) + ... + y(i-N)) \tag{5.1.2}$$

where $y_s(i)$ is the smoothed value for the ith data point, N is the number of neighbours on either side of $y_s(i)$ and $2N+1$ is the span (Mathworks, 2013b). In this method the span will be set to 5 so that longer term variance in demands caused by seasonal differences will not effect the smoothing process. Smoothing data helps to reduce the impact of large spikes so that they will not be considered the maximum demand selection.

Finally, the maximum of this smoothed curve is then output as the maximum demand. As an example, Fig. 5.2 shows the indicated MD level of this algorithm which is analogous to Fig. 2.3a which showed the maximum demand choice for a

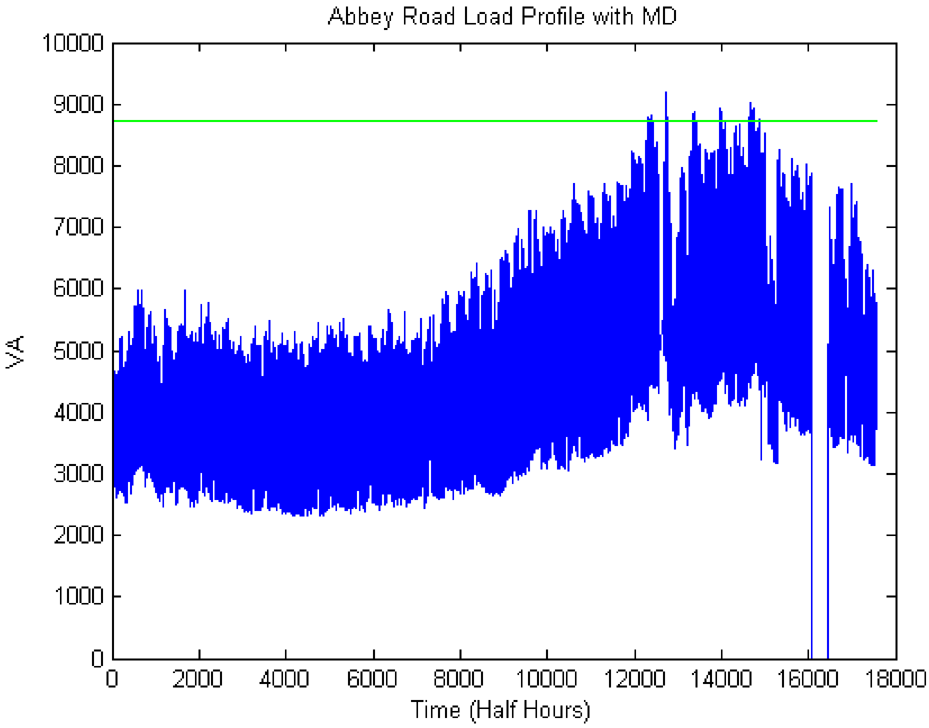


Figure 5.2: Abbey Road Annual Load Profile with Maximum Demand in Green

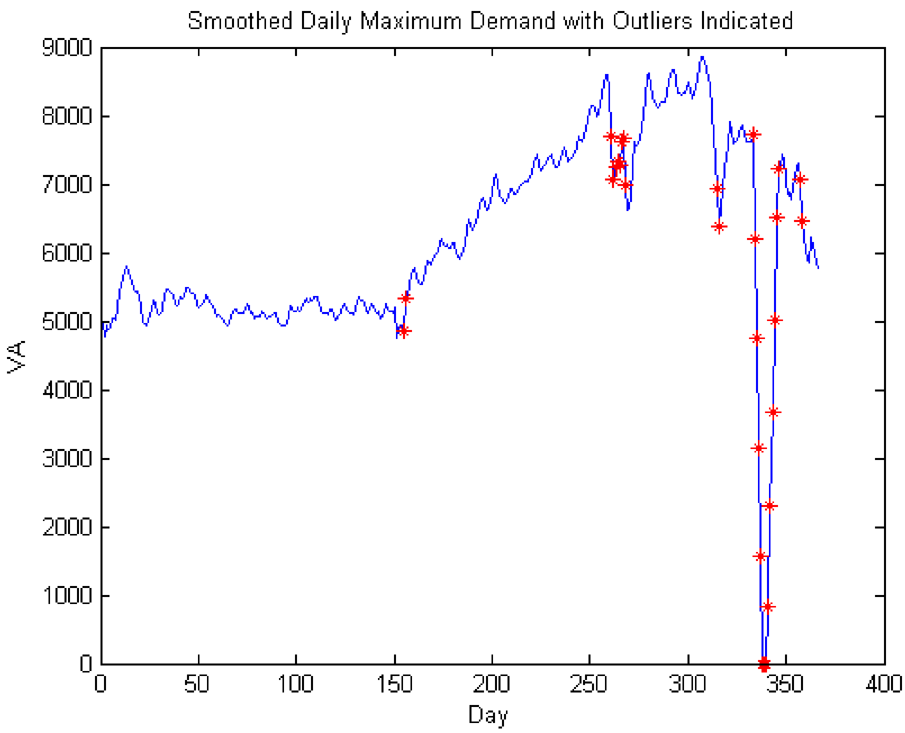


Figure 5.3: Abbey Road Daily Load Profile with Outliers Indicated

stable profile. The actual MD of this particular substation from the NPG DLEs was 8.82×10^3 VA while the MD chosen here is 8.73×10^3 VA for a difference of -1.01% indicating that in this example, the method is able to approximate the NPG MD well. Further comparisons are made in Section 5.1.4. Fig. 5.3 shows the outliers chosen for the same Abbey Road substation showing that it was able to detect the shifts in load and missing data. It is noted that the selection of outliers in itself can be subjective such as the first two points seen in Fig. 5.3. However, for a time series, an approach that detects changes from point to point rather than considering a given demand range is more appropriate here. Consider the outlying points that are detected between day 250 and day 300 in Fig. 5.3. There is clearly some anomaly occurring during this time period as seen in Fig. 5.2. If some range or percentile based approach was used instead, these points would not be considered outliers. However, this could have an effect on MD selection or on forecasting if they are not considered outliers. Further explanation of the wavelet method used is given in the following section.

5.1.2 Wavelet based outlier detection

Bilen and Huzurbazar (2002) proposed an algorithm to detect outliers using the following steps:

1. Apply the discrete wavelet transform using the Haar wavelet to the observed series $\{Z_t\}$ to obtain the first level of detail coefficients $D(J-1)$.

2. Estimate σ_1 from the data by taking mean absolute deviations from the median of wavelet coefficients:

$$\begin{aligned} \hat{\sigma}_1 &= AD(D(J-1)) \\ &= \frac{1}{n_1}\sum_{k=1}^{n_1} |d_k(1) - M_1| \end{aligned} \tag{5.1.3}$$

   where $d_k(1)$ is the wavelet coefficient at k and M_1 is the median of the first level coefficients.

3. Calculate the threshold limit τ_1 using the estimated σ_1:

$$\tau_1 = \hat{\sigma}_1\sqrt{2log(n)} \tag{5.1.4}$$

4. Find $S = \{s_1, ..., s_m\}$, the set of indices such that $|d_k(1)| > \tau_1$

5. Find the exact position of the outlier. The outlier is in position either $(2s_k)$ or $(2s_k - 1)$. Compute the sample mean of the original series without the observations at $(2s_k)$ or $(2s_k - 1)$:

$$\bar{Z}* = \frac{1}{n-2} \sum_{t \neq 2s_k, 2s_k - 1} Z_t \tag{5.1.5}$$

Then the location of the outlier is $(2s_k)$ if $|Z_{2s_k} - \bar{Z}*| > |Z_{2s_k - 1} - \bar{Z}*|$. Otherwise it is at $(2s_k - 1)$.

A brief discussion of wavelets for the discrete wavelet transform is presented here summarised from Ogden (1997). The Haar wavelet transform is used in this wavelet transform and frequently used in many discussions of wavelets. The Haar mother wavelet is defined as:

$$\psi(x) = \begin{cases} 1 & x \in [0, \frac{1}{2}), \\ -1 & x \in [\frac{1}{2}, 1), \\ 0 & \text{otherwise} \end{cases} \tag{5.1.6}$$

From this mother wavelet we can construct a family of wavelets by translation and dilation. Let j be the dilation index and k the translation index. Each wavelet born of the mother wavelet is

$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k) \tag{5.1.7}$$

This leads to a key theorem in wavelet theory:

Theorem 5.1.1 The set $\{\psi_{j,k}, j, k \in \mathbb{Z}\}$ is a complete orthonormal system for $L^2(\mathbb{R})$

We then carry on this idea by defining a function space V_j to be:

$$V_j = \{f \in L^2(\mathbb{R}) : f \text{ is a piecewise constant on } [k2^{-j}, (k+1)2^{-j}), k \in \mathbb{Z}\} \tag{5.1.8}$$

Then consider the scaling function in the Haar wavelet case also referred to as the father wavelet:

$$\phi(x) = \begin{cases} 1 & \text{if } x \in [0, 1), \\ 0 & \text{otherwise} \end{cases} \tag{5.1.9}$$

Then V_j is spanned by the set of functions $\{\phi_{j,k}, k \in \mathbb{Z}\}$ and is an orthonormal basis. If we wish to project an $L^2(\mathbb{R})$ function onto an approximation space V_j then we can write it in terms of the scaling functions: $P^j f = \sum_k c_{j,k}\phi_{j,k}$ and since $\{\phi_{j,k}, k \in \mathbb{Z}\}$ is an orthonormal basis for V_j then we can compute the scaling function with the integral $c_{j,k} =< f, \phi_{j,k} >= \int_{\infty}^{\infty} f(x)\phi_{j,k}(x)dx$.

It can also be shown that we can write the family of scaling functions as $\phi_{j,k}(x) = (\phi_{j+1,2k}(x) + \phi_{j+1,2k+1}(x))/\sqrt{2}$ and derive that in the Haar case the coefficients can be written as :

$$c_{j,k} = (c_{j+1,2k} + c_{j+1,2k+1})/\sqrt{2} \tag{5.1.10}$$

Similarly we can do this with the Haar wavelet detail functions so that the detail coefficients are defined as $d_{j,k} =< f, \psi_{j,k} >= \int_{\infty}^{\infty} f(x)\psi_{j,k}(x)dx$ and since $\psi_{j,k}(x) = (\phi_{j+1,2k}(x) - \phi_{j+1,2k+1}(x))/\sqrt{2}$ then we can see that:

$$d_{j,k} = (c_{j+1,2k} - c_{j+1,2k+1})/\sqrt{2} \tag{5.1.11}$$

This is a special case of the decomposition algorithm where we can find lower level coefficients from higher level ones. In theory, we would need to do numerical integration to obtain the initial wavelet coefficients but instead the original series is taken as our initial "guess" after which we can apply the decomposition algorithm by taking the difference between entries and dividing by $\sqrt{2}$. If there is an odd number of entries k then create a new entry Z_{k+1} such that $Z_k = Z_{k+1}$.

Then we need to identify a threshold value where the detail coefficients show a large change in values. Originally discussed by Donoho and Johnstone (1994), the value $\tau = \sqrt{\sigma(2\log(n))}$ is based on the result that

$$P\{\max_{1\leq i\leq n} |Z_i/\sigma_i| > \sqrt{2log(n)}\} \to 0 \text{ as n} \to \infty \tag{5.1.12}$$

Bilen and Huzurbazar (2002) determined that the outlier is at $(2s_k)$ or $(2s_k - 1)$ based on the deviation from the mean value. However, due to the downsampling of data, the algorithm here has been modified to include both $(2s_k)$ and $(2s_k - 1)$ as outliers.

5.1.3 Usefulness of Kurtosis Flag

The kurtosis measure is not used in the outlier removal process itself but is used more as a flag to give an indication about the peakedness and tail thickness of the load distribution. In order to gauge the usefulness of using such a measure, it is worthwhile looking at the distribution of values the kurtosis takes in a sample year.

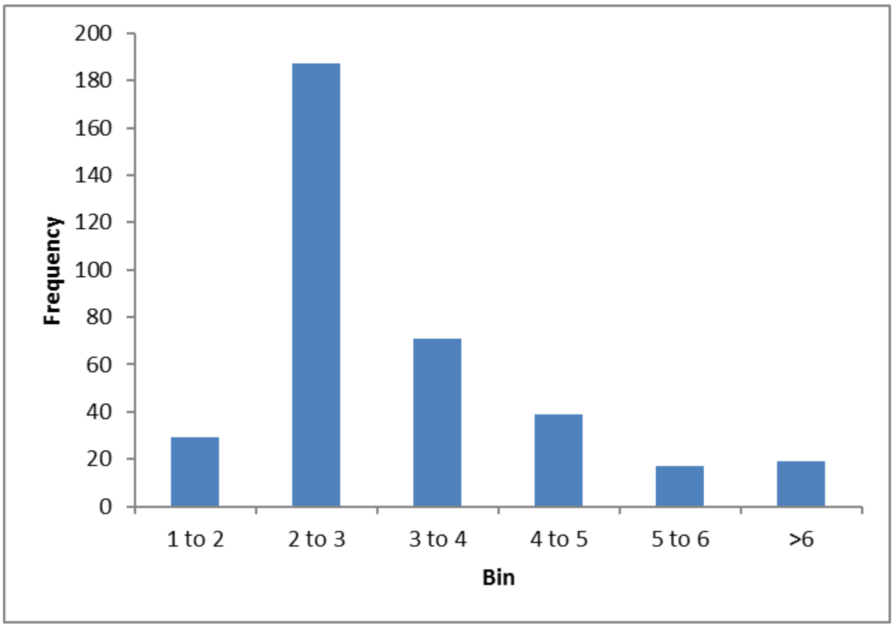


Figure 5.4: Yorkshire Load Profile Kurtosis Values

For the Yorkshire 2011 profiles, Fig. 5.4 shows there is a split distribution where less than half of the profiles have a kurtosis value greater than three. This does suggest that using kurtosis as an indication of outliers is useful since the majority of profiles have a kurtosis value smaller than the normal distribution. If instead the majority of cases had a kurtosis value greater than 3 then the flag would be less effective for the NPG practitioner.

5.1.4 Comparison of Methods

The advantages and disadvantages of changing methods from NPG's current methodology to the new wavelet outlier detection method will be considered here.

The main advantage of changing methodologies is that there is more room for automation in the new method. It is capable of removing short outlying periods that a user might previously have had to check manually. If used in conjunction with a load transfer shifting algorithm, that automation could further be increased. Even if human judgment is taken into consideration by considering a threshold value similar to NPG's current methods, the number of stations that need checking are reduced.

By using the kurtosis condition, substations that are prone to outliers or load transfers can be flagged immediately. This can provide a basis for immediate inspection but can also allow for those cases to be stored for future reference.

In the current method where the load duration curve is used, the past years MD value is relied upon to set a threshold value. Relying on this in itself could potentially introduce error under the assumption that the previous years calculation is accurate. Even if the previous years value is accurate, there could still be erroneous values within the threshold boundaries. Although the error introduced may not be large, relying on these threshold boundaries could mask some of the underlying outliers and load transfers. The new method does not rely on the previous years calculated value and attempts to locate outliers. By using the kurtosis value as a flag, the shape of the current load distribution is considered rather than relying on a historical calculated value.

The main disadvantage is that the current methodology is well understood and used so it may not be worth the time and money to change systems. If the amount of time saved by switching to a more automated method is not substantial enough then there may not be enough incentive to switch.

Even if a more automated approach is undertaken, there will still be a few erroneous sites which have an error that is too large to ignore. As an academic exercise, a few substations which have very high errors may not matter but for regulatory purposes, NPG cannot have any major avoidable errors. Thus, there would still need to be some human judgment involved even after adopting a more automated approach.

By looking at the error distribution graph in Fig. 5.5, more MDs are overestimated by the new automated algorithm in comparison to the MDs from NPG's methods. This would not necessarily be an issue in terms of reinforcement as some

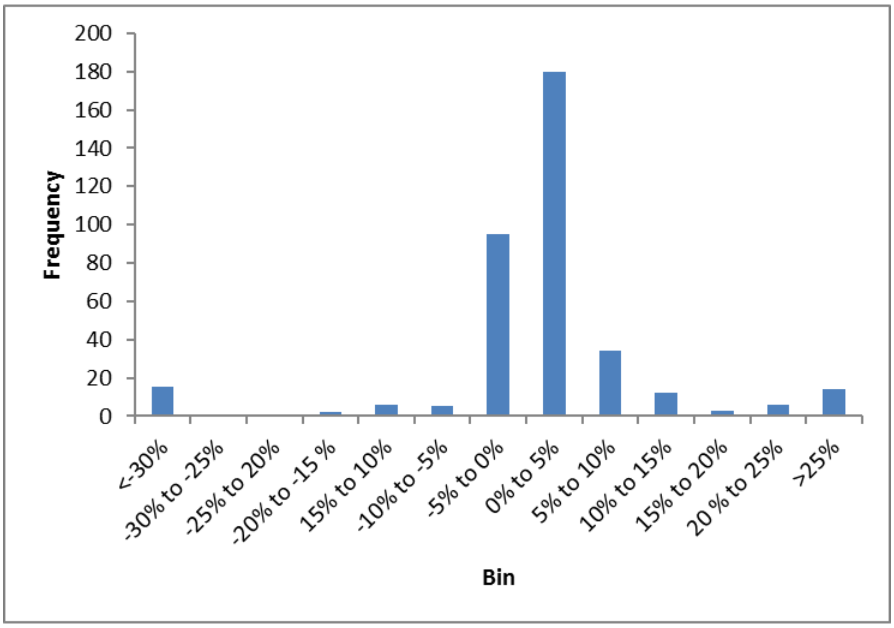


Figure 5.5: Yorkshire MD Error Distribution Between NPG DLE MD and Automated Method

overestimation would lead to a few substations being flagged for reinforcement even though it may not be necessary. In this case, the SCADA data would be considered more closely to determine if reinforcement is needed on that particular substation.

5.1.5 Initial Look at Load Transfers

In the documentation for determining maximum demand (CE Electric UK, 2009), there are five example cases of types of load profiles and the determination of maximum demand for each of them. One of these examples is a case where a load transfer temporarily occurs as a result of work on the network.

Consider the load profile in Fig. 5.6. There are two instances on this demand profile where there is a large deviation from the overall trend. These would most likely be categorised as load transfers and would not be taken into account when determining the maximum demand because it does not reflect the usual activity at the substation. In the following method, there are two main parameters which can be adjusted. One is the difference between two successive local minima and maxima which is stated here as 2σ. In this case, σ is the standard deviation of the load profile demand values. A short experiment is conducted later in this section to show

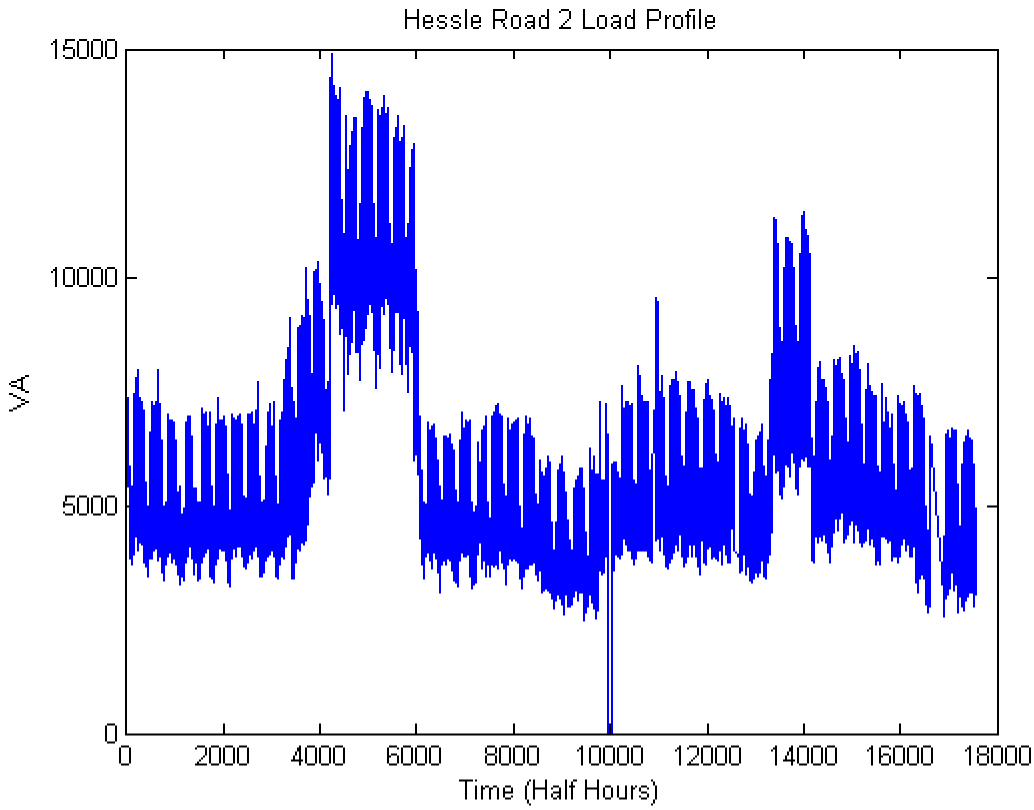


Figure 5.6: Hessle Road 2 Load Profile

that this is an appropriate value. The other adjustable parameter is the 3 days condition. This a subjective choice and is made on the basis that a time period less than 3 days is more likely to be simply random demand variations or outliers. If necessary, this parameter can be varied to suit the needs of the NPG practitioner. If both the 2σ and 3 day condition are fulfilled, then this method will treat that section of the load profile as a load transfer.

The following algorithm is an automated process for determining cases of load transfers in demand profiles.

1. Input the substation half hourly profile.

2. Find and calculate the daily mean demands for the profile.

3. Smooth the profile using a 10 point moving average.

4. Calculate the local minima and maxima of the curve.

5. If the difference between one local extrema i and the following extrema $i + 1$ is greater than 2σ and they are at least 3 days apart then:

   (a) Compare the next pairs of extrema $i + n$ and $i + n + 1$ where n ranges from 1 to # of extrema − 1 until a difference of 2σ is found. If such a pair exists then the load transfer is from i until $i + n + 1$.

   (b) If a load transfer is found, start the if condition from $i = i + n + 1$. Else, let $i = i + 1$. Continue from if condition until the last difference is checked.

Initially, the algorithm uses a daily mean demand curve which is smoothed by a process of moving averages. This helps to eliminate the shorter demand spikes in the profile which should not be considered when determining load transfers. The main process of the algorithm is in examining the difference between one local extrema and the following one. The extrema are found by If this difference is sufficiently large (in this case 2 standard deviations apart) then it is said to be the "start" of the load transfer. Afterwards, it is expected that the curve would return to its normal trend after some period of time so there must be a similar difference between two other extrema in the future. If both of these conditions are met, then there is a load transfer. This section will examine if the chosen threshold value of 2σ is appropriate.

Experiment For Load Transfer Method

The experiment to be carried out is the effect of threshold parameter 2σ on the accuracy of the load transfer detection algorithm. As there exists no strict definition or previously compiled load transfer log, a random selection of load profiles will be inspected prior in order to determine if there is a load transfer. The parameter to be varied is the multiple of standard deviation in the algorithm which determines the separation between local minima and maxima. The confusion matrix of these threshold variations will be presented along with the accuracy metric.

When judging if a load profile has a load transfer, missing data that lasts for more than one day will be taken to be a load transfer even though it may be a fault

Table 5.1: Confusion Matrix Values for Sigma Threshold Parameters

	TP	FP	FN	TN	ACC
1σ	32	0	13	5	74%
2σ	29	3	2	16	90%
3σ	23	9	1	17	80%

with SCADA instruments and not necessarily a load transfer. The algorithm does predict the time periods during which the load transfers take place but the accuracy will be guaged by a "yes" or "no" definition for the sake of simplicity. Table 5.1 shows the confusion matrix values for three possible threshold parameters.

The accuracy (ACC) of the 2σ threshold is $\frac{29+16}{29+3+2+16} = 90\%$. This threshold proved to be quite successful at detecting load transfers and had the highest accuracy justifying its use in the load transfer method.

The accuracy (ACC) of the 1σ threshold is $\frac{32+5}{32+0+13+5} = 74\%$. Lowering the threshold by 1 standard deviation does allow for the substations where there is a load transfer to have full accuracy due to the less stringent condition. However, this threshold value incurs more error due to the substations where there is no load transfer now being defined as having a load transfer.

The accuracy (ACC) of the 3σ threshold is $\frac{23+17}{23+9+1+17} = 80\%$. Increasing the threshold by 1 standard deviation gives a stricter condition on that which defines a load transfer. However, there were only 2 instances in the 2σ case where profiles that did not have a load transfer were incorrectly predicted as having one. Thus increasing the threshold provided little benefit and incurred more error by incorrectly classifying the substations that did have a load transfer.

Although it is difficult to define what is meant by a load transfer without a specification from the operators, with proper experimentation a reasonable method for detecting sharp changes was made. It was found that many substations had missing data over a long period of time which had to be defined as load transfers even though it is more likely that those periods are simply faults with the SCADA equipment. The biggest obstacle to creating an accurate algorithm in this instance was the dataset itself. Many substations had missing or erroneous data that may have affected where a load transfer was defined to start or end. Even if there is a strict definition for load transfers, the inaccuracy of the SCADA data would require the user to examine every questionable substation regardless.

5.2 Base Load Profile Determination

By combining the techniques used in the maximum demand algorithm and load transfer detection, the base load profile can be determined. The base load profile will shift outliers and load transfer periods such that they fall within the expected distribution of the profile. The outlier detection method is updated with a data replacement technique and the load transfer algorithm is also refined and updated here to account for more situations and analysed further using network diagrams.

5.2.1 Median Replacement

The methodology applied in the maximum demand algorithm will also be used here for outlier detection. The algorithm starts by doing the following two steps:

1. Input the substation half hourly load profile.

2. Find and store the maximum demands for each day in the load profile.

In the maximum demand algorithm, daily maximums were used to perform analysis which could be done because only maximums were of concern in that problem. However, here the raw data should also be considered because of the effect of using daily data. The disadvantage to using daily data is that when replacing outliers in the raw data, the entire day is replaced which will consist of 48 data points. Continuing to use daily data could still be advantageous as it will be able to isolate outlying days more easily than individual half hours in the raw data.

Then there is an option of smoothing the profile using moving averages. The advantage of using a smoothing process is that the longer patchier outliers will be detected more easily as the change in value will be sharper. The disadvantage of using smoothing is that the short erroneous spikes will be missed because the smoothing process will significantly reduce them.

Therefore, the smoothing process will be missed in this algorithm because of the load transfer algorithm which will be used in combination with the outlier detection method. The load transfer algorithm should pick up on the cases that the initial outlier detection method misses since it will handle sudden shifts in values that last for an extended period of time.

For the purposes of determining maximum demand, the outliers could be completely removed from the profile. However, if any further analysis needs to be carried out with the base load profile, eliminating data should be avoided. There must be consistency especially if comparison and analysis is carried out with other data sets. Therefore, instead of eliminating outlying points, they should be suitably replaced so that a consistent and continuous analysis can be carried out in the future.

As with detecting outliers, there are many different methods that can be chosen as a replacement technique. Menold et al. (1999) considers some of the possible techniques that can be used and their relative advantages and disadvantages. One possible technique is to use linear or nonlinear regression to a selected window. The disadvantage of these filters is that, dependent on the window size, they could be skewed by other outliers. In any of the techniques that could be considered, there must be a consideration for patchy outliers and the window size of values considered. Another possible technique is to take the last valid load value to be taken as the replacement. This may actually work well in certain cases but not as well in others. For example, in a monotonic sequence, taking the last valid value could be a good technique but in a periodically oscillating sequence the last value may be a poor indication of a correct value.

Here, this novel technique used will be to replace the detected outlier with the median value of loads in the same half hour period within a month interval of the outlier day. The median value is used in robust statistics as a measure of central tendency because it is not as susceptible to outliers unlike the mean. The main adjustable value here is considering the cases where the monthly time window around an outlier cannot be considered. A monthly time window for most days is considered to be 15 days before and 15 days after the outlying day. However, for the boundary cases before day 15 and after day 350, a different distribution is taken which is done in steps 3 and 4 in the following method. This monthly time period for median replacement is also subjective and can be adjusted dependent on what the NPG practitioner requires. After the wavelet method detects outliers, the outliers shall be replaced as follows:

1. Sort the daily outlier list in ascending order.
2. For each outlier Day t do the following:
3. If the outlying day occurs on or before Day 15 then:
   (a) Find difference $= 15 - \{\text{day number}\}$.
   (b) Have the daily distribution include Day 1 to Day $t + 16 +$ difference.
   (c) Remove Day t.
4. If the outlying day occurs on or after Day 350 then:
   (a) Find difference $= \{\text{day number}\} - 350$.
   (b) Have the daily distribution include Day $t - 15 -$ difference to Day 365.
   (c) Remove Day t.
5. Otherwise the daily distribution is Day $t - 15$ to Day $t + 15$ excluding Day t.
6. Store all the half hour periods in the day distribution (to relate back to raw data).
7. Store all the half hour indices by taking the half hour modulus 48.
8. Find the median of all the half hour periods with the same index in the raw data.
9. Replace each half hour period in Day t by their median values.

Menold et al. (1999) also briefly considers the window width to select when using outlier detection and replacement. There is no exact method for a good choice but in principle it is good to not take too large a window size so that the local dynamics of the problem are best considered. However, it is also good not to take too small a window due to successive patches of anomalous data. This is especially apparent in much of the NPG primary substation data where there are many cases of missing values that last several half hour periods.

5.2.2 Updated Load Transfer Algorithm and Shift Technique

The wavelet method for detecting outliers performs well at detecting sharp changes in data. However, it may still miss out on load transfers that occur for an extended period of time. Consider the following load profile:

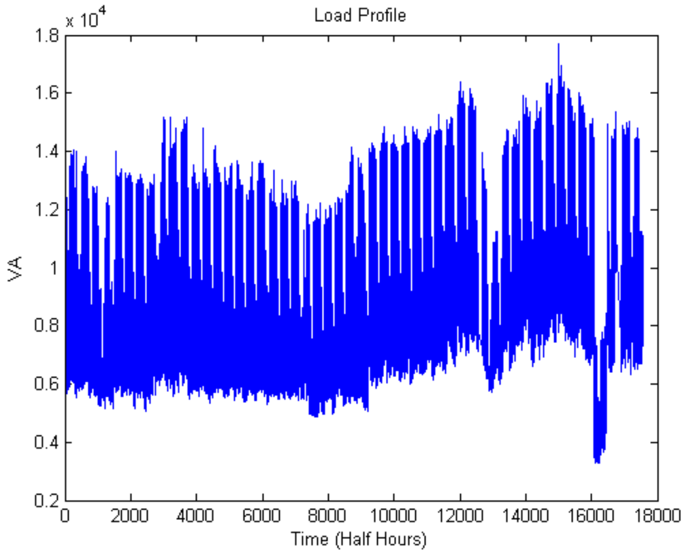


Figure 5.7: 2011 Arundel Street Load Profile after Outlier Detection

The outlier detection method was applied to this profile but it was not able to detect the load transfer which is clearly occurring towards the end of the profile. Therefore, it is worth using a load transfer shift method in combination with an outlier detection method. This novel load transfer algorithm has been revised and is given as follows:

1. Input the substation half hourly profile.

2. Find and calculate the daily mean demands for the profile.

3. Smooth the profile using a 10 point moving average.

4. Calculate the local minima and maxima of the curve.

5. If the difference between one local extrema i and the following extrema $i+1$ is greater than 2σ and they are at least 3 days apart then:

   (a) Store the remainder of the profile after the minima/maxima point.

   (b) If the load profile reaches a point where the extrema is exceeded then the load transfer ends at that point.

   (c) Else if it does not exceed the extrema, find the maximum of the remaining extrema points.

   (d) Check if that point is within 10% of the original extrema and if so, the load transfer ends at that point.

   (e) Go to the end of the load transfer period.

6. If neither of the above conditions is met, there is no load transfer. Continue checking extrema points until the entire profile is checked.

7. Store the difference related to the daily period.

8. Add or subtract the difference from the half hourly data where the first point is at the start of the first day of the load transfer (00:00) and the final point is at the end of the final day of the load transfer (23:30).

In the original algorithm after a change of 2σ was detected, a corresponding change of 2σ would be required in the opposite direction. This would often lead to inaccuracies if the actual end to the load transfer period happened gradually. In some cases where there were multiple load transfers, the period that the load transfer ended could have been far removed from the actual end. This updated version does not completely solve the load transfer problem but is able to handle more cases effectively.

Also in the original algorithm, the load transfer period was simply deleted but as with the outlier algorithm, a replacement technique is preferred here for consistency purposes. The use of daily MD data also causes some problems to arise when converting back to half hourly data as with the outlier algorithm. However, here the conversion is much more sensitive because of the fact the choice of replacement technique. That is, some half hour periods could be shifted at the start or end of the load transfer which are actually normal periods. Multiple replacement techniques were considered such as using a similar median filter or linear regression between two valid periods. The problem with using the median of some selected values is that the inclusion of the load transfer period in the window could make the filter very ineffective. However, if it is excluded, the problem then becomes what period of time should be chosen as a replacement. This window of time would also become more unrepresentative the longer the load transfer period is as more points start to become excluded. The use of linear or nonlinear techniques to replace points would also have similar issues at the end points of the half hourly data because of the use of daily data. The use of raw data was also considered but in order to detect sharp changes, using daily smoothed data is much more effective. In order to overcome the issues associated with using the daily data, the combined usage of the load transfer algorithm and outlier detection algorithm is done to compensate for their respective weaknesses.

There are many methods available in the literature regarding outlier detection and replacement and even some discussion on "patchy" outliers. Yet, there is little discussion on periods of time such as load transfers because in many instances they would not be considered anomalous data. Even here, there is an explanation as to why they occur but for the purposes of analysis, they should be ignored or shifted to be consistent with the rest of the data. Load transfers commonly occur when there is work on the network that causes load to move from one substation to another. Therefore, in conjunction with an algorithm, system diagrams should be investigated to see if the load transfer can be explained from a corresponding increase or decrease in connected substations.

A few instances of load transfers were investigated by looking at substation connections under a grid supply point. Using Arundel Street as an example, there were also decreases around the same time period for two other substations: Park Hill and Victoria Street. However, no corresponding increase was found within the same group of substations. It is possible that the increase could be occurring in a

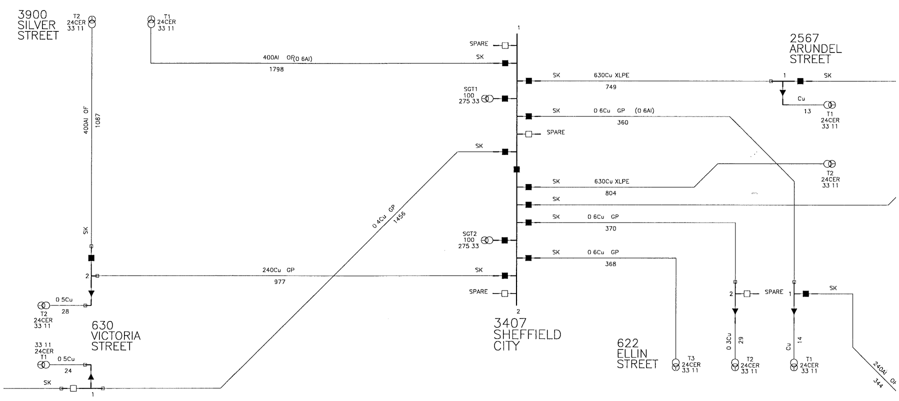


Figure 5.8: Network Diagram Around Arundel Street

substation not under the same supply point, but that makes it much more difficult to trace. While there is definitely some relationship between substations regarding load transfers, tracing it in a systematic way may overcomplicate the algorithm and introduce more error. Instead, there should be focus on what constitutes a load transfer given a yearly profile and being able to detect it.

Experiment on Effectiveness of Outlier Detection and Load Transfer Method

The first experiment will be gauging the effectiveness of the outlier detection and median replacement method. This will be done by calculating the kurtosis before and after the outlier detection. The kurtosis is calculated on the annual demand profile on substations in the Yorkshire region as in Equation 5.1.1. The kurtosis should give an indication to the peakedness and size of the tails. The percentage change in kurtosis will be measured between -5% and 5% in intervals of 1% and a histogram plotted. Both the raw half hourly data and daily maximum demand data will be considered. Then the data will be broken down into the 2011/12 and 2012/13 periods for a further breakdown of results for both the daily method and raw data method.

The second experiment will be gauging the effectiveness of the load transfer detection method together with the outlier method. That is, the load transfer algorithm will be run first with the outlier algorithm running immediately afterwards. Again, the effectiveness will be measured by calculating the kurtosis before and after the algorithm. The percentage change in kurtosis will be measured between -5% and

5% in intervals of 1% and a histogram plotted. Then the data will be broken down into the 2011/12 and 2012/13 periods for a further breakdown of results.

As the kurtosis is used in the maximum demand method to determine whether or not a substation is outlier prone, using it as a measure to determine the effectiveness of the method is appropriate here. The methods are deemed to be effective if there are more instances where the kurtosis value of the load profile is decreasing after performing the outlier replacement or load transfer shifting method. By plotting these kurtosis changes in a histogram form, a good performance would have higher frequencies towards the left (negative values) of the graph.

In both cases, the dataset to be used will be all of the Yorkshire primary substation half hourly demand data from April 2011-March 2013.

Outlier Detection and Replacement Using Daily MD Data

The outlier detection and replacement method was first applied to all of the Yorkshire primary substations using the daily maximum demand method which replaces 48 values in an outlying day.

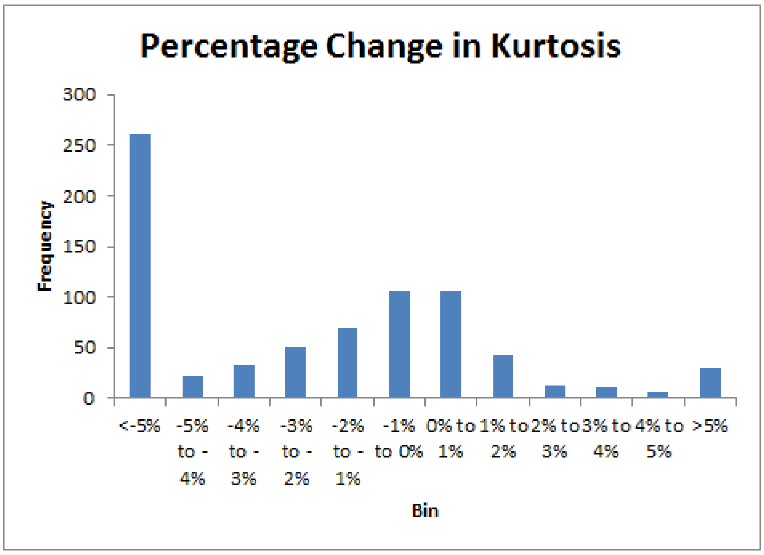


Figure 5.9: Percentage Change in Kurtosis for 2 years

The histogram of kurtosis changes for both years is given in Fig. 5.9. In the case for the use of this algorithm, there are more cases where the kurtosis decreases indicating a decrease in the size of the tails of the distribution. The highest bin range in the histogram also occurred where the kurtosis decrease by more than 5%. An argument could also be made against the algorithm in that there were a significant number of substations where the kurtosis increased.

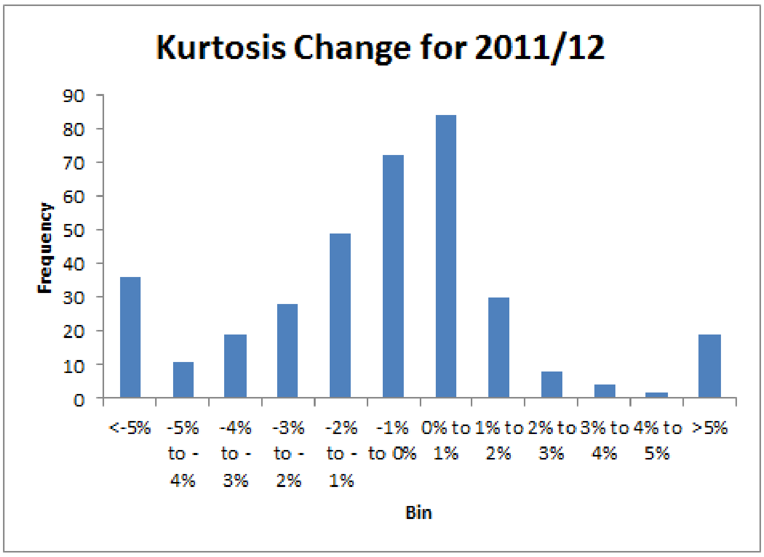


Figure 5.10: Percentage Change in Kurtosis for 2011/12

Afterwards, the same analysis was carried out for each individual year in order to see a breakdown of the aggregated results. For April 2011-March 2012, the results are given in Fig. 5.10 The algorithm seems to perform satisfactorily in this year as there are more cases where substations undergo a decrease in kurtosis.

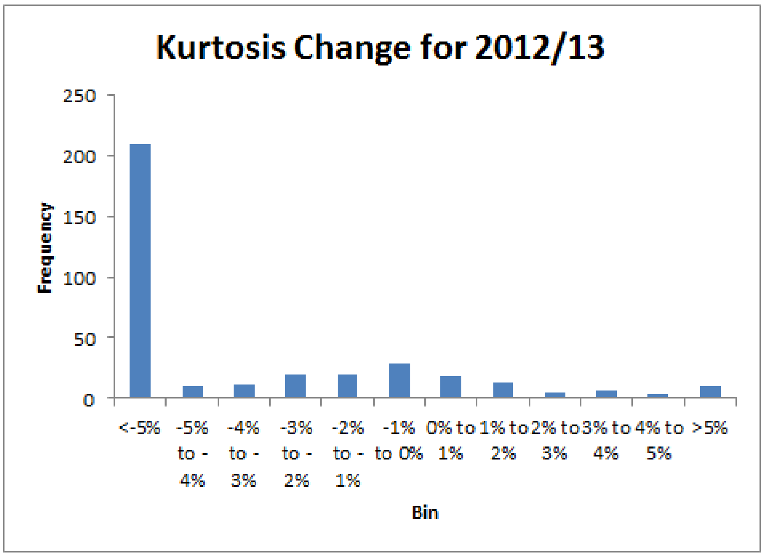


Figure 5.11: Percentage Change in Kurtosis for 2012/13

For April 2012-March 2013, the results are given in Fig. 5.11. This result proved to be the most surprising. For this year, the overwhelming majority of cases underwent a decrease in kurtosis upon applying the algorithm. These results suggest there is a fundamental difference in the more recently gathered data and older historic data. The algorithm proved to be very effective for the most recent data but still of overall benefit for both years.

Outlier Detection and Replacement Using Raw Half Hourly Data

The outlier detection and replacement method was then applied to the Yorkshire primary substations using only raw data.

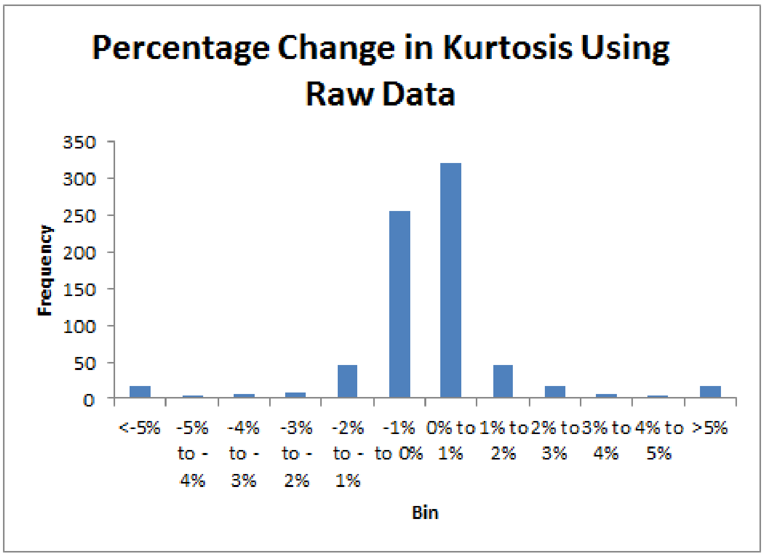


Figure 5.12: Percentage Change in Kurtosis for 2 years Using Raw Data

The histogram for both years is given in Fig. 5.12. Unlike when using the daily data to replace outlying days, the change in kurtosis value here is minimal as expected because fewer values are being updated. The way in which the wavelet method detects outliers could also lead to fewer changes. The method is good at detecting changes in data but if there are several half hour periods leading up to an outlier, the method may only detect a few of those points as being outliers.

When looking at both years, there are more substations which are undergoing an increase in kurtosis after the algorithm which suggests the algorithm is not very effective. The same analysis was then carried out for both individual years to see a breakdown of the aggregated results.

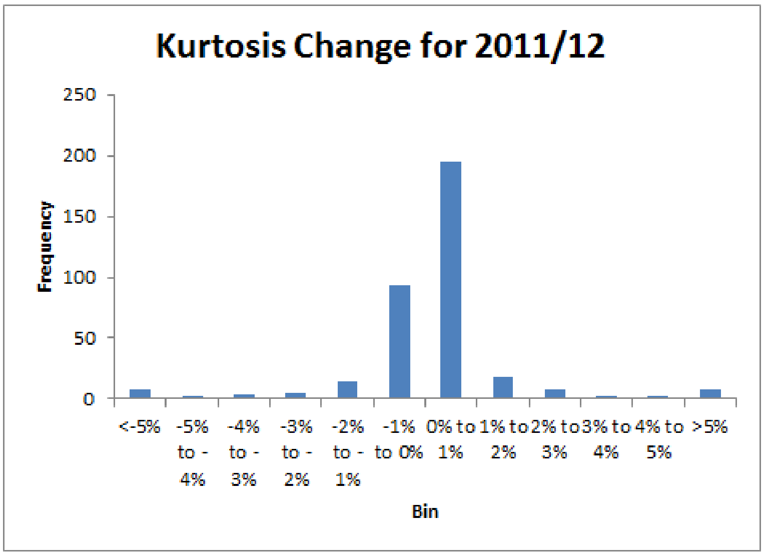


Figure 5.13: Percentage Change in Kurtosis for 2011/12 Using Raw Data

For April 2011-March 2012, the results are given in Fig. 5.13. Here, there are more substations where kurtosis is increasing showing the algorithm is not very effective for this year. Also, the percentage changes are mostly within $\pm 1\%$. As stated earlier, this is expected due to fewer values being updated in comparison to using daily data.

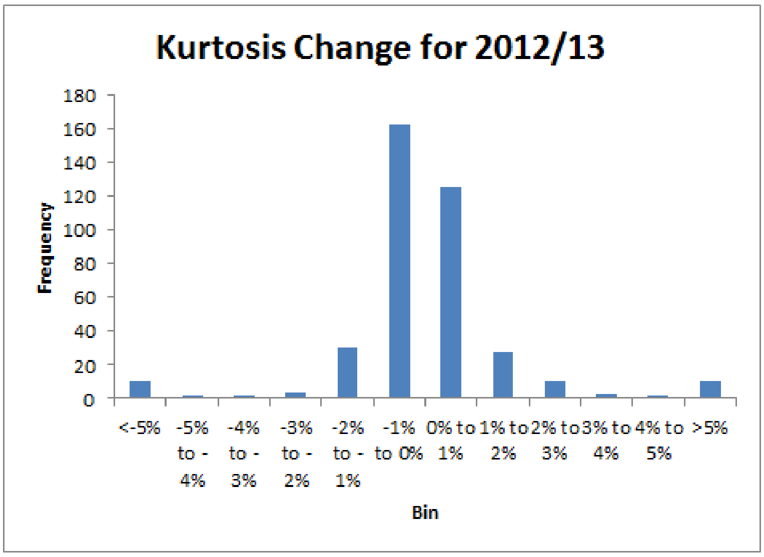


Figure 5.14: Percentage Change in Kurtosis for 2012/13 Using Raw Data

For April 2012-March 2013, the results are given in Fig. 5.14. For this year, there are more substations which undergo a decrease in kurtosis suggesting the algorithm is more effective for this year just as the daily data method was more effective for the 2012/13 period. However, there are still a significant number of cases where the kurtosis increases so the improvement is marginal.

Due to the minimal changes that result from using the raw data only, the daily MD data method will be used in the future. The daily method proved to be extremely effective for last years data and although it was not as effective for the 2011/12 period, there were still more cases where the algorithm reduced kurtosis rather than increased it.

Load Transfer Detection and Shift with Outlier Detection

The histogram for percentage change for both years during 2011-2013 using a combined load transfer and outlier technique is given in Fig. 5.15.

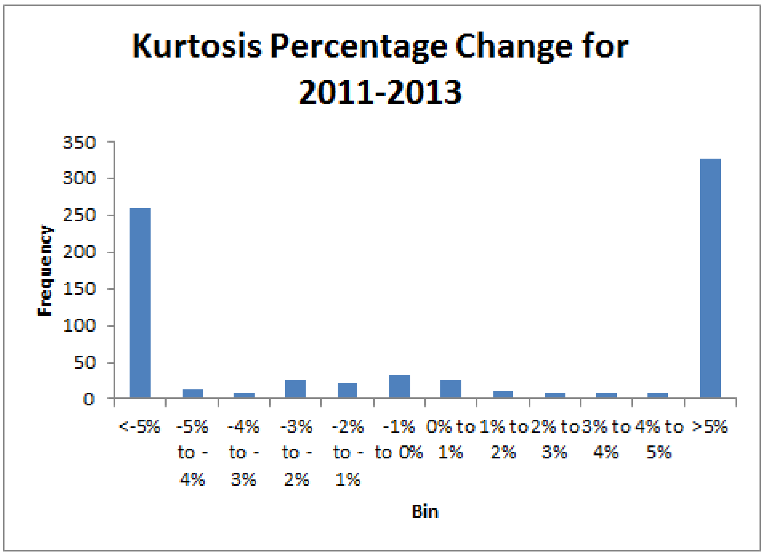


Figure 5.15: Percentage Change in Kurtosis for 2 years Using Combined Approach

The histogram suggests this combined approach performs much more poorly than simply using the outlier method alone. This suggests that the load transfer algorithm is incorrectly shifting certain periods of time and introducing more outliers.

The results are further broken down into the 2 years. The histogram for 2011-2012 is given in Fig. 5.16. In this year, the results are similar as for both years. There are a significant number of profiles which undergo an increase in kurtosis of

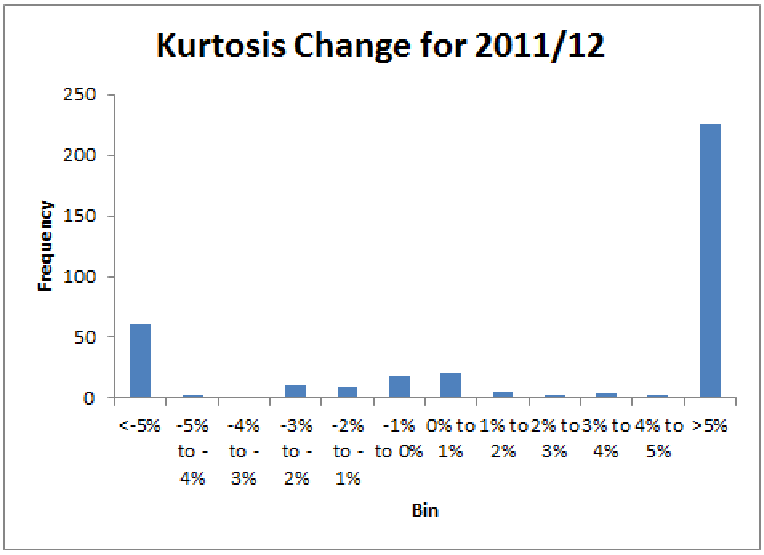


Figure 5.16: Percentage Change in Kurtosis for 2011/12 Using Combined Approach

more than 5% suggesting the algorithm is introducing more inaccuracies.

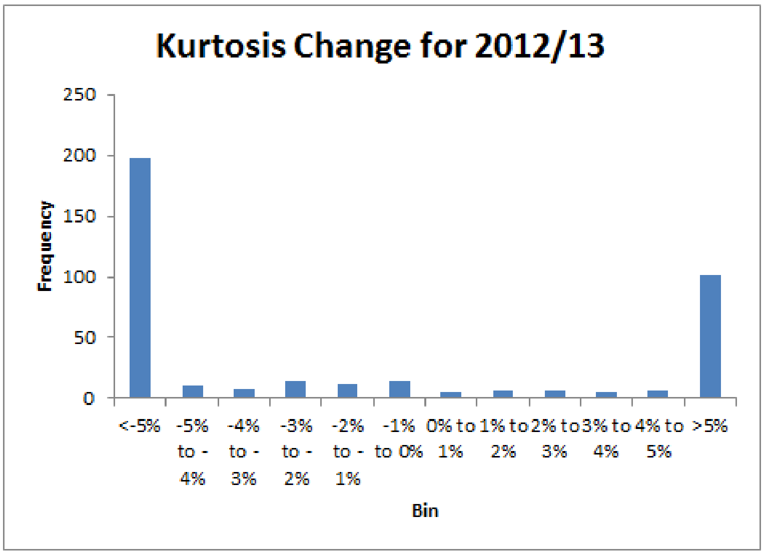


Figure 5.17: Percentage Change in Kurtosis for 2012/13 Using Combined Approach

For 2012-2013, the results are given in Fig. 5.17. In this year, there are actually more cases where the algorithm seems to improve upon the original load profile. However, in comparison to simply using the outlier detection method alone, more errors are introduced and there are significantly more cases where the kurtosis increases by more than 5%.

5.3 Conclusion

The main research question addressed in this chapter was RQ1 which is brought to a close here. Namely, internal demand variation was considered here in terms of anomalous data and the key method for determining maximum demand for NPG. For both techniques of outlier median replacement and load transfer shifting, simply removing anomalous data was easier to accomplish without introducing further complications. When data must be replaced or shifted, there is a potential for new problems to arise as was the case with some of the profiles here. Overall, the outlier detection and replacement method worked well by taking the median values of half hours around the period. This worked especially well with short outlying periods as opposed to patchy outliers lasting several days. The load transfer algorithm attempted to solve this issue as well as load transfer periods indicated by sudden changes in data. The problems arose from the need to use daily maximum demand data to find the load transfer period. This caused inaccuracies to arise regarding the exact start and end in the half hourly data. As well as this, the choice of 2σ as the threshold for change between two extrema still proved to be to arbitrary a choice. While this threshold worked on some profiles, it proved to be too sensitive on others. System diagrams showing connections between substations were also considered but they proved to be difficult to trace. It is possible that relying on these connections in an algorithm could introduce more error.

The outlier detection method will certainly be used in the future as a preprocessing technique before analysing data. The detection of load transfers is much more difficult however, and further research and techniques would need to be tested to improve upon it. Defining a load transfer purely based on the load profile is subjective and NPG currently manually checks many of these instances. In the CL-ANFIS method, the outlier detection method using daily MD will be used as it proved the most effective. The automated MD determination method will also be used in Chapter 7 in order to perform a comparison between the CL-ANFIS and the current NPG methodology.

Chapter 6

CL-ANFIS Forecasting Algorithm

After conducting these preliminary analyses on external data sets and the provided NPG load data in Chapters 4 and 5, an informed forecasting load algorithm was created. In this chapter, the novel contribution CL-ANFIS (Customer-Led ANFIS) is presented and discussed in all of its components. The novel approach of CL-ANFIS arises from a combination of the methods used and the resulting outputs. CL-ANFIS uses a combination of unsupervised learning and forecasting to create tailor made forecasts for different groups of substations. Then using the membership functions of ANFIS and clustering, qualitative attributes can be given to different groups of substations. CL-ANFIS will use the methods in Chapters 4 and 5 and expand upon them to develop a complete forecasting methdology. Section 6.1 details the motivations of the work. Section 6.2 gives the daily load profile clustering approach along with ANFIS. Section 6.3 gives an overview of how the load data is fed through the CL-ANFIS process. Section 6.4 goes through initial results of the CL-ANFIS by different clusters. Section 6.5 goes into greater depth of the forecasting errors using the demographic information from Section 4.3. Section 6.6 justifies the use and importance of qualitative rules and how they can be extracted from CL-ANFIS. Section 6.7 gives a more detailed analysis of the demographic customer labels defined in Section 6.6. Finally, Section 6.8 concludes the chapter with some thoughts on the results.

6.1 Motivations for CL-ANFIS

The need for electrical distribution companies to create an informed long-term annual forecast has increased due to the rapid change to a green economy. NPG in the UK is one such DNO interested in a more informed forecast. Currently, NPG produces a long-term annual forecast up to 8 years in the future for each of their primary substations in the North East and Yorkshire based on simple growth rates. Using simple growth rates has been partially influenced by their work showing that assuming the previous year's maximum demand provided similar or better results than their current process (CE Electric UK, 2009). However, this gives no insight into the drivers of demand or informs the infrastructure planning process. The problem is greatly driven by the scale of the forecasting task at the distribution level. On primary substations which operate at 11/33/66 kV levels, there can be hundreds of customers, some of which will have different energy usage patterns. This makes the problem difficult to address but by using the customer information supplied from NPG, some greater insight can be achieved by using the CL-ANFIS.

In terms of forecasting electricity demand, the literature is more focused on developing short-term forecasts. The motivation for the short-term forecast comes from a need to create point forecasts for controlling and scheduling power systems (Taylor and McSharry, 2007). Pricing markets around the world including UK, Spain, California and Colombia are based on day-ahead auction (Bunn, 2000). This requires the day-ahead forecasts to be as detailed and accurate as possible. The methods used focus on statistical techniques such as linear regression, Box-Jenkins, and ARIMA (Amjady, 2001; Zhang, 2003). They are often compared to more recent developments such as support vector machines (SVM) or artificial neural networks (ANN) (Chen et al., 2004; Hippert et al., 2001; Park et al., 1991). These consider which technique is more effective on a purely objective basis. An informed forecast using other factors such as type of day, economic or demographic information has become more common especially in relation to smart grid research (Ho et al., 1990; Mohamed and Bodger, 2005).

In this thesis, the long-term forecast is defined as a forecast that predicts demand at least 1 year ahead. The motivation for the long-term forecast comes from the need for capacity planning and providing justification for expanding the network. With additional load and generation connections made over a period of months and years, long-term forecasting becomes necessary. Particularly in the case where demand exceeds planned capacity, a decision to reinforce can be made. The methods used for long-term forecasting focus on density forecasts where estimates of probability distributions are given for future values (Gaillard et al., 2015; Hong et al., 2014, 2016; Hyndman and Fan, 2010; McSharry et al., 2005). In some markets, this probability distribution is necessary as in the case of the National Electricity Market (NEM) of Australia (Hyndman and Fan, 2010). This gives access to some additional information where there is uncertainty in the forecast. Additionally, demand side management has been used on long-term peak load values (Albadi and El-Saadany, 2008).

The requirement of producing a forecast for this project around a single annual maximum demand figure has driven the consideration of methods which are also used in short-term forecasting rather than a probabilistic method. This has led to a choice of the adaptive network based approach to fuzzy inference systems (ANFIS) (Jang, 1993; Sumathi and Paneerselvam, 2010; Takagi and Sugeno, 1985) as a forecasting tool as it offers the non-linear forecasting capabilities of ANN or SVM while offering some greater insight to its functionality through fuzzy membership functions (Chen et al., 2014, 2015).

6.2 Theory and Methods Used

In order to develop the forecast, clusters of substations are based on daily load profiles which is explained in subsection 6.2.1. This aims to provide supplemental knowledge and when combined with demographic information, can be used in conjunction with the forecast. Then the forecasting is carried out using ANFIS which is explained in Section 6.2.2. This is chosen as a non-linear forecasting tool where some transparency can be gained through the use of membership functions which can be linked to the demographic knowledge.

6.2.1 Creation of Representative Load Profiles

The justification for the use of K-means clustering is presented here based on work by Akperi and Matthews (2014a). The methodology will compare the clustering techniques: K-means, hierarchical and fuzzy C-means. In addition, it will make use of the principal component scores gained in Chapter 4 to draw further comparisons.

Justification of Clustering Technique - Method

Initially, each clustering technique will be analysed separately and representative load profiles for the summer period are considered. There are multiple ways a normalisation process can be done but here the following will be used. Each substation on the network has an annual demand profile consisting of readings taken every 30 minutes. The summer period is defined as June 1 2012-August 31 2012 for a total of 4416 readings. According to the investigation in Western Power Distribution (2013), it is more appropriate to normalise each daily load profile and then average over a period of time as opposed to averaging the period first and then normalising. This helps to highlight daily load patterns and is not as subject to seasonal variation. Normalising by the maximum of each day and then averaging each half hour point over the summer period is the procedure followed here to attain the representative load profiles. Therefore, the data used here for each of the clustering algorithms consists of 576 daily load profiles for each substation over the period of June 1 2012-August 31 2012 which are created by this normalisation and averaging process.

Before the clustering is done, the number of K clusters needs to be determined for each method. Depending on the criteria used, a different number of chosen clusters could be indicated.

In order to attain supporting evidence for the engineering based explanations of the cluster solutions, clustering validity metrics will be used. These metrics can be used to ascertain both the effectiveness of one clustering algorithm versus another and to determine the number of clusters that should be used. In this context, the metrics will be used to determine how many daily load profile groupings are appropriate for CL-ANFIS. For crisp clustering of K-means and hierarchical, a couple of popular indices are used here: Dunn's index and Davies-Bouldin.

Dunn's index is a popular method in the literature and identifies compact and separate clusters (Dunn, 1974). The Dunn index T is defined as

$$T = \min_{1\leq i\leq c}\left(\min_{1\leq j\leq c, j\neq i}\left(\frac{\min_{x_i\in X_i, x_j\in X_j}(d(x_i,x_j))}{\max_{1\leq k\leq c} y_k}\right)\right) \tag{6.2.1}$$

where $x_i \in X_i$ is a cluster group of vectors, c is the number of clusters used, $d(\cdot,\cdot)$ is the standard Euclidean distance metric, and $y_k = \max_{x_l,x_m\in X_k} d(x_l,x_m)$. In this context, x_i are the daily load profiles of the substations which are 48 dimensional vectors (48 half hour periods in one day). The best solution is the one with the highest Dunn's index.

The Davies-Bouldin (DB) criterion is defined as a ratio of within cluster and between cluster differences (Davies and Bouldin, 1979). Suppose S_i and S_j are dispersion measures which are the average distances between each point in the clusters and their respective centroids. That is, for cluster i, S_i is the average Euclidean distance between every daily load profile in the cluster and the centroid daily load profile after K-means. M_{ij} is the Euclidean distance between the ith and jth clusters using the centroid daily load profiles. Then the DB index $\bar{R}$ is

$$\bar{R} = \frac{1}{N}\sum_{i=1}^{N} max_{i\neq j}\{R_{i,j}\},\ R_{i,j} = \frac{S_i+S_j}{M_{i,j}}. \tag{6.2.2}$$

The best clustering solution is the one with the smallest DB index.

For fuzzy clustering, the allowance for partial membership requires a different clustering validity metric. One of the most popular metrics in use is the Xie-Beni metric (Xie and Beni, 1991) which is defined as

$$S = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n} U_{ij}^2\|V_i - x_j\|^2}{n\min_{i,j}\|V_i - V_j\|^2} \tag{6.2.3}$$

where c is the total number of clusters, n is the number of vectors (in this case substations), U_{ij} is an entry in the fuzzy membership matrix, V_i is a cluster centroid, and x_j is a daily load profile. The best solution is one with the lowest Xie-Beni index.

There are three main pieces of information that will inform the discussion for each method. First are the cluster centroids and how the load profile changes over the course of the day. The cluster centroids are generated directly for K-means and fuzzy C-means. For hierarchical clustering, the profiles belonging to each cluster can be averaged to generate a centroid.

Second are the population sizes of the clusters. These are easily gained for K-Means and hierarchical clustering. For fuzzy C-means, because of partial membership, a different method must be used. The membership values across the membership matrix U are summed to attain population sizes.

Third are the principal component values from Section 4.3 for each cluster centroid. This allows for an idea of the customer make-up to be understood for each cluster. For K-means and hierarchical clustering, these are the average values of PC1 and PC2 for each cluster. Associating the customer classification principal components from Section 4.3 is not as straightforward for fuzzy clustering because of the allowance for partial membership. However, the impact of the PCA can still be discussed with simple matrix multiplication. Let A be a $c \times m$ matrix of fuzzy cluster memberships where m is the total number of substations and c is the number of clusters and let B be the $m \times 2$ matrix which contains the first two PC scores for each substation. Then the matrix $W = AB$ is a $c \times 2$ matrix which contains the summation of the PC scores for each cluster weighted by the cluster membership of the substations.

After the discussion of each method, the clustering validity metrics will be discussed further in the context of the clusters and any differences between the methods will be highlighted and explained.

Choice of K Number of Clusters

The Dunn index in Fig. 6.1 shows that K-means and Ward's linkage criteria performed the worst. However, upon closer inspection it was found that the average, single and complete methods had most of the substations in a single cluster, making this misleading. The size of clusters for K-means and Ward's linkage are shown in their respective sections. For the choice of K, considering local maxima in a small neighbourhood, $K \in [9, 14]$ is appropriate for Ward's linkage. For K-means, there is no clear local maximum so based on $K \in [9, 14]$ being a good choice for Ward's

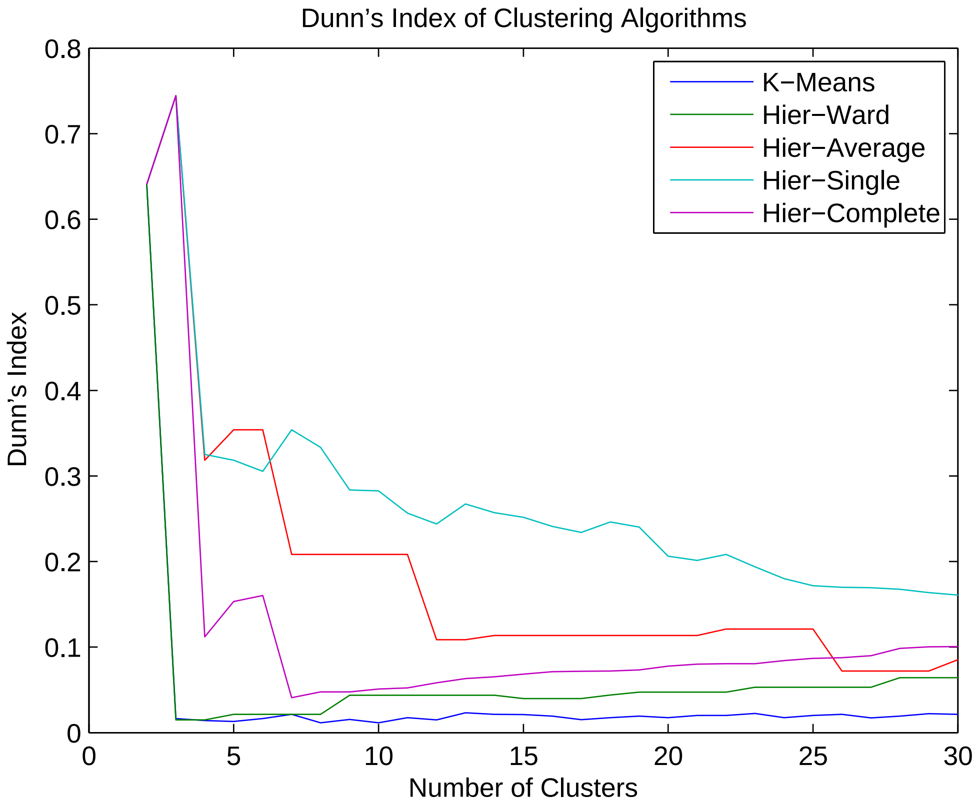


Figure 6.1: Dunn's Index for Crisp Clustering Algorithms

linkage, $K = 9$ clusters are chosen. The smallest number of clusters is preferable whenever possible for ease in understanding.

The choice of K according to the DB index in Fig. 6.2 is less clear. There is a local minimum for Ward's linkage at $K = 9$ but there is also one at $K = 6$. For K-means, there is a local minimum at $K = 8$ but the differences between for nearby values of K are small. Since the choice of $K = 9$ has the most supporting evidence, it is used for subsection 6.2.1 and 6.2.1.

For fuzzy clustering, the Xie-Beni index in Fig. 6.3 shows a general decrease as the number of clusters is increased which would suggest choosing a high number of clusters. As discussed earlier, this can be counterintuitive for the desire to achieve an engineering based explanation of load profile shapes. This problem is acknowledged in Xie and Beni (1991) and proposes a few ways to address this. One way to do this is by considering the maximum number of clusters $n - 1$ to be where there is a monotonically decreasing value of S afterwards. Then select the lowest value

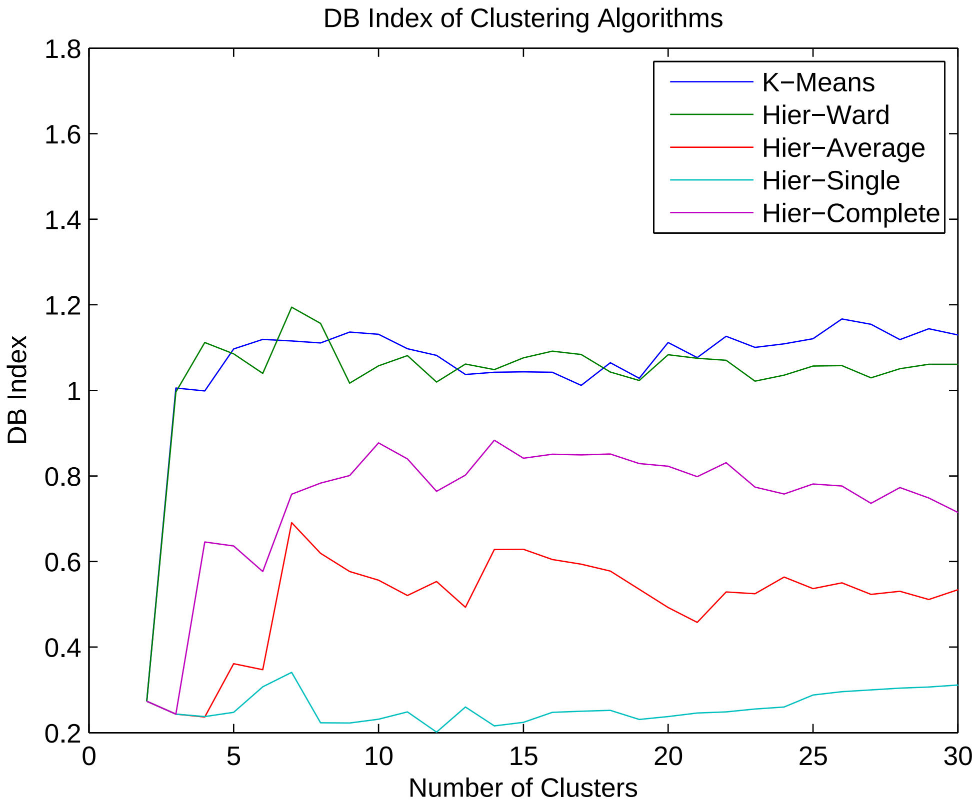


Figure 6.2: DB Index for Crisp Clustering Algorithms

of S in $c \in [2, n-1]$. However, there may not be a point at which the series is monotonically decreasing. Instead, using prior knowledge that $c \ll n$ a local minimum can be selected. Here, a value of 12 clusters is chosen as it is a clear local maximum.

With regards to clustering validity in general, the choice of a number of clusters is not simply the point at which the chosen metric gives the most favourable value. Instead, the application must be considered and choosing a local minima/maxima in a predetermined neighbourhood usually provides a good compromise between minimising/maximising the metric and an understandable number of clusters.

K-Means Clustering

After performing K-means clustering on the summer representative load profiles, Fig. 6.4 shows the candidate nine cluster centres and their cluster sizes in Table 6.1. Note that these clusters were reordered by cluster shape. Their original ordering is

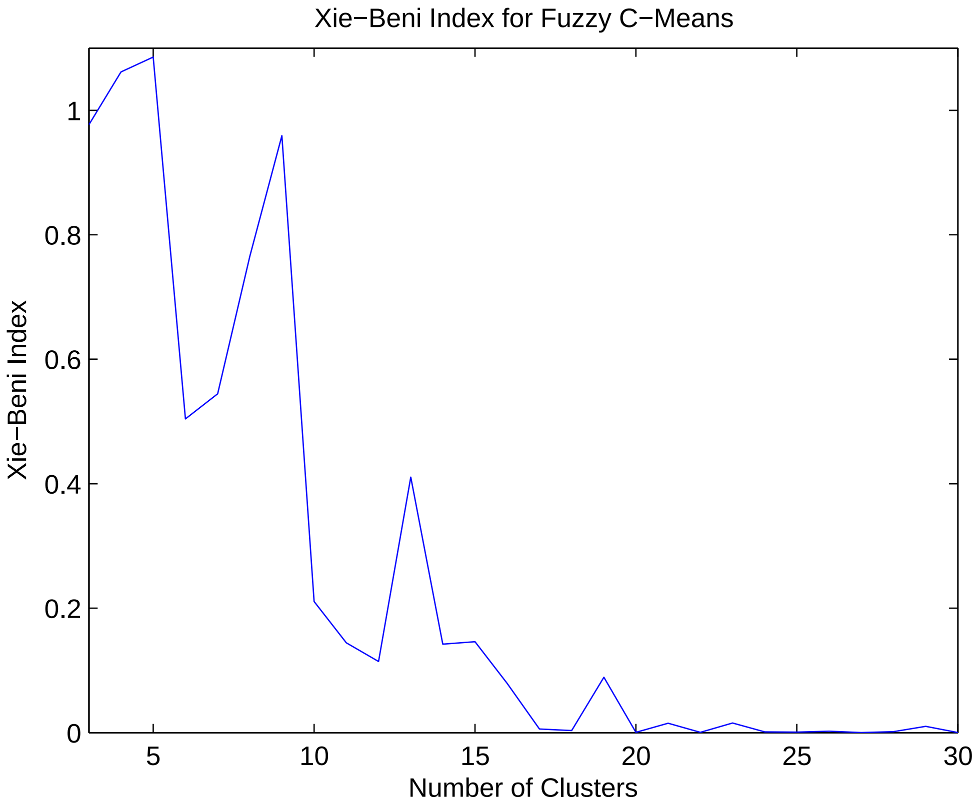


Figure 6.3: Xie-Beni Index for Fuzzy C-Means

Table 6.1: K-Means Clustering Populations and Principal Component Values

	Population	PC1 (Domestic vs. Industrial)	PC2 (Urban vs. Rural)
Cluster 1	60	0.101	0.155
Cluster 2	162	-0.164	-0.262
Cluster 3	72	2.969	-0.067
Cluster 4	127	0.297	0.224
Cluster 5	53	1.063	1.166
Cluster 6	27	-1.207	-0.474
Cluster 7	63	-2.231	-0.466
Cluster 8	8	N/A	N/A
Cluster 9	4	1.851	1.665

given in Table A.1. By cross referencing with the work in Western Power Distribution (2013) and using general knowledge of load profiles, it is possible to attain a

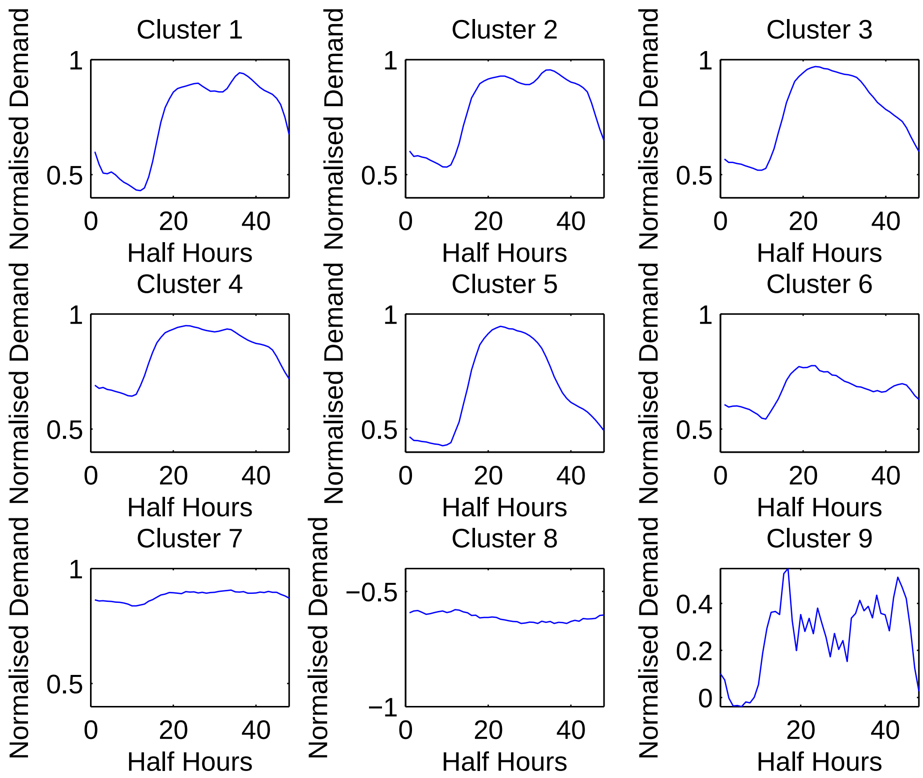


Figure 6.4: Clusters Gained By K-means With K=9

descriptive reference for these clusters. Note that N/A in these tables means there is no available customer data for these clusters because none of the substations have associated customer data.

Clusters 1 and 2 follow a typical domestic type load profile which peaks in the early evening around 6 p.m. and has a fairly flat profile throughout the middle of the day. In Western Power Distribution (2013), profiles of this shape are said to have high domestic dominance. In Table 6.1, the PC1 values for clusters 1 and 2 are small and close to 0 showing that there is influence from both domestic and commercial customers.

Clusters 3, 4, and 5 have a much earlier peak around midday with a steady decrease afterwards. In Western Power Distribution (2013), these types of load profiles are said to have a high commercial/industrial influence. Table 6.1 supports this since the PC1 values are more skewed to the positive end indicating a greater influence from commercial/industrial customers.

Cluster 7 is a mostly flat profile which is typically attributed to commercial and industrial customers using the same level of electrical demand throughout the day. However, this is not supported by the PCA as the value for PC1 is negative and indicates a higher influence from domestic customers. This suggests that either the first two components of PCA don't capture enough variance or that for this particular data set, these flat daily load profiles are attributed to domestic customers even if it is not generally the case.

Clusters 6, 8 and 9 are anomalous load patterns with small populations that do not fit into the other categories. Note that in cluster 8, these are substations on the network that are completely generation.

One of the disadvantages of the K-means algorithm is its sensitivity to outliers and noise. As the number of clusters is increased, the probability of an outlier distorting a cluster centroid decreases. Even in this example of nine clusters, there are a couple of clusters (8 and 9) which are sparsely populated and could be classified as outlier clusters. Therefore for K-means, the choice of K and population of each cluster is important when considering the effect of outliers. If only a few clusters are chosen then it may become a concern for the engineer.

The K-means algorithm mostly performed well here based on the fact that these profile shapes are commonly seen in the industry and that they can be supported by previous works with regards to customer types.

Hierarchical Clustering

Hierarchical clustering with Ward's linkage criteria is applied for this example and nine prospective cluster centroids are shown in Fig. 6.5. Note that these clusters are reordered by shape. These are generated by averaging the profiles of the substations in each cluster after the selection of nine clusters are made. According to Everitt et al. (2001), Ward's method will tend to find clusters which are the same size. The cluster populations shown in Table 6.2 do not corroborate with this however as some of the clusters are not well populated. Four of the nine clusters contain 95% of the total population.

The load profiles that are well populated do show some of the shapes given by

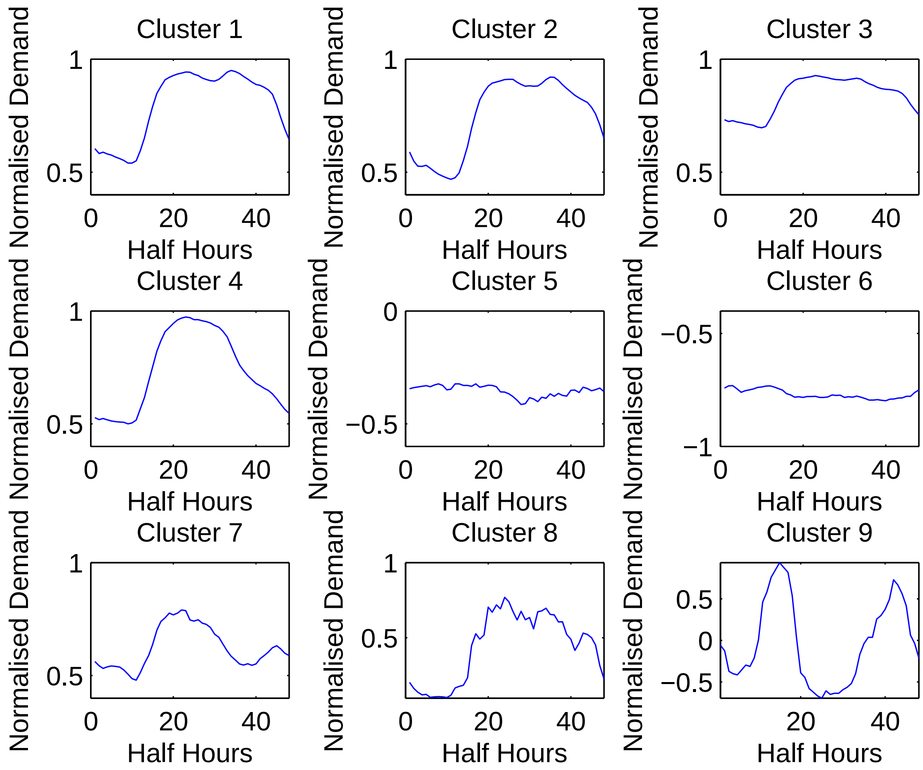


Figure 6.5: Clusters Gained By Hierarchical Clustering

Table 6.2: Hierarchical Clustering Populations and Principal Component Values

	Population	PC1 (Domestic vs. Industrial)	PC2 (Urban vs. Rural)
Cluster 1	172	-0.061	-0.238
Cluster 2	159	1.636	-0.076
Cluster 3	142	-1.268	-0.016
Cluster 4	73	1.047	1.354
Cluster 5	3	N/A	N/A
Cluster 6	5	N/A	N/A
Cluster 7	15	-1.066	-0.147
Cluster 8	6	-0.918	1.437
Cluster 9	1	-1.851	1.665

other clustering methods. In particular, clusters 1 and 2 show a load profile with two peaks. This suggests that at least in part that these are clusters that contain the prototypical domestic load profile with a peak around midday, a slight decrease in the afternoon followed by the daily peak in the evening. However, the two peaks are nearly identical which suggests that these profiles were offset in time.

Another issue is that these profiles are extremely similar in shape which is not desirable considering only nine clusters out of a total 576 substations are considered. When checked against the figures from the PCA analysis in Table 6.2, it suggests that cluster 1 contains more load profiles with a domestic influence and cluster 2 contains more profiles with an industrial/commercial influence. This is not made clear in the shapes of the load profiles which makes this method less useful for both future forecasting purposes and in a more general sense for understanding.

The main issue with hierarchical clustering is that once two items are grouped in the dendrogram, they are no longer considered in future iterations of the algorithm. This is of great importance here because distinct and populated clusters are required for understanding.

Fuzzy C-means Clustering

Fuzzy C-means clustering is unique among these methods in that substations do not have to belong to a single cluster. Instead, they are given membership values in $[0, 1]$ that denote the degree to which the daily load profile shape matches the centre point with a value of 0 being the weakest and 1 being the strongest. Similarly to other methods used, the load profile centres in Fig. 6.6 can be explained using general knowledge of load profiles but also using the preliminary work done in Section 4.3. Note that these clusters are reordered by shape. By performing a summation across the membership matrix U, Table 6.3 shows the number of substations belonging to each cluster rounded to the nearest whole number.

The first principal component indicates the contribution of commercial/industrial customers where a positive value indicates a greater dominance of commercial/industrial customers while a negative value indicates a greater dominance of domestic customers. The clusters where this value is the highest are in clusters 1, 3, 4, 5, 6, 8 and 9. All except cluster 1 have no distinct second peak and are consistent with industrial load profiles in industry and as shown in Western Power Distribution (2013).

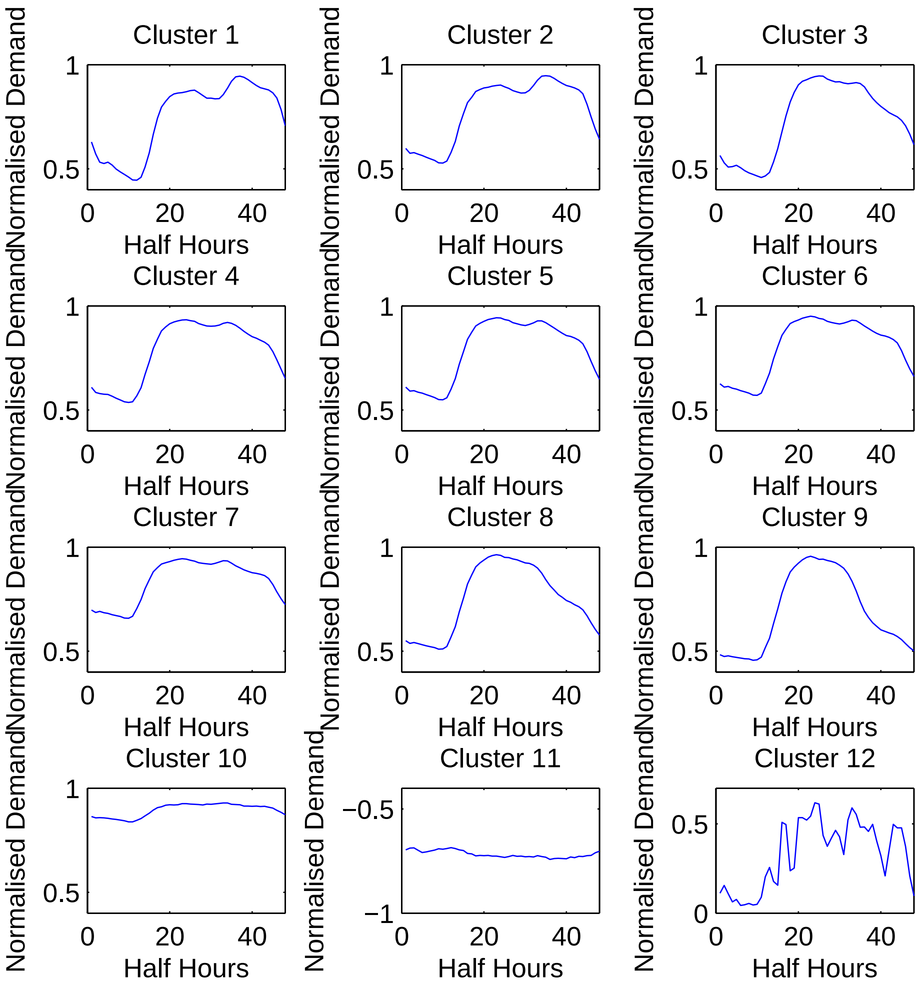


Figure 6.6: Clusters Gained By Fuzzy C-means Clustering

The load profile in cluster 1 is more often associated to more domestic load profiles so this result is surprising. This could be due to the fuzzy clustering algorithm itself since these substations could be outliers in the principal component space which are then exacerbated by the matrix multiplication.

The lowest PC1 values are in clusters 2, 7 and 10 where cluster 2 does exhibit a domestic load profile but 7 and 10 do not. Again, this could be due to the allowance for partial membership as 7 and 10 are much flatter profiles than what is expected of a domestic load profile.

The second principal component indicates the contribution of rural (positive values) and urban (negative values) domestic customers. The highest positive values

Table 6.3: Clusters Gained By Fuzzy C-means Clustering

	Population	PC1 (Domestic vs. Industrial)	PC2 (Urban vs. Rural)
Cluster 1	48	34.523	-3.274
Cluster 2	68	-8.645	-8.588
Cluster 3	49	58.302	3.469
Cluster 4	65	21.579	-3.109
Cluster 5	72	12.539	-1.201
Cluster 6	75	9.090	-2.847
Cluster 7	63	-28.894	-5.132
Cluster 8	56	41.423	20.608
Cluster 9	35	15.092	22.458
Cluster 10	32	-31.580	-5.398
Cluster 11	7	0.213	0.191
Cluster 12	5	0.580	0.821

in clusters 8 and 9 show load profiles with the earliest decrease in demand more consistent with industrial load profiles. This follows naturally as one would expect more rural housing in these areas.

Overall, it seems that although fuzzy clustering can be more insightful because of relaxing the hard membership criteria, it is less useful when attempting to gain a general overview of substation behaviour. This is compounded by the fact that PCA already loses information when reducing the high dimensionality of the original space. For the analysis here, fuzzy clustering of load profiles is not seen as appropriate.

Comparison between Methods

Analysing the crisp clustering techniques first, the main difference that can immediately be seen is in the distribution of the population between clusters. For K-means, the population is well spread out with only two clusters (8 and 9) containing less than 10 substations. As noted earlier for Ward's linkage, 95% of the substations are grouped in four out of nine clusters. This disparity can be attributed to the way in which the algorithms group data. The methods both use Euclidean distance as a metric for clustering but K-means allows for substations to be reassigned after updating cluster centres. This allows for more distinct cluster groups to be formed. For example, clusters 3 and 4 are given as distinct clusters in K-means but there is no analogous cluster for the K-means cluster 3 in hierarchical clustering.

Fuzzy clustering does offer the potential for more insight but the method offered the least corroboration with the principal component analysis in Section 4.3. The method produced cluster centres of a similar nature to the crisp methods but with inconsistent customer values. For the purposes here, the partial membership proved to be more of a hindrance because it made analysing the total cluster more difficult.

Regarding the clusters in general, it is desirable for each cluster centroid to be unique in shape with the clusters generally having significant numbers in population size. Apart from a good computational performance criteria, there is also a need for human interpretation. If the clusters appear similar visually or the population is mostly contained in a small number of n clusters where $n \ll K$ then the clustering has lost much of its purpose. Arguably, if the initial data set contains many instances in close proximity to each other then this cannot be helped. However, here there is prior knowledge that on a distribution network of this level that there will be unique profiles each with a healthy population size. The load profiles shown in Western Power Distribution (2013) are evidence of this. Based on this reasoning and the best support of the customer data, K-means is seen as the most appropriate method.

Clustering Technique Conclusion

This section analysed various clustering techniques on daily load profiles for the purposes of understanding the impact of these load profiles on the distribution network. Also, using the work in Section 4,3, the customer make-up at these substations can be attributed to the load profiles. Much of the work in this area in the past has been done on an individual customer basis which makes supporting the shape of the load profile with engineering judgment a much easier task. At the distribution level, a mixture of different customers can make analysis more difficult but the preliminary work done in Section 4.3 generally aided the analysis. Out of the algorithms considered here, K-means proved to be best fit for purpose. The linkage methods for hierarchical clustering that only compared one object in one cluster to one in another did poorly in evenly populating clusters. Fuzzy clustering did provide more insight than either with fuzzy membership but did not corroborate as well as K-means with the customer work done in Section 4.3. Overall, this clustering work provides a good basis for general engineering understanding and for future forecasting techniques.

Overall Customer Information

K-Means clustering was chosen in Section 4.3 and Section 6.2 as the technique for creating representative load profiles based on work in Chicco et al. (2006). The K-means approach was chosen over other clustering methods due to the distribution of substations which created visually distinct cluster groups.

In Fig. 6.4 the daily representative load profiles show the energy usage of varying types of customers. The choice of $K = 9$ was made based on the error metrics Dunn's index and Davies-Bouldin index. The justification for the type of customer associated to each clustered load profiles arises from industry knowledge and supplemental data regarding numbers of customers for each substation. Note that clusters 8 and 9 will not be referred to in the remaining parts of this paper due to them consisting of generating substations and extreme outliers.

The supplemental customer data detailed in Section 4.3 is commissioned by NPG which associates 35 customer categories to each substation on the network, 20 of which are domestic and 15 of which are commercial/industrial. Associated to each of the 35 categories is a count of the number of customers belonging to that type where one customer can only belong to one category. The domestic customers are labelled 1-20 based on the five descriptors: fuel (mains gas connection), location, size, tenure and age. The commercial/industrial customer types have 15 descriptors based on buildings and services such as offices, shops, schools and transport.

Fig. 4.8 shows the customer space for the first two principal components after being transformed from 35 customer descriptors as shown in Section 4.3. This immediately gives qualitative descriptors for each substation dependent on their PCA score. The first principal component contrasts domestic customers against commercial/industrial customers and the second principal component contrasts urban areas against rural areas.

6.2.2 ANFIS

The adaptive network based approach to fuzzy inference system (ANFIS) developed by Jang (1993) and detailed in Takagi and Sugeno (1985) and Sumathi and Paneerselvam (2010) relies on the concept of fuzzy sets and will be used as the basis for the algorithm. ANFIS has the benefits of an artificial neural network structure which is able to forecast non-linear data while offering some greater insight through its membership functions.

Input Selection

Jang (1996) described how to perform input selection given many input candidates. If grid partitioning is used to divide the input space into sections representing every possible combination of inputs, then a small number of inputs should be used so as to avoid the curse of dimensionality. That is, for n number of inputs each with m numbers of membership functions we have m^n fuzzy rules generated. It was found that ANFIS can produce satisfactory results after one epoch of training due to the hybrid learning approach of gradient descent and least squares estimation (LSE) (Jang, 1996). Therefore, an input selection method is based on an assumption that the ANFIS model with the lowest root mean squared error (RMSE) after one epoch will provide good results when trained. This is an assumption based on the least squares method being the main driving force behind the learning in ANFIS as opposed to gradient descent. For example, if there is a problem with 10 candidate inputs and the three best inputs are to be chosen from them then $C_3^{10} = 120$ ANFIS models are constructed each trained with just one epoch. Then the ANFIS model with the lowest RMSE is selected and used in future training. Also note that one epoch training of 120 ANFIS models involves less computation that 120-epochs of training a single ANFIS model. Therefore, this method is not computationally expensive.

This method will be used to select the appropriate time-lagged inputs for CL-ANFIS. That is, it will choose which days in a given time window will be used to predict future daily demand values. See Section 6.3 for further details.

Application to Power Systems Demand Data

In order to produce the forecast, historical demand data will be used. The demand data is gathered on a half hourly basis but to reduce the amount of unnecessary data, only daily maximum demand is considered. After preprocessing is completed, demand data is given to ANFIS at historical time intervals which are chosen based on input selection. The Takagi-Sugeno (Takagi and Sugeno, 1985) type rules of ANFIS can assist in the understanding of the forecast by giving qualitative labels to membership functions. For example, a candidate rule would be of the form

$$\begin{aligned} \textbf{If } & x_1 = x(t-s_1) \text{ is LOW,} \\ & x_2 = x(t-s_2) \text{ is MEDIUM and} \\ & x_3 = x(t-s_3) \text{ is HIGH} \\ \textbf{then } & f = p_0 + p_1 x_1 + p_2 x_2 + p_3 x_3 \end{aligned} \tag{6.2.4}$$

where x_i are historical loads at past times $t - s_i$ which are chosen by input selection to give a final future load value f with learned parameters p_0, p_1, p_2, p_3. Note that qualitative labels are subjective and a further discussion of this concept is carried out in Sections 6.6 and 6.7.

Fig. 6.7 shows an example of membership functions for one input. In this example, there are three membership functions which can be given the qualitative labels of LOW, MEDIUM and HIGH. Depending on where the historical loads are, the output will be affected by the placement of the membership function. The benefits of the membership functions of ANFIS and the Takagi-Sugeno rule base is further explored in Section 6.6.

Membership Functions

When considering the error differences between the clustered model and unclustered model, the membership functions of the models can be used as a way to explain this.

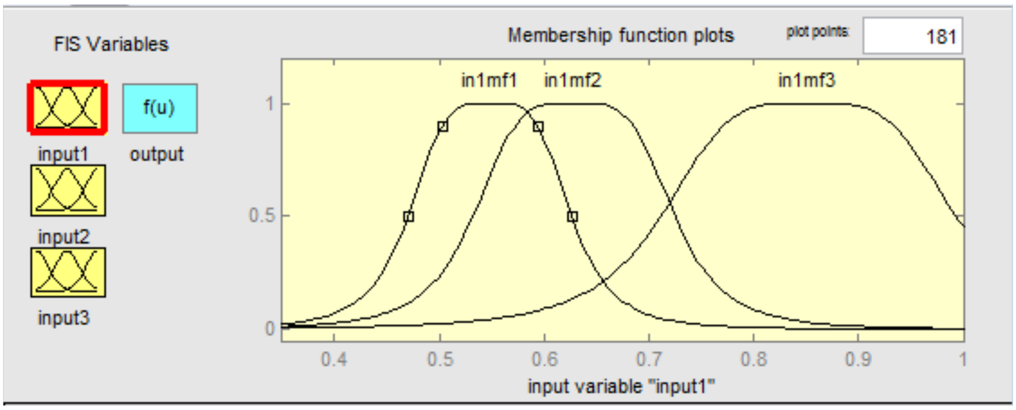


Figure 6.7: Membership Function Example

One of the similarity measures that can be used is one given by Wang (1997):

$$S(A,B) = \frac{1}{N}\left[\sum_{i=1}^{n} \frac{\min(\mu_A(x_i),\mu_B(x_i))}{\max(\mu_A(x_i),\mu_B(x_i))}\right] \tag{6.2.5}$$

where $\mu_{A,B}(x_i)$ is the membership function of the fuzzy sets A, B on the element x_i. The right hand side can be seen as the average of all the elements $x_i \in X$.

Wang (1997) ran trials of other similarity measures in the literature and showed that a simple comparison between two fuzzy sets could often result in different figures depending on the similarity measure used. Therefore, it is difficult to determine which is necessarily the best. Rather than use this as an absolute measure of similarity, it will be regarded as supplementary information that will assist in explaining the primary results of the differences in forecasting error. In the case of all the similarity measures, the relative impact of a membership function for a particular input is not shown. Here an adjusted weighted similarity measure will be used:

$$\tilde{S}(A,B) = \frac{1}{2}(\bar{\mu}_A(x_i) + \bar{\mu}_B(x_i))S(A,B) \tag{6.2.6}$$

By using this, the importance of each membership function is brought out so that those with smaller fuzzy membership values have the impact of their similarity lowered.

There are three membership functions for three inputs for each clustered model resulting in 9 membership functions. Therefore, for each cluster there is a comparison of the clustered 9 membership functions to the 9 membership functions of the general model. These $S(A,B)$ values are averaged across all substations in the clusters. Rather than using these $S(A,B)$ values in the evaluation, the impact of the input values should come across in the comparison. For example, let $S_1(A_1,B) = 0.8$ and $S_2(A_2,B) = 0.5$ with $\frac{\sum \mu_{A_1}(x_i)}{n} = 0.6, \frac{\sum \mu_{A_2}(x_i)}{n} = 0.9, \frac{\sum \mu_B(x_i)}{n} = 0.7$. Then $\tilde{S}_1(A_1,B) = 0.52$ and $\tilde{S}_2(A_2,B) = 0.4$. With the original metrics, the membership function A_1 had a higher similarity measure with B than A_2. However, using the weighted similarity measure, the difference between $\tilde{S}_1$ and $\tilde{S}_2$ is reduced because of the impact of the elements on A_1 and A_2. In terms of the comparison between the clustered and unclustered ANFIS models, this means the membership functions will be compared by the relative impact the demand values have on the membership functions rather than simply the similarity of the membership functions alone.

From a qualitative perspective, the parameters of the generalised bell function can also be used for information. For example from $f(x,a,b,c) = \frac{1}{1+|\frac{x-c}{a}|^{2b}}$, the c parameter indicates the centre of the curve and the a parameter indicates the width. If used in conjunction with the Takagi-Sugeno type rules, a qualitative description of a particular substation can be built.

6.2.3 Benchmark - Holt-Winters Exponential Smoothing

As an additional benchmark for the proposed method, Holt-Winters (HW) exponential smoothing will be used. The demand series is written in the form (Hyndman and Athanasoupoulos, 2013):

$$F_{t+m} = (s_t + lb_t)c_{t-s+m} \tag{6.2.7}$$

$$s_t = \alpha\frac{y_t}{c_{t-s}} + (1-\alpha)(s_{t-1} + b_{t-1}) \tag{6.2.8}$$

$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1} \tag{6.2.9}$$

$$c_t = \gamma\frac{y_t}{s_t} + (1-\gamma)c_{t-s} \tag{6.2.10}$$

where F_{t+m} is the value of the series y at $t+m$, s_t is the base signal, b_t is the linear trend component and c_t is the seasonal correction factor. The parameters α, β, γ are the associated parameters for s_t, b_t, and c_t respectively. The parameter $\alpha \in [0,1]$ controls the extent to which more recent or distant observations have an

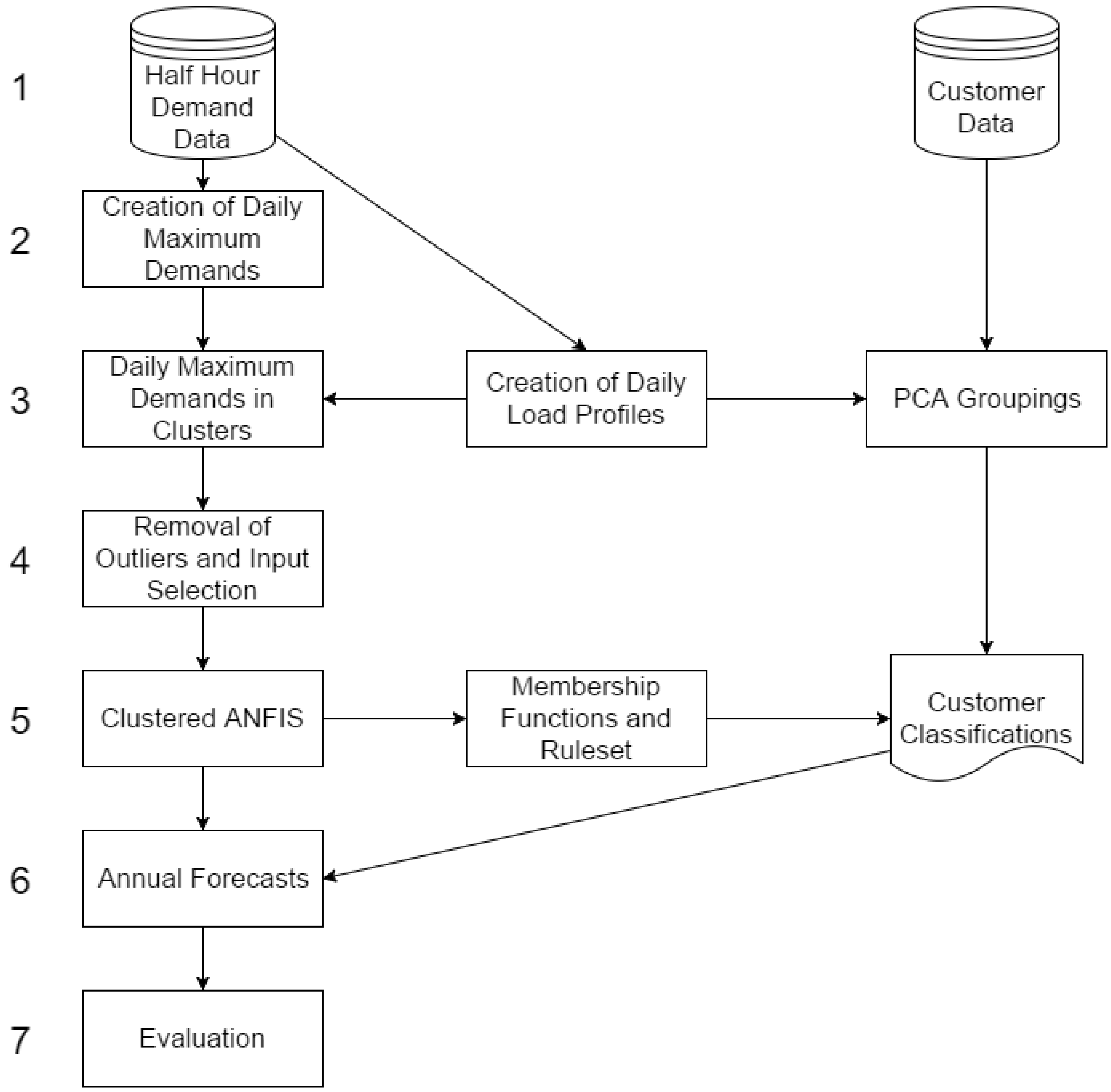


Figure 6.8: Flowchart for CL-ANFIS Model

effect on the forecast. The parameter $\beta \in [0, 1]$ indicates the extent to which the linear trend component b_t is updates over time. The parameter $\gamma \in [0, 1]$ controls the extent to which the seasonal component c_t is affected by more recent or distant observations. The optimal parameters are found by minimising the squared one-step ahead prediction error.

6.3 Cluster-based Methodology

In order to demonstrate the effectiveness of the clustered groupings on CL-ANFIS, this experiment will compare the forecast accuracy of using one unclustered ANFIS model as a benchmark against the CL-ANFIS models which are built using the clustered substations. The model flowchart for the CL-ANFIS is given in Fig. 6.8 addressed here by the numeric labels. At stage 1, both sets of half hour demand data and demographic data are used as inputs. The demographic data is the set of 35 customer descriptors detailed in Section 4.3.

In stages 2 and 3, the daily load profile clusters as seen in Fig. 6.4 are created as described in Section 4.1. At stage 4, the profiles are prepared for forecasting through outlier detection and input selection. Normally outliers are often removed or ignored from data sets but here they are adjusted so the time series is not disjointed too greatly especially in the case of consecutive outliers.

Then the input selection for the time lagged variables is carried out with candidates over a monthly period. The chosen inputs for the annual demand profile x are $x(t-30), x(t-b), x(t-c)$ where $b \in [c+1, 29], c \in [1, 28]$. The choice of range of inputs here is a compromise between simply improving the accuracy of the model against achieving a qualitative understanding of the forecast. In this local neighbourhood of 30 days (± 10 days), it was found that changing the size of this window did not have a noticeable effect on the error rates of the input selection ANFIS models.

In order to pass to ANFIS, a fuzzy inference system using grid partitioning is initialised with a candidate primary substation for each of the eight clusters. After the FIS structures are built for each of the clusters for the period of 2010-13, they are evaluated on the test data for the years 2013-15. This lead to the clustered CL-ANFIS model created in stage 5. Then after the annual forecasts are obtained at stage 6, error evaluation is conducted in stage 7.

In order to measure the forecast accuracy, a couple of forecasting error metrics are used. One of the simplest error metrics is the mean absolute error (MAE). It is defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |p_i - a_i| \tag{6.3.11}$$

where n is the total number of data points, p_i are the predicted values and a_i are the actual values.

Another error metric that can be considered is the RMSE. It is defined as:

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^{N} |p_i - a_i|^2 \right]^{\frac{1}{2}} \tag{6.3.12}$$

According to Willmott and Matsuura (2005), the RMSE is an ambiguous measure of average error. This is due to the fact that RMSE is bounded with $RMSE \in \left[MAE, n^{\frac{1}{2}} \cdot MAE\right]$ and as the variance of the frequency distribution of the error magnitudes increases, RMSE increases faster than MAE in a sometimes non-monotonic fashion. However, RMSE is consistently used as an error of measure in several model based studies (Willmott et al., 1985). This provides additional considerations to the more traditional error of MAE.

The half hour demand data consists of 305 substations each of which have five years of daily demands available after stage 2 is completed. Three of these years (April 2010- March 2013) of daily maximum demand will be used for training CL-ANFIS and two years (April 2013- March 2015) will be used for testing. The error metrics will be calculated on the out-of-sample testing set for the 2 year forecast for each individual substation. To summarise these metrics, the mean and median of these errors will be given per cluster.

The Holt-Winters method (Kalekar, 2004) will also be used as another benchmark for the CL-ANFIS method. In the literature, HW has been used as a forecasting tool and shown to be effective for load forecasting (Bianco et al., 2010; Taylor, 2010). Based on the literature (Amin-Naseri and Souroush, 2008; Goia et al., 2010), it is expected that the clustered CL-ANFIS models will outperform the larger unclustered ANFIS model. In the event where this does not occur, demographic information will be used for supporting conclusions.

6.4 Results of Clustering Method

First, an absolute comparison of the clustered CL-ANFIS results against the large ANFIS model will be discussed. The clusters shall be examined in order of their population size so their relative impact can be discussed. A comparison of the MAE and RMSE errors will be made between the CL-ANFIS and ANFIS models in

Table 6.4: Median Errors of Three Forecasting Models

Cluster	1	2	3	4	5	6	7
Size	13	97	6	99	38	15	37
CL - ANFIS MAE	**0.085**	**0.127**	**0.109**	**0.115**	**0.137**	0.376	0.782
ANFIS MAE	0.199	0.139	0.121	0.183	0.185	0.296	0.209
HW MAE	0.145	0.249	0.135	0.200	0.185	**0.201**	**0.179**
CL - ANFIS RMSE	0.279	**0.247**	0.290	**0.236**	0.269	0.531	1.596
ANFIS RMSE	0.367	0.298	0.268	0.380	0.395	0.526	0.369
HW RMSE	**0.182**	0.283	**0.164**	0.245	**0.221**	**0.252**	**0.222**

Table 6.5: Weighted Similarity Measure

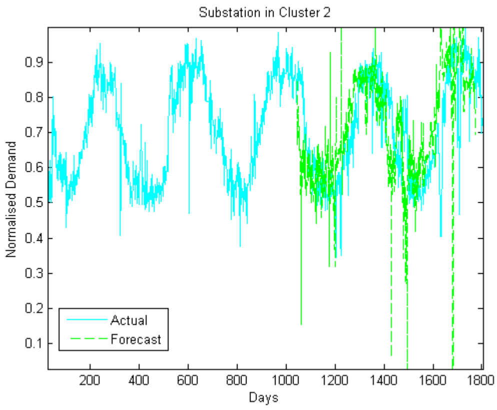
Cluster	Weighted Similarity
1	0.106
2	0.144
3	0.098
4	0.122
5	0.091
6	0.101
7	0.093

Table 6.4 and relating them back to the load profiles given in Fig. 4.8. This should give some further insight into any discrepancies in forecasting error between the two models.

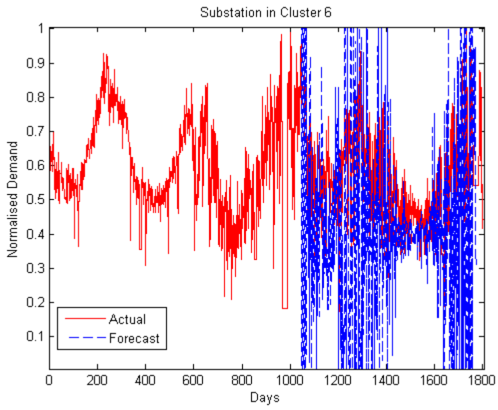
Then, a short comparison will be made between the CL-ANFIS method as a whole and the Holt-Winters benchmark. In this way, the advantages and disadvantages of the CL-ANFIS will become apparent.

6.4.1 Clusters 2 and 4 - The "Traditional" Load Profiles

Firstly looking at cluster 2, this daily load profile is dominated by domestic households according to the PCA in Fig. 4.8 and exhibits the traditional double peak that



(a) Case I - Good Forecast in Cluster 2



(b) Case II - Poor Forecast in Cluster 6

Figure 6.9: Difference between Load Profiles in Principal Component 1

is obtained from typical working hours of about 09:00-17:00. Looking at the error results in Table 6.4, this cluster showed minimal difference in terms of the clustered CL-ANFIS model and the unclustered model. As the substations show some of the most standard load profiles, this is not necessarily surprising. The CL-ANFIS MAE value of 0.127 in Table 6.4 for cluster 2 is slightly higher than a desired value of around 0.1. In the current practice of NPG, an error of $\pm 10\%$ is automatically accepted and anything higher is checked. In this sense, the method presented here is comparable as roughly 50% of substations are manually checked. In Section 6.5, the substations with significantly high error will be checked more closely but for this more well behaved cluster, the forecast is promising and example of which is shown in Fig. 6.9a. In Table 6.5, the weighted similarity measures are given and it is seen that cluster 2 has the largest value between the membership functions of the clustered and unclustered ANFIS. This supports the minimal difference in their error values.

In Table 6.4, cluster 4 has similar error values to cluster 2 although there is a more noticeable improvement in the CL-ANFIS model over the unclustered model. The daily load profile is skewed more towards commercial type customers, hence the flatter profile of cluster 4 in Fig. 4.8. As clusters 2 and 4 contain most of the substations, the unclustered ANFIS will tend towards membership functions that are similar to these clusters. However, the clustered CL-ANFIS can give membership functions which are more tailored to the individual cluster so there is an improvement in this case. In Table 6.5, cluster 4 has the second highest similarity measure, supporting this thought.

6.4.2 Cluster 5 - The Non-Peaked Load Profile

The cluster 5 daily load profile in Fig. 6.4 is unique from the others as it does not display anything similar to the two peaks in other clusters. This type of load profile is more consistent with businesses which operate for a set amount of hours. This is supported by the PCA which showed this cluster is dominated by commercial customers in a rural setting. The clustered errors in Table 6.4 are comparable to that of 2 and 4 which were acceptable. However, the improvement in error of the clustered model over the unclustered model is more apparent as there is a 5% MAE difference. The population size of this cluster is significantly smaller so the load profiles of these substations are not reflected in the larger unclustered model. Therefore, the CL-ANFIS model is much more suited here and a noticeable improvement is seen in both MAE and RMSE. This idea is reflected in Table 6.5 which shows this cluster as having the lowest similarity measure between the membership functions of the two models. This shows that although this cluster has an atypical load profile, it is possible to produce an accurate forecast if they are addressed separately.

6.4.3 Clusters 6 and 7 - The Flat Load Profiles

Clusters 6 and 7 are the only two clusters where the clustered CL-ANFIS model performed significantly worse than the unclustered model. Firstly, looking at the daily load profile of cluster 6 in Fig. 6.4, it almost resembles a more typical load profile as in cluster 1 or 2 but at lower demand values. The PCA shows that this cluster is dominated by domestic customers in an urban environment. All of the error values for cluster 6 in Table 6.4 are quite high with a median MAE value of 0.376 for CL-ANFIS. This poor forecast is reflected in Fig. 6.9b where the forecast fails to resemble the actual events. One possible reason for this high error is that some of these substations can be more erratic with an unpredictable load profile.

Therefore when addressing this cluster in the unclustered model, the forecast was approximately 8% more accurate as it is more suited to the general case. The small population of this cluster hinders the CL-ANFIS model in this case. Table 6.5 shows a lower similarity measure for cluster 6 explaining the discrepancy in errors.

For cluster 7, the daily load profile shown in Fig. 6.4 is nearly flat. The PCA however showed cluster 7 was also influenced by domestic customers in an urban setting. There is already some discrepancy here as a flat load profile would only be expected of industrial customers. As cluster 7 is more well populated than cluster 6, the cluster centre is averaging out different substations which are more erratic. The error values for cluster 7 in Table 6.4 are unreasonably high indicating that either the model is completely inappropriate or the substations are so erratic that a good forecast for them is difficult. Similarly to cluster 6, there is a notable improvement for the unclustered model although the error is still much higher than desired. These two clusters appear to contain the more unpredictable substations and this will be discussed further in Section 5. Table 6.5 reflects this discrepancy in the low weighted similarity measure for cluster 7.

6.4.4 Comments on the Remaining Clusters

Due to the fact that the remaining cluster groups 1 and 3 are not as well populated as the ones discussed in the previous subsections, conclusions drawn from them are not as reliable. However, it is noted that for both clusters, there is an improvement by using the clustered CL-ANFIS model.

6.4.5 Comparison of the CL-ANFIS Model against Holt-Winters

Holt-Winters exponential smoothing was applied to the data as an additional benchmark against the CL-ANFIS model. The forecasting errors are supplied in Table 6.4. When comparing the forecasting errors in Table 6.4, the main observation is that Holt-Winters had larger MAE values for cluster 1-5 but smaller errors for clusters 6 and 7. For those substations which have more erratic demand profiles, a method such as Holt-Winters may be more suitable as it is less likely to give large fluctuations during fitting in comparison to CL-ANFIS. There are some instances where the RMSE values are less than the MAE values but since for any individual value, RMSE$\geq$MAE, it must mean that taking the median value here skews the results to a lower than expected RMSE value. Moreover, the high error substations are isolated in the clusters of the CL-ANFIS model. This causes even larger errors as supported by the difference in errors for clusters 6 and 7 across Table 6.4. However, for clusters 1-5, the lower error shows that the CL-ANFIS model is more effective than a traditional forecasting tool in a majority of cases.

6.4.6 Summary of Results

Purely based on the errors, it is shown that the CL-ANFIS algorithm is an improvement over the unclustered model and benchmark in the cases of clusters 1-5. Due to this and the qualitative impact explored in Section 6, this model is suggested for future use. In the cases of cluster 6 and 7, the CL-ANFIS method performs worse than the unclustered model and benchmark. This is believed to be due to the erratic nature of these load profiles not having any similar patterns. Hence the generalised unclustered model will perform better in those cases. This is examined in closer detail in Section 6.5.

6.5 Analysis of Forecast Errors Using Demographic Information

In order to understand the difference in forecasting errors between the low error clusters and more problematic clusters, the supporting demographic information and daily load profiles will be examined. One of the ways this can be checked is by using the principal component scores found in Section 4.3 for each of the substations. By looking at the variance spreads of these scores, it will be possible to see if there are similar customers for each cluster. It would be expected that for the lower error clusters 1-5, there will be a low customer variance in each of the clusters. For cluster 6 and 7 which have high errors it would be expected that there will be a larger customer variance. However, it was found that there was no obvious correlation between customer variance and forecast accuracy. In Table 6.6, clusters 6 and 7 actually have a low variance in each of the principal components despite their high forecast error. Cluster 4 was also shown to have a high variance in PC1 but it was one of the clusters with a low forecasting error.

Another possible contributor to unpredictable load profiles is a low number of customers on the network. Each substation has associated to it a number of customers defined as domestic and a number defined as commercial/industrial. For the purposes of this analysis, a low number of customers will be conditional on both the number of domestic and the number of commercial/industrial customers being less than one standard deviation away from the mean of the number of domestic and commercial/industrial customers across all clusters. This metric is chosen as it will select the substations which are the low outliers regardless of cluster in a normally distributed customer population. The mean and standard deviation for the number of domestic customers are 6191 and 3789 respectively leading to the low customer condition of $6191 - 3789 = 2402$. The mean and standard deviation for the number of commercial/industrial customers are 437 and 263 respectively leading to the low customer condition of $437 - 263 = 174$. The result of this is the conditional rule:

IF the number of domestic customers is less than 2402 AND the number

Table 6.6: Variance of Principal Components per Cluster

Cluster	Var(PC1)	Var(PC2)
1	1.15	4.00
2	2.86	3.63
3	1.77	0.71
4	10.02	3.89
5	5.49	0.82
6	0.82	1.31
7	2.80	2.08

Table 6.7: Low Customer Substations and their Forecasting Error

Cluster	Low Customer Pop.	Average MAE
1	1	0.043
2	2	0.134
3	0	N/A
4	5	0.233
5	2	0.501
6	4	0.330
7	12	2.067

of commercial/industrial customers is less than 174 THEN the substation has a LOW number of customers.

In total, there are 26 substations out of 305 which can be classified as having low customer numbers according to this definition. The numbers for each cluster are shown in Table 6.7. More than half of these are in either cluster 6 or 7 which perform the worst in forecasts. Clusters 6 and 7 also contribute more to their respective clusters due to their smaller population size making up roughly a third of their clusters. The average MAE values in Table 6.7 give some indication that these substations are difficult to forecast on although a comparison against their representative load profiles is needed to gain a clearer understanding.

Closely associated with customer numbers would be the actual maximum demand (MD) values of the substations. A substation with a higher maximum demand value is more likely serving more customers which causes its annual load profile to be more predictable. The average MD values in Table 6.8 show that cluster 6 is clearly made up of more substations with smaller MD values which explain the higher forecasting errors. However, although cluster 7 has a smaller error on average than many of the other clusters, it is not as distinct as the others. Also note that in Table 6.8, cluster 1 has a much higher average MD value over its median MD value. This is due to the small population size and a few high value MDs skewing the average.

After applying the clustering methodology to cluster 6 and 7 again, Fig. 6.10

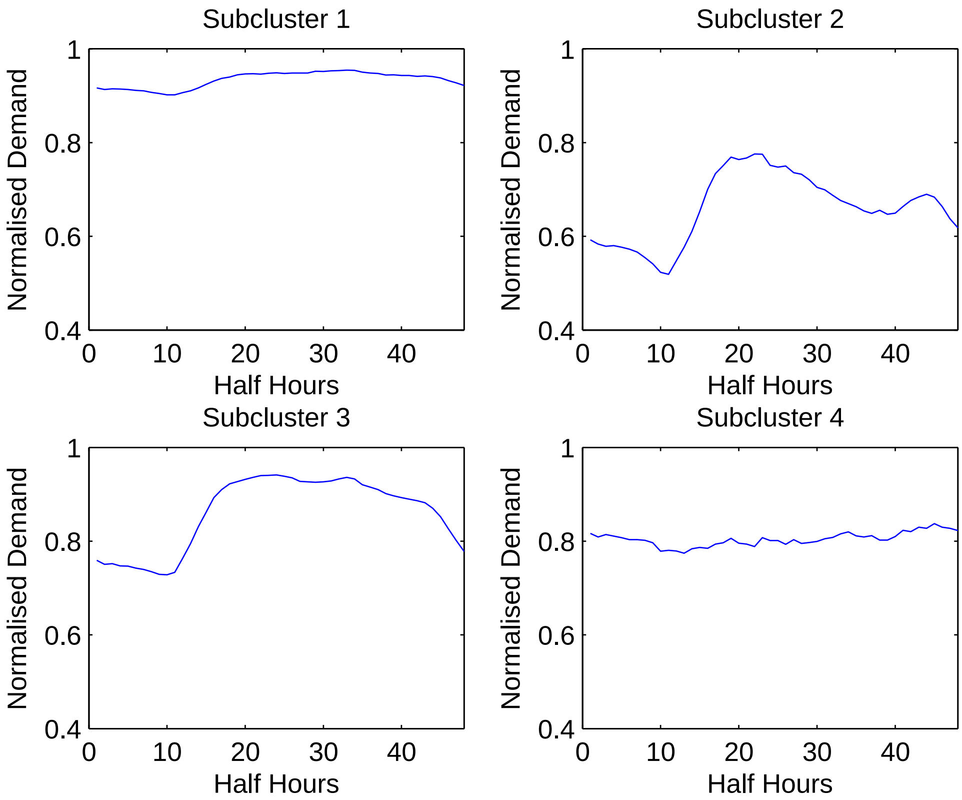


Figure 6.10: Subclusters of Clusters 6 and 7

shows that although the load profile shape of cluster 6 remains unchanged in subcluster 2, cluster 7 is further broken down into more groups. This supports the idea that cluster 6 does consist of profiles that are difficult to forecast on due to their lower demand values but cluster 7 has a mixture of the low demand substations along with the higher demand substations which are more predictable.

The substations in subcluster 3 in Fig. 6.10 show a more typical load profile similar to the cluster centre in cluster 4 in Fig. 4.8. The average distance of these substations in subcluster 3 away from cluster centres 4 and 7 are computed as 0.367 and 0.524 respectively. In general, this shows the clustering method is accurate but there are some substations where the difference between the two centres are within a margin of 0.1. When these substations are sent through the FIS structure for cluster 4, the median MAE is 0.348. This is still a much higher error than required but it is an improvement over the cluster 7 FIS. However, it is still not an improvement over the error using the unclustered ANFIS model. This suggests that cluster 7 consists of

Table 6.8: Maximum Demand Values across Clusters

Cluster	Average MD (kVA)	Median MD (kVA)
1	26783	14224
2	15044	15644
3	15291	15498
4	17660	17170
5	16469	16578
6	12596	10419
7	14906	15579

a mixture of substations, some of which forecast more accurately in a general model. Unlike cluster 6 where the higher error can be explained by low demand substations, there is no clear link between the substations in cluster 7. These substations have abnormal load profiles which are not necessarily linked to customers they serve or the amount of demand. Therefore, it is recommended that these are treated on a case-by-case basis as many of them are under the current DNO methodology. Even in the case of the high error substations, they would be highlighted in a similar way to the NPG method.

6.6 The Qualitative Importance of CL-ANFIS

One of the motivations for using this clustered approach was not only to improve the accuracy of the forecast but to give a generalised view of substations that belong to these clusters in a qualitative sense. Using the available information of the cluster shape, the membership functions in the clustered CL-ANFIS model and rules, a qualitative picture of the profile can be obtained.

First, using the cluster shapes in Fig. 6.4 and the work done in Section 6.2, an initial categorisation of any substation can be given. The qualitative descriptors are given in Table 6.9.

For each cluster, there are three descriptors. Firstly, the sizes of small, medium and large describe the population size for that cluster. This is based on the initial

Table 6.9: Qualitative Descriptors

Cluster	Category
1	Medium Domestic Rural (MDR)
2	Large Domestic Urban (LDU)
3	Medium Commercial Urban (MCU)
4	Large Commercial Rural (LCR)
5	Medium Commercial Rural (MCR)
6	Small Domestic Urban (SDU)
7	Medium Domestic Urban (MDU)

clustering done in Section 6.2 due to the fact that some substations had to be removed for the forecasting experiment because of a lack of data. The small label has a population less than 50, the medium has 50-100 and the large has greater than 100. Secondly, there is the domestic or commercial descriptor. This is given by the daily load profile shape in Fig. 6.4 and the work done in Section 4.3. This describes whether the cluster is dominated by domestic households or commercial/industrial customers. Thirdly, there is the descriptor of urban or rural. This is also based on the work done in Section 4.3 and describes if the substation is in an urban or rural setting. This shows by using the clustered approach, the engineer can already start to form an idea of the type of substation involved.

Next, the membership functions for each cluster can aid the qualitative description. In the forecasting model, each cluster has three inputs with three membership functions each for a total of nine membership functions. They are of the form: $f(x,a,b,c) = \frac{1}{1+|\frac{x-c}{a}|^{2b}}$. The parameter c gives the centre of the curve and a describes the width. The b parameter controls the shape of the curve and does not vary significantly between membership functions. Considering each one of these membership functions individually would be time-consuming so general labels should be formed to aid in the description. Since the substation demands are normalised, all values are in $[0, 1]$ and because there are three membership functions per input, the following

Table 6.10: Frequency of a and c parameters

	a (width)			c (centre)		
	THIN	NORM	LARGE	LOW	MED	HIGH
Cluster 1	2	2	5	3	1	5
Cluster 2	5	3	1	0	6	3
Cluster 3	2	6	1	0	5	4
Cluster 4	3	4	2	0	5	4
Cluster 5	4	1	4	2	3	4
Cluster 6	1	6	2	3	3	3
Cluster 7	2	4	3	2	3	4

labels are suggested based on the centre c parameter:

$$\text{LOW: } [0, 0.33] \text{ MEDIUM: } (0.33, 0.66] \text{ HIGH: } (0.66, 1] \tag{6.6.13}$$

The a parameter which describes the width of the membership function can be described in a similar manner although the range to consider is not immediately obvious. Upon observation across all membership functions per each input: $a \in [0.02, 0.2567]$. By dividing this interval into three equal parts, we can generate the following labels:

$$\text{THIN: } [0.02, 0.099] \text{ NORMAL: } (0.099, 0.178] \text{ LARGE: } (0.178, 0.257] \tag{6.6.14}$$

By examining Table 6.10, the most evident observation to make is the similarity of the c distribution between clusters 2, 3 and 4 and the similarity between 5, 6 and 7. The c parameter gives an indication of the location of inputs for each cluster. This suggests that substations in the clusters which have lower forecasting errors have annual demand profiles which are less varied. The substations in cluster 2-4 rarely have instances where demand falls below 33% of the maximum resulting in membership functions that are scaled higher. The fact that there are some membership functions in cluster 6 and 7 which can be classified as LOW suggests there are a significant number of substations where demand does fall below 33% of the maximum showing instability. However, clusters 1 and 5 also have membership functions in the LOW categorisation but this is reflected in the load profiles shown in Fig. 6.4 where the early morning demand falls lower than that of clusters 2-4.

Table 6.11: Average Distribution of Load for c Parameter

	LOW	MED	HIGH
Cluster 1	0.101	0.557	0.342
Cluster 2	0.020	0.553	0.427
Cluster 3	0.057	0.446	0.497
Cluster 4	0.020	0.496	0.484
Cluster 5	0.072	0.501	0.427
Cluster 6	0.212	0.504	0.283
Cluster 7	0.042	0.458	0.500

If the inputs from the membership functions are analysed and grouped by the bands in Eqn. 6.6.13 then Table 6.11 shows the average distribution. Then for cluster 1, the associated rule would be:

> **If** substation belongs to cluster 1 **then** 10.1% of the load is LOW, 55.7% of the load is MEDIUM, and 34.2% of the load is HIGH

This is of course one such way that the Takagi-Sugeno rules and membership functions can be interpreted. There can be a much more granular level of detail if needed as the membership functions can be retrieved for each cluster. The fuzzy rules themselves can be used to aid in the qualitative description. For the clustered CL-ANFIS model, there will be 27 rules for each of the seven clusters as there are 3 membership functions for 3 historical time inputs. These can give the individual outputs for any set of historical load data as well as giving the qualitative descriptors relative to the membership function. As well as providing the forecast, this method has provided the DNO with qualitative labelling and insight into the algorithm. The DNO will have to apply these rules and review the impact they have on governance and forecasting.

6.7 The Impact of the Customer Classifications on Load Profiles

The motivation for a more in-depth analysis of customer load profiles comes from the need to develop an informed long-term forecast for the DNO. Due to the size of the network and number of customers on an individual substation, other external data sources such as economic or weather data are often too general to have an impact on the annual long-term forecast. This customer database is specific to the network so it is more reliable. If this database can be updated regularly, it would be more useful than other external data sources.

In the literature, there has been work done on examining customer data but often on a smaller household scale. Räsänen et al. (2010) used hourly load data for customers on a household scale to create better estimates of load curves. Elkarmi (2008) considers demand by customer sector in order to assess the benefit of demand side management programs. Several techniques have been considered to analyse customer demand such as clustering (Akperi and Matthews, 2014a; Chicco et al., 2006) and self-organising maps (Räsänen et al., 2008). However, further analysis is needed to justify and explore the above methods. Popular techniques for analysing data sets of this kind are contained in data mining including classification algorithms like naïve Bayes and decision trees (Duda et al., 2000). More recent developments such as artificial neural networks (ANN) (Kanitpanyacharoean and Premrudeepreeechacharn, 2004) can also be used for classification.

6.7.1 Data Description

For a more robust test of the qualitative labelling like those given in Table 6.9, the daily load profile clustering has been redone using the annual load profiling period of April 2012 - March 2013. This is done both for the summer period of June 1 2012 - August 31 2012 and the winter period of December 1 2012 - February 28 2013. Similarly to Section 6.2.1, the choice of $K = 9$ is based on the metrics of Dunn's index and Davies-Bouldin index. The experimental daily load profiles for the summer and winter periods are given in Fig. 6.11 and Fig. 6.12 respectively.

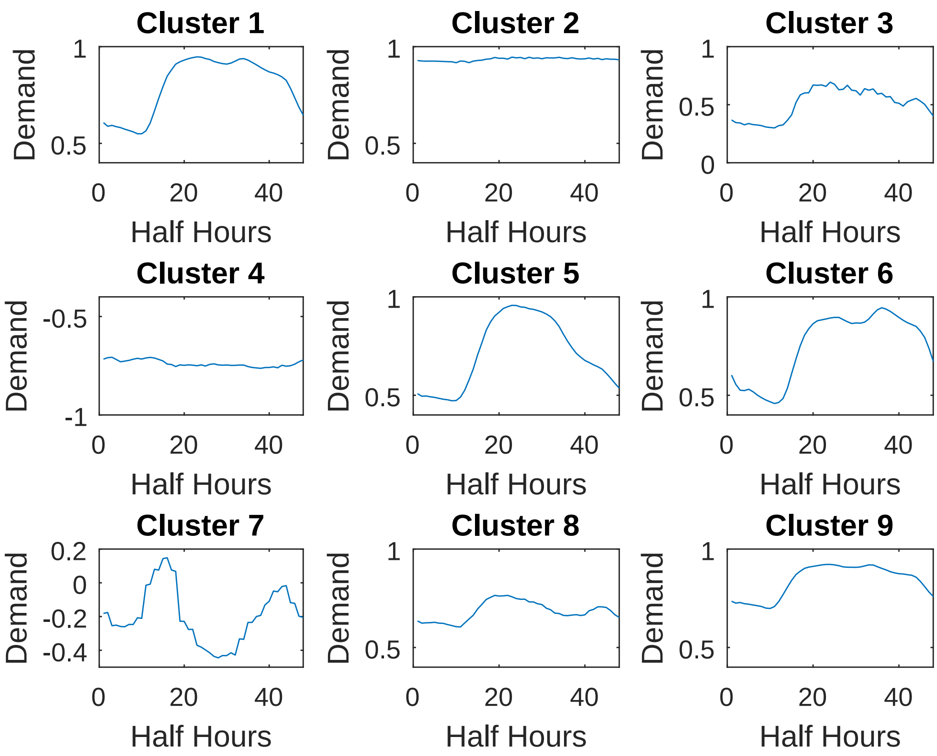


Figure 6.11: Summer Clusters Gained By K-means With K=9

Table 6.12: Summer Classification Labels

Cluster	Category
1	Large Commercial Urban (LCU)
2	Small Domestic Rural (SDR)
3	Small Commercial Urban (SCU)
5	Medium Commercial Rural (MCR)
6	Medium Commercial Urban (MCU)
7	Small Commercial Rural (SCR)
8	Small Domestic Urban (SDU)
9	Medium Domestic Urban (MDU)

The approach to creating the classification labels seen in Tables 6.12 & 6.13 is the same as for Table 6.9 on a different subset of load data.

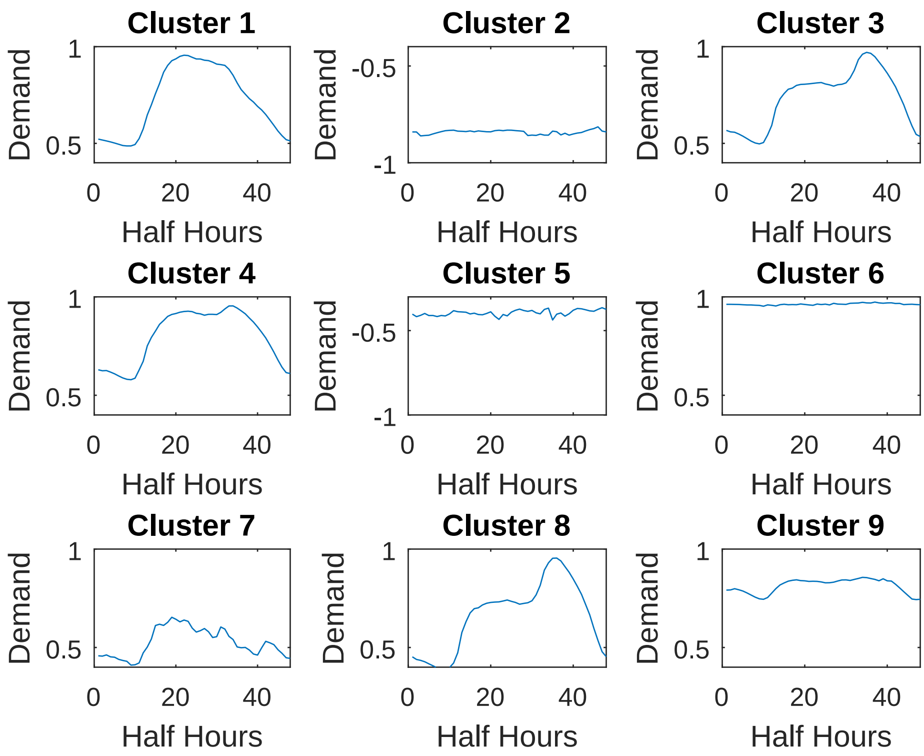


Figure 6.12: Winter Clusters Gained By K-means With K=9

Table 6.13: Winter Classification Labels

Cluster	Category
1	Medium Commercial Rural (MCR)
3	Large Domestic Urban (LDU)
4	Large Commercial Urban (LCU)
6	Small Domestic Urban (SDU)
7	Small Domestic Urban (SDU)
8	Medium Domestic Rural (MDR)
9	Medium Domestic Urban (MDU)

6.7.2 Methodology and Experiment

In order to test the effectiveness of the classification labels in Tables 6.12 & 6.13, an experiment using some of the classification algorithms discussed in Chapter 3 is performed. The evaluation will be considered for both a summer and winter period so that both sets of labels in Tables 6.12 and 6.13 are used. All classification algorithms are used in MATLAB in the “Statistics” toolbox. For K-nearest neighbours, the parameters are set to the default apart from the k number of neighbours which is chosen based on the lowest error following 10-fold cross validation. For naïve Bayes, the box kernel distribution will be used which is more adept to continuous variables and does not require normality. For decision trees, Gini’s diversity index will be used as the binary splitting criterion and a pruning stage will occur after the tree is built.

The main experiment involved here is the gradual buildup of the classification labels and the effect on classification error. The training data is the principal component scores for each of the 35 descriptors given in Section 4.3.2. Firstly, the classifiers will be used on only domestic/commercial labels which constitute Class1. These are tested first as they constitute the first principal component of the PCA so they explain the most variance in the data. Then the rural/urban descriptors will be added for the Class2 labels. Finally the population size descriptors will be added so that the full labels known as Class3 in Tables 6.12 and 6.13 are tested. Additionally, the summer and winter periods add the suffix “S” or “W” in order to distinguish the time periods that are considered here. The justification for using this methodology is to determine at what stage the qualitative principal component description is no longer useful. If the classification methods can accurately determine the labels based on the full set of PC scores then it can be said that the qualitative labelling has been applied correctly. If it cannot, then the process has not been effective.

At each of the three stages, both the training error and 10-fold cross validation error will be recorded. The training error is

$$\mathrm{err}_{\mathrm{train}} = \frac{1}{n}\sum_{i=1}^{n} L(f(x_i), g(x_i)) \tag{6.7.15}$$

which is the average of the classification 0-1 loss function

$$L(f(x), g(x)) = \begin{cases} 0, & \text{if } f(x) = g(x) \\ 1, & \text{otherwise} \end{cases} \tag{6.7.16}$$

Then the k-fold cross validation is:

$$\mathrm{CV}_k = \sum_{k=1}^{K} \frac{n_k}{n} \mathrm{err}_{\mathrm{train}} \tag{6.7.17}$$

where $n_k = \frac{n}{K}$ is the kth subset of n training observations. In this instance, $K = 10$ is used.

By comparing the methods across the various classification methods and between seasons, a good understanding of their effectiveness and role in future work will be gained.

6.7.3 Overall Algorithm Results

Table 6.14: Classification Method Errors for Summer

	Class1S		Class2S		Class3S	
	T. Err.	CV Err.	T. Err.	CV Err.	T. Err.	CV Err.
KNN	0.159	0.198	0.260	0.368	0.422	0.553
DT	0.034	0.245	0.108	0.389	0.157	0.615
NB	0.073	0.186	0.202	0.440	0.372	0.672

Table 6.15: Classification Method Errors for Winter

	Class1W		Class2W		Class3W	
	T. Err.	CV Err.	T. Err.	CV Err.	T. Err.	CV Err.
KNN	0.209	0.301	0.361	0.510	0.432	0.598
DT	0.058	0.318	0.129	0.559	0.174	0.579
NB	0.142	0.389	0.363	0.573	0.404	0.692

As there are many comparisons to be made, it is important for the discussion to be conducted in a systematic approach. The results will be discussed in the context of one classification algorithm at a time. Tables 6.14 and 6.15 contain the general algorithm errors across both seasons. These will be discussed first. Then there will be two more tables per classification algorithm which will contain the

Table 6.16: KNN Errors for Summer

Class1S	Error	Class2S	Error	Class3S	Error
D	0.537	DU	0.611	SDU	0.850
				MDU	0.640
		DR	0.769	SDR	0.692
C	0.045	CU	0.093	SCU	1.000
				MCU	0.592
				LCU	0.130
		CR	0.382	MCR	0.500

Table 6.17: KNN Errors for Winter

Class1W	Error	Class2W	Error	Class3W	Error
D	0.145	DU	0.298	SDU	0.833
				MDU	0.480
				LDU	0.369
		DR	0.513	MDR	0.628
C	0.302	CU	0.370	LCU	0.267
		CR	0.352	MCR	0.537

misclassification errors for each individual label for both seasons. This will give some further insight as to how the algorithm is performing at each stage. These discussions will be supplemented by algorithm specific figures.

6.7.4 K-Nearest Neighbours

In Table 6.14, KNN has an increasing error in both the training and cross-validation errors as more information is added to the labels. Comparatively to the other algorithms, it had performed the worst in terms of training error although its cross validation error is comparable or even improved in some instances. The same holds true for the winter errors in Table 6.15 although they are all larger across all classification methods.

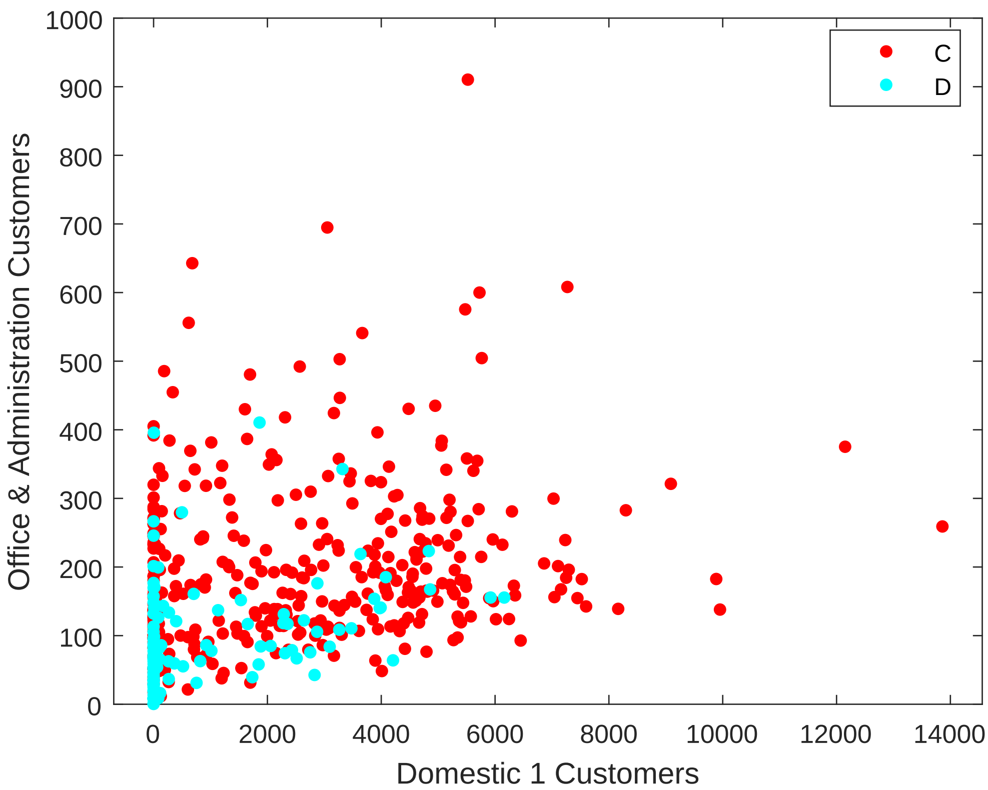


Figure 6.13: Scatter Plot of Number of Domestic 1 Customers against Number of Office & Admin Customers with Class1S labels

In order to gain some clarity, Tables 6.16 and 6.17 will be examined to look at the KNN errors in more detail. In Table 6.16, it is seen that the majority of errors in Class1S arise from the domestic label. The error implies that in many instances, substations which should be classified as domestic actually have commercial nearest neighbours in the 35 dimensional space. This is most likely due to the fact that the commercial customer label outnumbers the domestic customer label by a ratio of more than 3:1 and classification by nearest neighbour across all descriptors evenly results in few distinctions being made in the predictors where differences may lie. For example, by examining the scatterplot in Fig. 6.13 of the most populous domestic descriptor Domestic 1 against the most populous commercial descriptor Office & Admin, there is no clear distinction in the concentration of commercial and domestic customers.

An examination of Class2S in Table 6.16 yields higher error than Class1S which is expected based on the increase in class errors across Table 6.14. The labels DU and DR still had high error but of note was the fact the rural labels DR and DCR had a higher classification error than their urban counterparts DU and CU. This was due to a combination of the fact that both DR and CR are not as well populated as the urban customers and in general the rural domestic customers in the original space are less populous. Therefore, the nearest neighbour search would be expected to perform worse with a less distinct customer base.

At the final stage of Class3S in Table 6.16, the error is once again higher than the previous class. The addition of population size does not increase the accuracy and makes classification more difficult by further decreasing the population size of the labels. Indeed, a recurring pattern is that labels in Class3 with the same Class2 label will have the smallest error in the "L" or "M" label as opposed to "S". In terms of daily load profiles, this is expected due to substations with a smaller population size having less predictable demand.

Now by analysing Table 6.17, it can be seen how the winter period affects the classification errors of KNN. Contrary to the summer period, more commercial substations are misclassified in Class1W than domestic. Again, this can be related to the populations of the classes "C" and "D" where "D" is more populous than "C". Because the difference in population is not as great as in the case of the summer, the error difference is negligible.

In Class2W of Table 6.17, the errors have increased as in the case of summer. As previously, DR has a higher error than DU but the CR error is comparable to CU. This result is surprising but it could be due to these populations being distinct enough in their customer make up such that their overall population size does not have any adverse effects.

In Class3W of Table 6.17, the errors have increased across the table and similarly to the summer, the larger population sizes tend to have the lowest error. Overall, it can be said that the population size of the classification label is key in terms of the accuracy of KNN. KNN performed poorly for the summer through all class levels due to this population imbalance. For the winter, because the population differences between C and D were not as high, this led to improved results. Across

Table 6.18: Naïve Bayes Errors for Summer

Class1S	Error	Class2S	Error	Class3S	Error
D	0.102	DU	0.021	SDU	0.000
				MDU	0.333
		DR	0.385	SDR	0.385
C	0.064	CU	0.104	SCU	0.167
				MCU	0.418
				LCU	0.260
		CR	0.838	MCR	0.779

Table 6.19: Naïve Bayes Errors for Winter

Class1W	Error	Class2W	Error	Class3W	Error
D	0.239	DU	0.349	SDU	0.000
				MDU	0.760
				LDU	0.492
		DR	0.269	MDR	0.282
C	0.000	CU	0.259	LCU	0.148
		CR	0.815	MCR	0.815

both summer and winter though, the errors across Class 3 are too high to be useful apart from those classes with an L label.

6.7.5 Naïve Bayes Results

The overall errors in Table 6.14 and 6.15 show that it performs the worst against the other classification methods in terms of cross validation error across both seasons. The training error is more favourable when compared to the other classifiers, therefore the algorithm is either overfitting or the underlying assumptions of the method are inappropriate for the data. As the implicit assumption of the independence of the algorithm make overfitting unlikely Bayes classification method is most likely ill-suited to a problem of this type due to its high bias. However, it is still performs

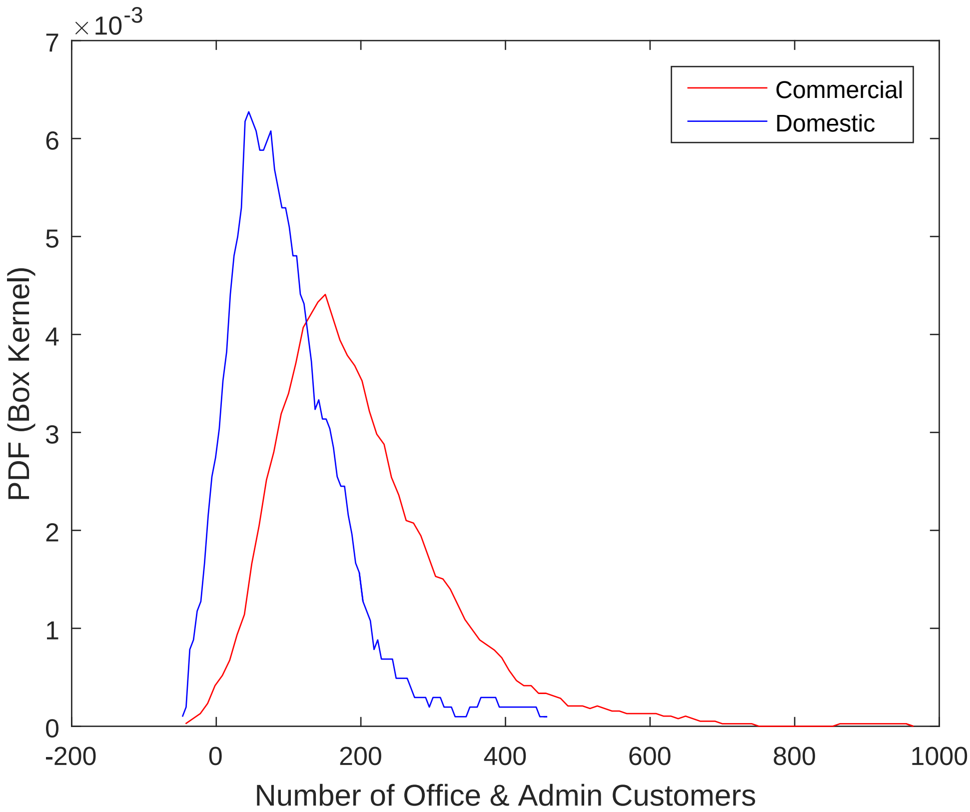


Figure 6.14: PDFs of Office & Admin Customers

well when the classification problem is in its simplest Class1S and Class1W forms.

For the naïve Bayes classifier, Tables 6.18 and 6.19 will be used to gain insight between the seasonal periods and across the class labels. In Table 6.18, Class1S shows a minimal difference between the errors of C and D. The errors themselves are also small showing this classifier to be effective at this stage. Indeed, Fig. 6.14 shows the probability density functions (PDF) between the commercial and domestic customers for the most populous commercial descriptor (Office and Admin). It can be seen here that a larger number of customers indicates a higher likelihood of a commercial classification label. Note that there are some artifacts for negative values of customers having a PDF value greater than 0 due to implementation of the method in MATLAB. Using the pdfs across all descriptors should give a relatively accurate picture in the case of two classification labels.

When considering the four labels in Class2S, naive Bayes was far less accurate. In particular, the rural labels DR and CR performed worse than the urban labels

Table 6.20: Decision Tree Errors for Summer

Class1S	Error	Class2S	Error	Class3S	Error
D	0.074	DU	0.263	SDU	0.600
				MDU	0.173
		DR	0.231	SDR	0.615
C	0.022	CU	0.045	SCU	0.667
				MCU	0.163
				LCU	0.065
		CR	0.132	MCR	0.118

DU and CU. For DR, it could reasonably be attributed to variance due to its small population but CR is well populated. The CR population was mostly misclassified as CU as its pdf for most descriptors was captured by CU.

When considering Class3S in Table 6.18, the errors have mostly increased. Unlike in the case of KNN, the errors do not necessarily seem to be tied to their population sizes. The label SDU has no misclassifications while the more greatly populated MDU misclassifies a third of its substations. Overall, the summer classifications follow the general rule that the simpler classification labels perform the best.

Table 6.19 shows the Bayes classification errors for winter. It can immediately be seen that in Class1W, there was no error for the commercial label. That is, no domestic customers were incorrectly labelled as commercial customers. In Class2W, as with the summer period, the CR label was the most prone to errors with the majority being incorrectly classified. The other three labels in Class2W showed a rise in error although not as significant as the CR error. Finally in Class3W, it can be seen that SDU and SCU have lower errors to their more heavily populated counterparts. This is the opposite effect of what is seen in Tables 6.16 and 6.17 for KNN and reflects the strength of Bayes in good performance with less training data.

Table 6.21: Decision Tree Errors for Winter

Class1W	Error	Class2W	Error	Class3W	Error
D	0.040	DU	0.066	SDU	0.611
				MDU	0.340
				LDU	0.100
		DR	0.295	MDR	0.231
C	0.085	CU	0.096	LCU	0.119
		CR	0.204	MCR	0.111

6.7.6 Decision Tree Results

The overall errors in Table 6.14 and Table 6.15 show that the decision tree classification method performed the best compared to the other methods in terms of training error. Although the cross validation error was not always the lowest, it was at least comparable to the others.

Table 6.20 shows the classification errors for decision trees during the summer. The Class1S errors are low for both D and C customer labels showing that decision trees perform well at this general problem. In Class2S, the errors have increased as with the other classification methods but none of the errors are above 0.3 which is considerably better than either KNN or naïve Bayes. This is only a consideration of the training error but it is still of note. Also in Class2S, the summer commercial customers were easier to classify than the domestic customers. In Class3S, the more well-populated labels had lower errors than their low population counterparts contributing to the concept that the more well populated clusters have less variance in their customer profiles and are therefore easier to classify.

The winter errors for decision trees in Table 6.21 are similar to the summer errors. For Class1W, the classification errors are both low and there is no distinction between the C and D labels. For Class2W, there is an increase in overall error and notably, the rural classification labels have higher errors than the urban labels. Surprisingly for Class3W, many of the classification errors are quite low. Only SDU and MDU have classification errors above 0.3. As the corresponding cross validation error is not low, it indicates that some overfitting has occurred with this particular decision tree model.

6.7.7 Comparing Methods

As this is an investigation of the effectiveness and usefulness of the customer classification labels, it is not necessarily important how well the classification methods perform on their own. Rather the knowledge gained about the labels through analysing the classification methods is needed.

By considering the errors in Tables 6.12 & 6.13, a general comparison can be made. In terms of the lowest classification error, decision trees had the lowest training error across all classes and both seasons. However, trees did not have the lowest cross-validation error in all cases, suggesting that there is some overfitting. With k-nearest neighbours, although the method had the highest training error in many cases, it also had the lowest cross-validation error for Class2S, Class3S, Class1W and Class2W. Decision trees only had the lowest CV error for Class3W and naïve Bayes had the lowest CV error for Class1S. Therefore, overall KNN performed the best by the classification error metric. KNN is the least likely to overfit without doing an extensive amount of parameter tuning as the whole space of parameters are searched unlike in the case of naïve Bayes and decision trees where subsets of the space are searched along subsequent steps.

Due to this investigation, it is clear that simpler classification labels such as "D" and "C" can sometimes be more insightful than the detailed description. This is due to the fact that the errors for the Class1 and Class2 cases are significantly lower. In some cases, a lack of a good data population size makes some of the Class3 labels irrelevant. It is proposed that the Class1 labels of D and C as well as the Class2 labels DU, DR, CU and CR be implemented into the forecasting method for additional qualitative understanding. The classification labels in Tables 6.12 and 6.13 are related to respective clusters which are used to build multiple ANFIS models for improved forecasting.

6.8 Conclusion

The research questions addressed in this chapter used RQ1 and RQ2 to support RQ3, RQ4 and RQ5. That is, the CL-ANFIS method considers both internal factors such as demand variation and external factors such as customer type as posed in RQ1 and RQ2. Through the use of rules and membership functions, it also provides greater transparency as posed by RQ3 and presents them in a qualitative fashion as posed by RQ4. Some preliminary evaluation work is also done to address RQ5. Discussions on RQ3 and RQ4 are closed in this chapter. By considering the mixture of customers on the substation and the rule space generated by CL-ANFIS, insight can be obtained in addition to the numeric forecast. Rule sets using the membership functions can be obtained with qualitative labels such as low, medium and high. This qualitative aspect in the forecasting process is important for the DNOs in the planning stages. Therefore, it is recommended that the DNO considers and adopts this methodology in their load estimation and forecasting process.

The long-term forecast in electrical distribution is not often addressed because of the difficulty and scope of the problem. However, for the purposes of planning on the network and greater justification for reinforcement on substations, an informed long-term forecast is still required. It was shown that the clustered CL-ANFIS forecast performed better than the naïve forecast of simply applying the unclustered ANFIS model and the benchmark Holt-Winters. For clusters 1-5, the improvement was clear with an average forecasting improvement by 6%. For clusters 6 and 7, CL-ANFIS did not outperform the unclustered model and Holt-Winters. However, this could be explained through a further analysis of the average customer population size and maximum demand values of substations in these clusters.

A closer examination of the qualitative labelling methods showed that in general, the simplest labels of domestic and commercial were well defined. Adding greater complexity by the addition of the urban/rural and small/medium/large descriptors attempts to offer greater transparency but in some cases, they are not fit for purpose. Therefore the recommendation is to rely mostly on the domestic/commercial customer distinction within the clusters to be the most accurate labelling and the others are to be used more sparingly. Through a combination of these qualitative labels, the historical membership functions and the Takagi-Sugeno fuzzy ruleset, a good basis of understanding can be formed of the substations in each cluster.

Chapter 7

Evaluation and Testing

After the creation of the CL-ANFIS model in Chapter 6, testing must be conducted to show the benefits of the model against the NPG method and the robustness of the model.

Some of the initial errors based on three years of training and two years of testing are given in Chapter 6. Here, a more in-depth approach will be taken to discuss the errors. It is important to show from both an academic and commercial perspective that this algorithm can provide an impact on the status quo at NPG either by the accuracy of the predictions, the insight of qualitative labelling or a combination of the two.

In Section 7.1, the daily load profile clusters will be examined more closely in terms of their deviations in order to determine their validity. In Section 7.2, the robustness of the algorithm will be checked based on the amount of input data given to the algorithm. In Section 7.3, a direct comparison against the NPG DLEs with a focus on the maximum demand figures will be made. Finally, Section 7.4 will summarise the discoveries.

7.1 Examination of Daily Load Profile Clusters

The clusters found in Fig. 6.4 are produced through K-means clustering after the determination that it would be appropriate method for the data. As part of the evaluation process, it is important to analyse the variance within these cluster groups.

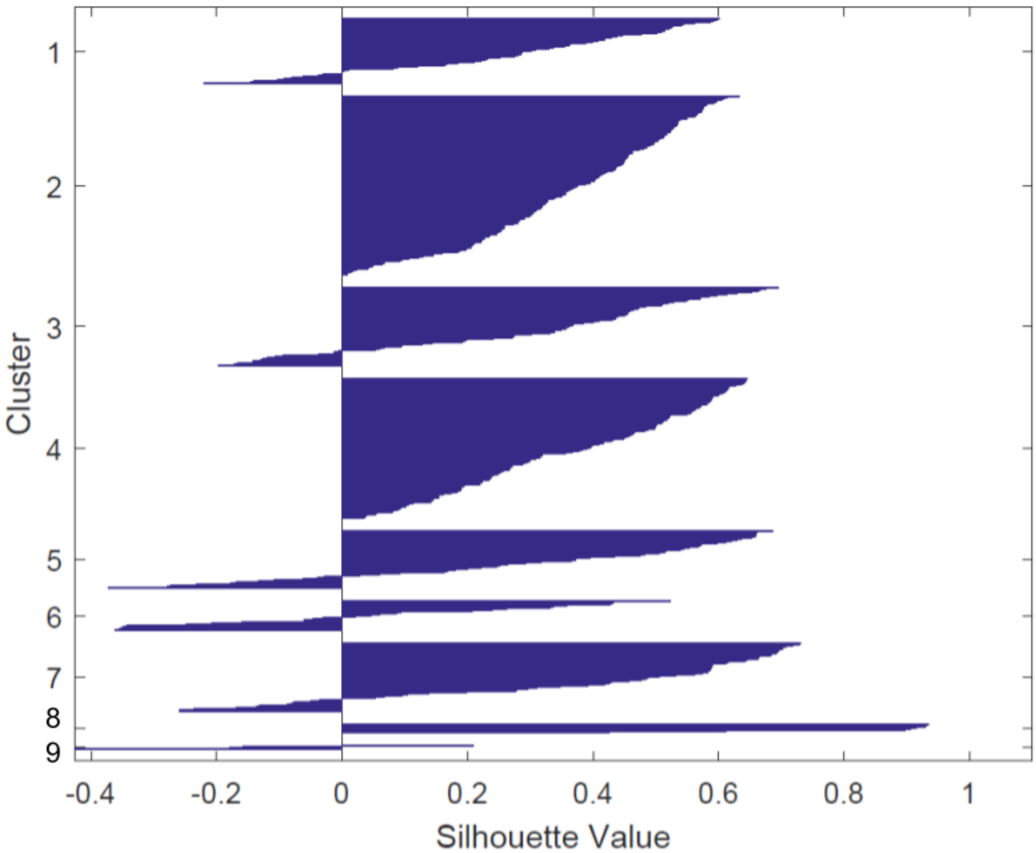


Figure 7.1: Silhouette Plot for Daily Load Profile Clusters

The optimisation of the K-means problem can be expressed as:

$$min_{C_1,...,C_K}\left\{\sum_{k=1}^{K} W(C_k)\right\} \tag{7.1.1}$$

where $C_1, ..., C_K$ are the sets of observations for K clusters and $W(C_k)$ is the amount that the data points within the cluster differ from each other. In this algorithm, $W(C_k)$ is taken to be the pairwise Euclidean distance function.

A methodology specific to clustering is known as the silhouette criterion given by Rousseeuw (1987). First let a data point i be in cluster A. Then $a(i)$ is the average dissimilarity of point i to all other objects in cluster A. Then let C be a cluster different from A. Then $d(i, C)$ is the average dissimilarity of point i to all other objects in cluster C. Then for all $C \neq A$,

$$b(i) = \min_{C \neq A} d(i, C) \tag{7.1.2}$$

Table 7.1: Average Variance of Daily Load Profile Cluster Values

Cluster	Av. Variance
1	0.00411
2	0.00200
3	0.00317
4	0.00192
5	0.00520
6	0.01148
7	0.00454
8	0.05168
9	0.10402

Then the silhouette value $s(i)$ is

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{7.1.3}$$

where $s(i) \in [-1, 1]$. This $s(i)$ ratio intuitively implies that if $a(i) << b(i)$ that $s(i)$ will be close to 1 and the data point i can be considered to be clustered well. That is, the in-cluster dissimilarity is smaller than the out-of-cluster dissimilarity so the point i is closer in value to other points in A. Alternatively, if $a(i) > b(i)$ then $s(i) < 0$ which implies that i is closer to some $C \neq A$ and is therefore misclassified.

From Fig. 7.1, most of the values are larger than 0. Clusters 1-5 contained substations which forecasted well based on the work presented in chapter 5. Out of those, clusters 1, 3 and 5 do have a small portion of substations which are deemed to be misclassified. Clusters 6 and 7 contained substations which performed poorly based on the ANFIS forecasting presented in Chapter 5. Approximately 50% of the substations in cluster 6 had a negative silhouette value which implies many were misclassified while cluster 7 only had a small portion of negative silhouette values. This gives some initial insight into the performance of K-means clustering and the level of impact a poor solution might have.

An average variance for each daily load profiles can now be calculated. See Appendix A for figures on the spread of the load profiles per each cluster. Table 7.1

Table 7.2: Average Variance of Daily Load Profile Cluster Values After Negative Silhouette Values are Removed

Cluster	Av. Variance
1	0.00362
2	0.00200
3	0.00318
4	0.00192
5	0.00519
6	0.01321
7	0.00454
8	0.05168
9	0.00000

shows the average variance of each clusters daily load profiles. Clusters 2 and 4 clearly have the smallest variance which is expected both from their accurate forecasts in Chapter 6 and the silhouette values in Fig. 7.1. Clusters 8 and 9 have the highest variance which is expected as there are less than 10 substation in both and cluster 8 is exclusively generating substations. For those substations which are forecasted on, clusters 6 and 7 show a higher variance in general in comparison to clusters 1-5. Again, this is corroborated by the difficulty in forecasting for them in Chapter 6.

In order to see the effect of those substations with a negative silhouette value, they can be excluded from the calculation of variance in Table 7.1. The results of this are given in Tables 7.2. Clusters 2, 4 and 8 have the same variance as before due to the fact that they have no negative silhouette values. However, out of those remaining, only cluster 1 saw a notable improvement in terms of reduced variance (as cluster 9 only contained one instance out of four total, it is excluded from this consideration). The variance of the other clusters either remained the same or even increased in some instances. By visualising the load profile clusters in Appendix A, it is possible to see why this is the case. The deviation of those load profiles with negative s values from the mean is not significant in most instances. This leads to the conclusion that the clustering algorithm performed well and the few misclassified substations are not significantly different from the cluster they were placed in.

7.2 Evaluating CL-ANFIS with Subset Input Data

As part of the evaluation, a check of the robustness of the algorithm must be conducted. In this section, the term "robust" is used to mean the effectiveness of CL-ANFIS subject to varying amounts of training data. The same evaluation and methodology conducted in Chapter 6 will be completed here where a fraction of the dataset will be used to train and test the algorithm in multiple instances.

As the CL-ANFIS algorithm is in fact multiple ANFIS models based on clusters, an explicit ratio must be applied to each cluster. For those clusters which are not well populated, this should have an adverse effect on the results.

7.2.1 Method for CL-ANFIS Evaluation with Subset Input Data

The CL-ANFIS method will be evaluated for robustness on the individual cluster groups. In order to train the CL-ANFIS models, a subset of the original training data used in Section 6.3 will be used. For each cluster, 10 subsets of the original data set will be evaluated between 10% to 100% of the original data set in 10% increments. That is, for a cluster of size n substations and $m = 1, 2, ..., 10$ intervals, $(0.1 \times m) \times n$ substations will be used for training and evaluation using the same methodology as outlined in Section 6.3. In order to take into account some of the variance in selecting substations randomly, this method will be run 10 times with different random initialisations such that different subsets of substations are selected for evaluation. The mean absolute error (MAE) and root mean squared error (RMSE) will also be calculated in the same way as outlined in Section 6.3.

Table 7.3: Median MAE values of Robustness Evaluation for CL-ANFIS [Averages based on 10 different random initialisations]

	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Clus. 1	0.045	0.070	0.088	0.280	0.096	0.101	0.159	0.103	0.061	0.085
Clus. 2	0.076	0.059	0.077	0.108	0.097	0.077	0.192	0.116	0.067	0.127
Clus. 3	0.052	0.057	0.084	0.262	0.287	0.102	0.086	0.124	0.108	0.109
Clus. 4	0.109	0.076	0.132	0.085	0.133	0.094	0.127	0.165	0.120	0.115
Clus. 5	0.182	0.342	0.445	0.295	0.230	0.196	0.160	0.167	0.173	0.137
Clus. 6	0.628	0.274	1.063	0.140	0.181	0.466	1.854	1.091	0.120	0.376
Clus. 7	0.204	0.112	0.084	0.936	0.421	0.201	0.266	0.538	0.488	0.782

Table 7.4: Median RMSE values of Robustness Evaluation for CL-ANFIS [Averages based on 10 different random initialisations]

	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Clus. 1	0.105	0.220	0.237	0.574	0.285	0.236	0.434	0.342	0.151	0.279
Clus. 2	0.148	0.147	0.149	0.239	0.274	0.177	0.426	0.278	0.158	0.247
Clus. 3	0.093	0.164	0.218	0.457	0.497	0.291	0.334	0.405	0.226	0.29
Clus. 4	0.234	0.183	0.257	0.199	0.235	0.197	0.217	0.493	0.247	0.236
Clus. 5	0.260	1.321	1.407	0.544	0.498	0.387	0.307	0.339	0.336	0.269
Clus. 6	1.290	0.776	1.749	0.355	0.388	1.167	3.051	1.766	0.186	0.531
Clus. 7	0.706	0.233	0.154	1.778	1.030	0.621	0.566	1.362	1.063	1.596

7.2.2 Results and Discussion of CL-ANFIS Evaluation with Subset Input Data

Before discussing the results, it is important to note the relative cluster sizes that are taken in these 10% increments. Table 6.4 gives the data set size of each cluster. Taking 10% increments of a small set of data will lead to overfitting in some cases (as in the case where there is only one substation for Clusters 1 and 3) and insufficient data in other cases. Only clusters 2 and 4 had a reasonable data set size of near 100 substations where taking 10% increments updated the training set by a sufficient

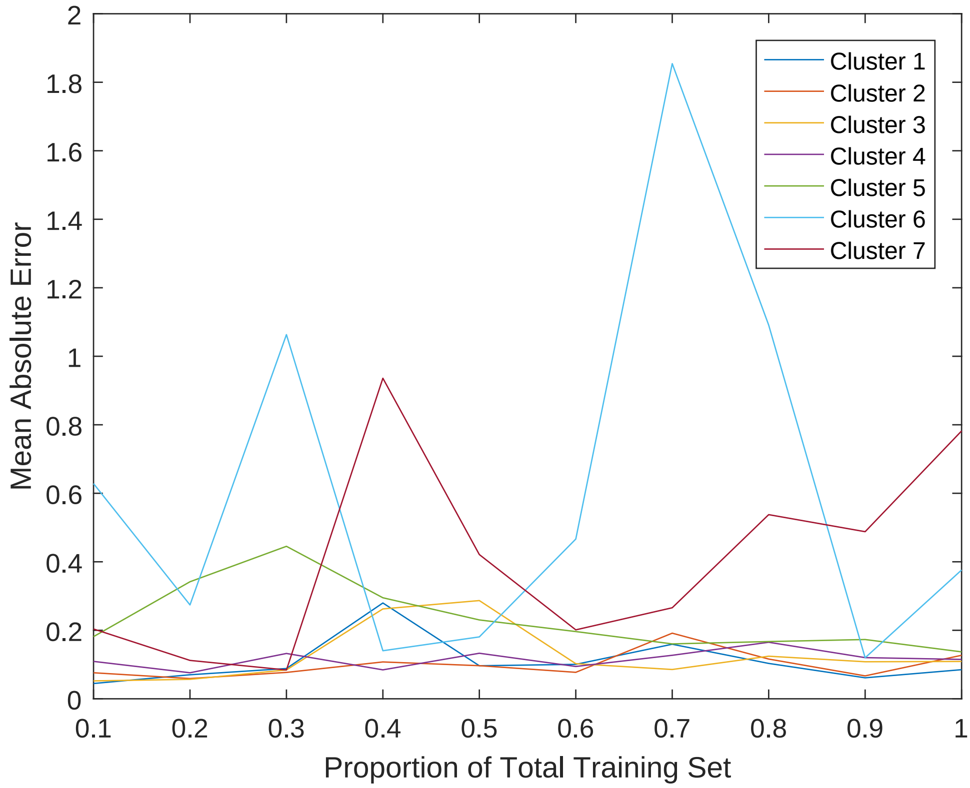


Figure 7.2: Median MAE values of Robustness Evaluation for CL-ANFIS

amount to make meaningful conclusions. In contrast for the smaller sized clusters, adding 1-3 substations with these 10% increments could lead to large fluctuations in forecasting error if the load profile patterns are significantly different.

Taking the cluster sizes into account, the results in Table 7.3 and Table 7.4 show that for clusters 1-5, the errors have a small variance across the subsets of the training data. This shows robustness and stability in that inputting more training data will not greatly change the forecasting errors.

Cluster 1 shows an unusually forecasting error at 40% of total training set size but otherwise shows a flat error rate. This suggests that the additional substation added at 40% was significantly different from the first substation used in the training set but after 50% of the training set was used, the error returned to a consistent rate. At such a small cluster size, adding substations with a substantially differing load profile can affect the overall CL-ANFIS model. However, the fact that there was stabilisation afterwards is consistent with the total cluster having low forecasting

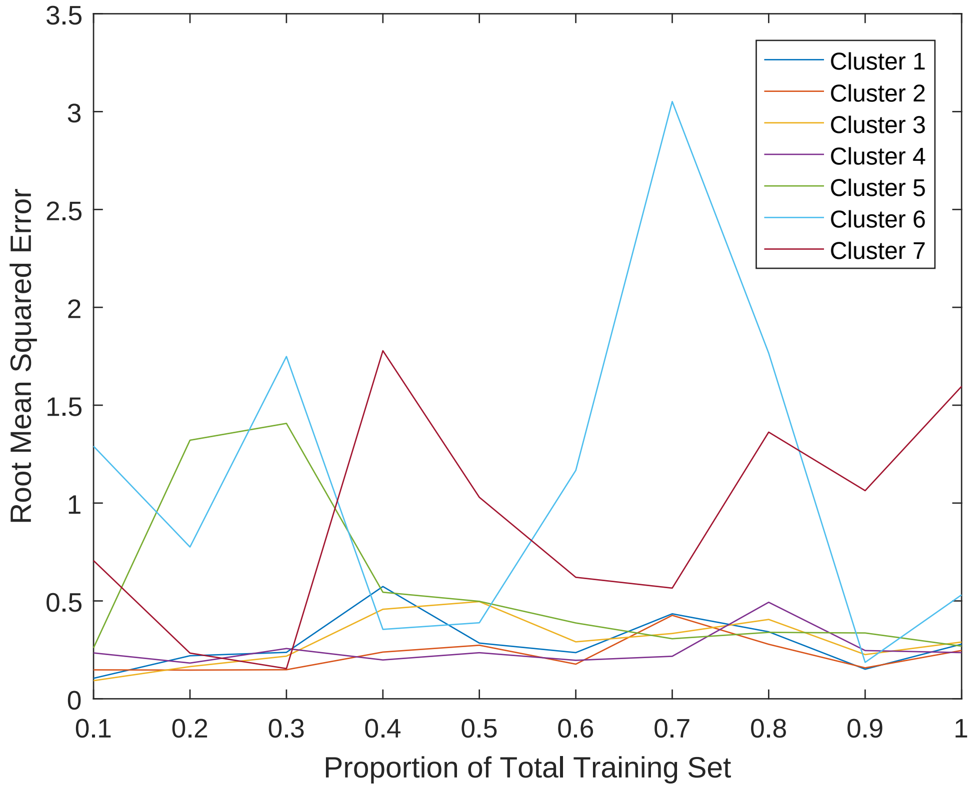


Figure 7.3: Median RMSE values of Robustness Evaluation for CL-ANFIS

error.

Cluster 2 outputs a relatively flat error rate throughout Table 7.3 and Table 7.4. Due to the sufficiently large cluster size, it is suggested to treat these results as representative of the effect on adding new substations to the network. As cluster 2 had one of the lowest forecasting errors, it is encouraging to see this invariance in error rates.

Cluster 3 also has a relatively small variance in error rate but the cluster size is the smallest of all the clusters containing only 6 substations. So upon adding 10% more data, there will be 1 or 0 substations added to the total cluster size. This shows that the substations added to the cluster do not materially affect the forecasting performance of the ANFIS model. This further supports not only the accuracy of the model but also similarity of annual load profiles in the cluster.

Cluster 4, along with Cluster 2 has one of the largest cluster sizes. Because of this, the relatively flat error rate indicates that adding substations with similar daily load profiles does not materially affect the forecasting accuracy of the cluster group.

Cluster 5 begins with a higher error rate and then returns to a lower one in higher proportions of training data. This suggests that adding new substations to this cluster improves the performance significantly.

Clusters 6 and 7 have a much greater variance of errors as seen in Fig. 7.2 and Fig. 7.3. As there are some sporadic increases in error at a greater amount of training data, it is clear that adding more substations which might fall into these categories does not necessarily mean the forecasting accuracy of CL-ANFIS will increase. However, there are also points at which for both clusters 6 and 7, there are acceptable levels of forecasting error. This suggests that there are optimal points at which the cluster might have a mixture of substations with an acceptable forecasting performance but this will by definition change the daily load profile definition of the cluster. There are two suggested approaches to this:

1. Find the optimal mixture of substations on this cluster which maximises the cluster size and minimises the MAE/RMSE values.

2. Address the substations in these clusters on a case-by-case basis.

As the sizes of these clusters are relatively small and volatile to new additions of substations, it is therefore recommended that those substations which fall into these clusters be treated on a case-by-case basis.

7.3 Comparison against NPG DLEs

As a final analysis, the output of the CL-ANFIS model should be compared to the status quo of the NPG DLEs. This includes both a direct comparison of the forecast MDs as well as a description of the qualitative information the engineers can expect to have.

7.3.1 Method for Comparsion against NPG DLEs

The forecast in the NPG DLEs is based on the maximum demand figure (MD) as detailed in Chapter 2. In the DLEs, there is an eight year long term forecast reported based on these MDs. Due to the availability of training data, a direct evaluation of an eight year forecast cannot be conducted. However, a forecast can still be produced for a shorter time period and an example will be presented in this section.

In order to do a comparison of the CL-ANFIS method, there should be a direct comparison of MD figures. After the complete annual forecast profile is determined with CL-ANFIS there are two possible approaches for the selection of MD:

1. Replicate the MD selection method of NPG
2. Use automated MD selection from Chapter 5

As there is a substantial amount of human input to the selection of MD for NPG, for the purposes of this analysis the automated MD method detailed in Section 5.1.1 will be used.

The forecast to compare will be the output of the experiment detailed in Chapter 6. That is, the CL-ANFIS is trained on 3 years of demand data for the period April 2010 - March 2013 and a 2 year forecast is produced for the period April 2013 - March 2015.

The DLEs from NPG that will be used in this evaluation will be the ones reported for the period of April 2012 - March 2013. This is analogous to the training data used to produce the 2 year forecast in Chapter 6. The two MD figures reported in the forecast for 2014 and 2015 will be compared against the MD figures obtained from the 2 year forecast.

7.3.2 Results and Discussion of NPG Comparison

The results show that in most instances the NPG maximum demand forecast of 0% or 0.5% growth outperforms the CL-ANFIS forecast with the automated maximum demand selection method.

This is not necessarily a surprising result given the amount of human bias and adjustments that are made during the MD forecast and MD selection process. The main areas of discrepancy that could be favouring the NPG MD forecast include:

Table 7.5: Median Percentage Errors for MD Forecast

Cluster	CL-ANFIS 2014	NPG 2014	CL-ANFIS 2015	NPG 2015
1	0.102	0.046	0.329	0.043
2	0.233	0.076	0.342	0.073
3	0.076	0.075	0.088	0.122
4	0.120	0.074	0.148	0.086
5	0.126	0.079	0.124	0.083
6	0.412	0.026	0.216	0.077
7	0.434	0.069	0.254	0.113

1. Human selection of maximum demand in about 50% of all substations

2. New demand or generation connections added to the forecast

3. Demand being transferred from one substation to another

Regarding the first point, the human selection of maximum demand from the actual demand profile will severely bias the results as human correction is made post receiving the substation load profile from SCADA.

Secondly, new demand or generation connections are typically added to the NPG forecast in the near future (within 1 or 2 years of the current year MD). Given the small sample size of data, these new connections are not practical to consider for the forecast using external information such as economic information. This gives the NPG forecast a further advantage as maximum demands are adjusted in these cases where known new connections arise.

Thirdly, demand transfers are difficult to forecast for. It is possible to adjust load profiles where a load transfer might be detected as discussed in Section 5.2.2 but this is only for historical load profiles. Again, the sample size here is too insignificant to account for these situations but when they do occur they can make a material impact on the forecast which is accounted for when NPG does their maximum demand selection.

For the 1 year ahead forecast, Clusters 1, 3, 4, and 5 give forecasting results in an acceptable range for CL-ANFIS. The 1 year ahead forecasts align with the forecasting results in Table 6.4 with the exception of cluster 2. This shows that although the in-sample error of the forecast is within a reasonable error bound for NPG, there is a discrepancy between the MD selection of NPG and the automated MD selection used here. There are some demand peaks in the substations in Cluster 2 which are not detected by the wavelet method detailed in Section 5.1.2. If the change is not sharp enough, it may not be considered by the method as an outlier but it could still be materially different enough from the MD selected by NPG to cause a great difference in the final MD selected.

Cluster 6 and 7 showed high error rates for the CL-ANFIS both over 40% when compared to the actual MD for 2014. This is not unexpected and falls in line with previous forecasting errors in Chapter 6 and Section 7.2.

For the 2 year ahead forecast, only Clusters 3, 4 and 5 give acceptable forecasting errors. For Clusters 1-5, the forecasting percentage errors either increased or stayed approximately the same between 2014 an 2015 predicted MDs. This would be expected as the forecast will produce continually worse results the further into the future a prediction is required. In particular, Clusters 3-5 still give favourable forecasting results within 15% of the actual maximum demand for 2015. Clusters 1 and 2 however have a noticeable increase in forecasting error which is above 30% in both cases. Due to the small cluster size for Cluster 1 training the model, this is not necessarily surprising.

On the other hand, the forecasting errors for Clusters 6 and 7 decreased upon going from 2014 to 2015. This is due to CL-ANFIS showing a more regular forecasting load profile over longer periods of time. There are less likely to be any errant spikes in the forecast leading to lower maximum demands.

The final items to note in the comparison between the NPG forecast and CL-ANFIS method is the ancillary information available. As mentioned in Chapter 2, the NPG DLEs has some ancillary information such as firm capacity, power factor, new customer connections and new customer generations. From CL-ANFIS the main ancillary information are as follows:

- Clusters of substations with similar daily load profiles

- Qualitative customers classifications with varying levels of depth from a PCA study
- Takagi-Sugeno type rules from ANFIS
- Load distributions from ANFIS membership functions

By having such information available from this work, forecasting rates and methods can become standardised for substations with similar drivers.

7.4 Conclusion

The main research question addressed and closed in this chapter is RQ5. That is, an evaluation of CL-ANFIS is carried out analysing clusters, effect of training data and comparison to NPG practices. The main conclusions from this evaluation exercise are that the CL-ANFIS methodology is robust within self comparisons but against the current MD forecasting methodology conducted by NPG, it does not necessarily offer an enhanced numerical forecast. This is due to the many post processing techniques engineers at NPG use after SCADA information is gathered, most of which is human engineering judgment. Approaching this problem purely from a statistical or machine learning perspective will not offer an enhanced numerical forecast in some cases because of this.

However, it does offer a comparable forecast in many instances within the allocated 10% MD change NPG allows for which is good for those particular substations where no human intervention is needed. The other enhancement which NPG does not currently use is the customer information from CL-ANFIS. This gives qualitative information based on the daily load profiles of the individual cluster groups and the customer information sourced by NPG.

Chapter 8

Conclusions

The goal of this research was to produce a robust forecasting methodology which would provide a greater understanding of substation load profiles. This has been addressed in two main ways through the proposed CL-ANFIS. One is through a more robust forecasting method than the current implementation at NPG and the other is a deeper understanding of substation load profiles based on demographic information.

8.1 Addressing Research Questions

The research aim of this thesis was to create a forecasting methodology which would also provide a greater insight of substation load profiles through demand or external data sets. This was primarily addressed through the research questions:

RQ1: How can statistical and machine learning methods be used to analyse internal factors such as demand variation?

RQ2: How can statistical and machine learning methods be used to analyse external factors such as temperature and customer type?

RQ3: How can more transparency be gained about current and future trends after analysing internal and external factors?

RQ4: How can the forecast and insights gained from statistical analysis be used to enhance the business as usual for NPG?

RQ5: How can an evaluation of a new method be conducted, given the multiple influencing factors which affect the load?

8.1.1 RQ1: Internal Factors

In order to complete a comparison between the CL-ANFIS and the NPG forecasting method, the demand data had to undergo several pre-processing methods. The main internal demand factors addressed here include the following:

1. Determination of Maximum Demand Values

2. Handling outlying loads

3. Handling load transfers

A simple but comparable MD method was created that mimics the MD selection method of NPG but in a fully automated way. In Chapter 5, it was shown that this method provided similar results to the NPG MD values as most cases were in a ±5% band of the NPG MD. This was later useful when a comparison between CL-ANFIS and the NPG method was made in Chapter 7.

By using a combination of an existing outlier detection method using wavelets and some engineering judgment, a method was created to find and replace outlying values within load profiles. This is of key importance when creating a forecast but it is also important in removing some of the human bias in a maximum demand selection method.

In addition, there was a discussion of load transfers in Chapter 5 which are not necessarily outliers but can have an unwanted effect on a forecast. Some preliminary work was done which was not used in the CL-ANFIS method but could be carried forward to develop governance on how load transfers should be defined.

8.1.2 RQ2: External Factors

Three supplementary data sets were presented in this thesis:

1. Met Office Integrated Data Archive System Temperature data

2. Office of National Statistics' Output Area Classifications

3. NPG Substation Demographic Information

Relating any external data sets to a substation was reliant upon matching a postcode at a particular substation to the nearest matching postcode in the external data set. This proved problematic for both the MIDAS in Section 4.1 and OAC data sets in Section 4.2 as the related classifications could not be made specific enough. While there was a negative correlation between temperature and demand data, the relationship was not explicit enough on an annual level to warrant usage.

The OAC data offered no relationships between the chosen demand metrics and the OAC groups. Attempting to use popular machine learning classification methods showed there was no overall good performance in classifying substations correctly. This showed some clear discrepancies in relating different catchment zones of postcodes between the NPG substations and the public OAC data. That is, the customers that a substation serves and the customers that are captured in an OAC group can be extremely different even when the postcodes are in approximately the same area.

The NPG demographic information in Section 4.3 proved to be the most useful as it did not suffer from either missing data or difficulty in relating substation postcodes to the correct area. In the case where aggregated demand is used, this showed that having a tailor-made database of customer information is extremely useful. This data set helped form the basis for the CL-ANFIS method.

8.1.3 RQ3: Transparency of Information

By transparency, a qualitative description that can aid NPG practitioners is needed. This is covered in detail in Sections 6.6 and 6.7. That is, using a combination of the customer descriptions in Section 4.3 and the membership functions within CL-ANFIS a ruleset about substation clusters can be created. Furthermore, the validity of the qualitative descriptors in Table 6.9 is evaluated in Section 6.7. It was found that using the more descriptive labels of Domestic or Commercial can sometimes be preferable to a more granular description that includes size or a urban/rural description. Ultimately it was shown that NPG has three main areas from which to draw more insight using the CL-ANFIS method which are:

- Takagi-Sugeno type rules from ANFIS directly
- Cluster rule types extracted from bell-shaped membership functions
- Qualitative demographic labels by cluster

These give various forms of rules from both the pure algorithm perspective as well as interpreted rules using the customer information provided by NPG. Takagi-Sugeno type rules may not have immediate impact on NPG practices but do give some insight into the behaviour of the ANFIS method on its own. However, the demographic labels and cluster rules from the membership functions provide NPG practitioners with practical and easy to understand qualitative descriptions.

8.1.4 RQ4: Enhancing Business as Usual for NPG

Enhancing the business as usual for NPG comes in a variety of methods covered by the previous research questions. Namely, the new methods that enhance business as usual include:

- Maximum Demand Determination (Section 5.1.1)
- Outlier Replacement (Section 5.2.1)
- Load Transfer Shifting (Section 5.2.2)
- Clustering substations into groups in CL-ANFIS (Section 6.2.1)
- Forecasting using CL-ANFIS (Section 6.3)

Each of these methods can immediately enhance the forecasting and governance structure in the current NPG method. The maximum demand determination allows for automated MD selection with comparable results to reported DLE MDs. Outlier detection and replacement can be used to aid in both the forecasting process and the determination of maximum demand. Load transfer shifting allows for an automated approach to both detecting and adjusting for load transfers. The clustering approach in CL-ANFIS allows substations with similar daily load profiles and customer demographics to be grouped together. This aids in both the forecasting process and for qualitative explanations. The forecasting with CL-ANFIS provides automated forecasts based on annual profiles rather than just the MD based approach of MD. This allows for a more robust statistical approach that improves with more input data.

8.1.5 RQ5: Evaluation of CL-ANFIS Method

Evaluation of the CL-ANFIS method was carried out in Section 6.4 by traditional metrics including MAE and RMSE. The results showed an improvement over an unclustered solution and more naïve forecasting approach for most clusters. Further evaluations of the qualitative aspect was carried out in Section 6.7 using a variety of supervised learning techniques and misclassification errors. Evaluation continued in Chapter 7 considering the effectiveness of the clustering solution and the effect of varying amounts of input data on the CL-ANFIS method. A final comparison against the NPG method was also made in Chapter 7 but due to much of their method occurring manually, it was difficult to make a fair and objective comparison. Some clusters performed comparably to the NPG method in a completely automated method and it is expected that errors would decrease even more significantly if manual adjustments were made in the same way as the current NPG method.

8.2 Recommendations for Future Works

There are three main areas that will be considered for future works which include:

1. How the current method could be adjusted
2. Consideration of future data
3. Alternative methodologies

8.2.1 Adjustment of the Current Method

Within CL-ANFIS, there are many variables which can be adjusted in terms of the parameters of the learning method as well as the form of the input data. The main parameter tuning includes the following:

- Number of principal components to consider for customer information (Section 4.3)
- Kurtosis value of outlier indication for maximum demand method (Section 5.1)
- Difference between extrema in load transfer method (Section 5.1)
- Time window for median outlier replacement (Section 5.2)
- Number of K clusters used for grouping daily load profile clusters (Section 6.2)
- Number of historical load values to consider for CL-ANFIS (Section 6.3)
- Time window of historical load values to consider for CL-ANFIS (Section 6.3)
- Number of years of electrical demand for training CL-ANFIS (Section 6.3)

8.2.2 Consideration of Future Data

In the future, a greater breadth of data is expected due to the rollout of smart grid technology. One of the limiting factors of this research was the analysis of demand data at an aggregated substation level. If all individual customer data was available through the smart grid technology, then a more detailed customer analysis could be completed. Also, the long-term forecast proved to be more inaccurate further into the future. This was mostly due to the availability of demand data used to build the forecast at the time. As more historical demand data becomes available, the more accurate a long-term forecasting method will become.

8.2.3 Alternative methodologies

The aim of using the CL-ANFIS method was to simultaneously produce a forecast whilst allowing for the most flexibility in terms of the implementation as well as the interpretation. In terms of the implementation, traditional time series methods such as ARIMA (Hong et al., 2014, 2016) could potentially be used or other forms of a neural network structure (Chen et al., 2004; Hippert et al., 2001; Park et al., 1991). Similarly, for the unsupervised learning, the choice of using daily demand profiles to group substations is based on the current state of research in the area. Clustered groups could be created in other ways such as annual demand of substations or location based clustering. Ultimately, as more data becomes available and as machine learning techniques become more sophisticated, the CL-ANFIS method presented here can be enhanced or altered as needed.

Bibliography

B. Akperi and P. Matthews. Analysis of Clustering Techniques on Load Profiles for Electrical Distribution. *2014 IEEE International Conference on Power System Technology (POWERCON)*, 2014a.

B. Akperi and P. Matthews. Analysis of Customer Profiles on an Electrical Distribution Network. *2014 49th International Universities Power Engineering Conference (UPEC)*, 2014b.

M. H. Albadi and E. F. El-Saadany. A summary of demand response in electricity markets. *Electric Power Systems Research*, 78(11), 2008.

M. R. Amin-Naseri and A. R. Souroush. Combined Use of Unsupervised and Supervised Learning for Daily Peak Load Forecasting. *Energy Conversion and Management*, 49(6), 2008.

N. Amjady. Short-Term Hourly Load Forecasting Using Time-Series Modeling with Peak Load Estimation Capability. *IEEE Transactions on Power Systems*, 16(3), 2001.

R. K. Anderson. *Visual Data Mining: The Visminer Approach.* Wiley, 2013.

Andrew Ng. CS229 Lecture Notes. `http://cs229.stanford.edu/notes/cs229-notes3.pdf`, 2017.

N. O. Andrews and E. A. Fox. Recent developments in document clustering. Technical report, Technical report, Computer Science, Virginia Tech, 2007.

C. Basu, H. Hirsh, and W. Cohen. Recommedation as Classification: Using Social and Content-Based Information in Recommendation. 1998.

V. Bianco, O. Manca, S. Nardini, and A. A. Minea. Analysis and forecasting of nonresidential electricity consumption in Romania. *Applied Energy*, 87(11), 2010.

C. Bilen and S. Huzurbazar. Wavelet-Based Detection of Outliers in Time Series. *Journal of Computational and Graphical Statistics*, 11(2):311–327, 2002.

S. Bond and F. Insalaco. Area Classification of Super Output Areas and Datazones, 2007.

T. Brijs, G. Swinnen, K. Vanhoof, and G. Wets. Using association rules for product assortment decisions: a case study. *KDD '99 Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 254–260, 1999.

D. W. Bunn. Forecasting Loads and Prices in Competitive power markets. *Proceedings of the IEEE*, 88(2), 2000.

Cambridge Econometrics. Cambridge Econometrics. `https://www.camecon.com/`, 2017.

CE Electric UK. Code of Practice for Production of Load Estimates, 2005.

CE Electric UK. Review and alignment of the Load Estimates production processes within NEDL & YEDL, 2009.

J. R. Chang, L. Y. Wei, and C. H. Cheng. A hybrid anfis model based on ar and volatility for taiex forecasting. *Applied Soft Computing*, 11(1):1388–1395, 2011.

B. Chen, P. Matthews, and P. Tavner. Wind Turbine Pitch Faults Prognosis Using A-priori Knowledge-Based ANFIS. *Expert Systems with Applications*, 40(17), 2014.

B. Chen, P. Matthews, and P. Tavner. Automated On-line Fault Prognosis for Wind Turbine Pitch Systems Using Supervisory Control and Data Acquisition. *IET Renewable Power Generation*, 9(5), 2015.

B. J. Chen, M. W. Chang, and C. J. Lin. Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001. *IEEE Transactions on Power Systems*, 19(4), 2004.

G. Chicco, R. Napoli, F. Pigilone, P. Postolache, M. Scutarui, and C. Toader. Emergent electricity customer classification. *IEE Proceedings - Generation, Transmission and Distribution*, 152(2), 2005.

G. Chicco, R. Napoli, and F. Pigilione. Comparisons Among Clustering Techniques for Electricity Customer Classification. *IEEE Transactions on Power Systems*, 21 (2), 2006.

W. W. Cohen. Learning Rules that Classify E-Mail. 1996.

A. M. Creighton. Load Related Drivers & Interventions. *CE Electric UK*, 2011.

C. Creighton and S. Hanash. Mining gene expression databases for association rules. *Bioinformatics*, 19(1):79–86, 2003.

J. C. Cuaresma, J. Hlouskova, S. Kossmeier, and M. Obersteiner. Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy*, 77 (1):87–106, 2004.

D. L. Davies and D. W. Bouldin. A clustering separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1979.

Department of Energy and Climate Change. 2050 pathways analysis. `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/42562/216-2050-pathways-analysis-report.pdf`, 2010.

Department of Energy and Climate Change. 2050 Pathways. `https://www.gov.uk/2050-pathways-analysis`, 2013.

D. L. Donoho and I. M. Johnstone. Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, 81(3):425–455, 1994.

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification.* Wiley-Interscience, New York, NY, 2000.

J.C. Dunn. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3), 1974.

EA Technology. About EA Technology. `https://www.eatechnology.com/about-us/about-ea-technology/`, 2017.

EDW Technology. Electricity System. `http://www.edwt.org/images/content/Electricity_System.png`, 2017.

F. Elkarmi. Load research in electric power system planning, operation, and control - the case of Jordan. *Energy Policy*, 36, 2008.

Energy Networks Association. Electricity Distribution Map. `http://www.energynetworks.org/info/faqs/electricity-distribution-map.html`, 2014.

B. S. Everitt, S. Landau, and M. Leese. *Cluster Analysis*. Arnold, London, 2001.

V. Figueiredo, F. Rodrigues, Z. Vale, and J. Borges Gouveia. An Electric Energy Consumer Characterization Framework Based on Data Mining Techniques. *IEEE Transactions on Power Systems*, 20(2), May 2005.

A. Fink. *Conducting Research Literature Reviews*. Sage Publications, 2005.

M.A. Friedl and C.E. Brodley. Decision Tree Classification of Land Cover from Remotely Sensed Data. *Remote Sensing of Environment*, 61(3):309–409, 1997.

J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.

P. Gaillard, Y. Goude, and R. Nedellec. Semi-parametric models and robust aggregation for gefcom2014 probabilistic electric load and electricity price forecasting. *International Journal of Forecasting*, 2015.

Gas and Electricity Markets Authority. ELECTRICITY ACT 1989 - Standard conditions of the Electricity Distribution Licence Econometrics. `https://epr.ofgem.gov.uk//Content/Documents/Electricity%`

`20Distribution%20Consolidated%20Standard%20Licence%20Conditions%20-%20Current%20Version.pdf`, 2017.

A. Goia, C. May, and G. Fusai. Functional Clustering and Linear Regression for Peak Load Forecasting. *International Journal of Forecasting*, 26(4), 2010.

M. A. Hall, I. H. Witten, and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 3rd edition, 2011.

J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 3rd edition, 2011.

T. Hastie and R. Tibshirani. Cross-validation and bootstrap. `http://statweb.stanford.edu/~tibs/sta306bfiles/cvwrong.pdf`, 2009.

E. Hawkins and R. Sutton. The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, 90, 2009.

H. S. Hippert, C. E. Pedreira, and R. C. Souza. Neural Networks for Short-Term Load Forecasting. A Review and Evaluation. *IEEE Transactions on Power Systems*, 16(1), 2001.

K. L. Ho, Y. Y. Hsu, C. F. Chen, and T. E. Lee. Short term load forecasting of taiwan power system using a knowledge-based expert system. *IEEE Transactions on Power Systems*, 5, 1990.

T. Hong, J. Wilson, and J. Xie. Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid*, 5(1), 2014.

T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, 32(3), 2016.

R. J. Hyndman and G. Athanasoupoulos. *Forecasting: principles and practice*. 2013.

R. J. Hyndman and S. Fan. Density Forecasting for Long-Term Peak Electricity Demand. *IEEE Transactions on Power Systems*, 25(2), 2010.

IEEE. IEEE Trial-Use Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions, 2000.

Ž. Ivezić, A.J. Connolly, J.T. Vanderplas, and A. Gray. *Statistics, Data Mining and Machine Learning in Astronomy*. Princeton University Press, 2014.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014.

R. Jammazi and C. Aloui. Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Economics*, 34(3): 828–841, 2012.

J. S. R. Jang. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems, Man, and Cybernetics*, 1993.

J. S. R. Jang. Input Selection for ANFIS Learning. *Proceedings of the IEEE Conference on Fuzzy Systems*, 1996.

I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.

P. S. Kalekar. Time series Forecasting using Holt-Winters Exponential Smoothing. Technical report, Kanwal Rekhi School of Information Tech., IIT Bombay, Powai, Mumbai, India, 2004.

W. Kanitpanyacharoean and S. Premrudeepreeechacharn. Power Quality Classification Using Wavelet Transformation and Artificial Neural Networks. *IEEE PES Power Systems Conference and Exposition*, 2004.

S. Kaplan. *Risk management: Expanding horizons in nuclear power and other industries*, chapter Risk Assessment and Risk Management - Basic Concepts and Terminology, pages 11–28. Hemisphere Publishing Corporation, 1990.

S. Kooths, T. Mitze, and E. Ringhut. Forecasting the emu Inflation Rate: Linear Econometric vs. Non-Linear Computational Models Using Genetic Neural Fuzzy

Systems. *Applications of Artificial Intelligence in Finance and Economics*, 19, 2004.

L. Li, C. R. Weinberg, T.A. Darden, and L. G. Pedersen. Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the ga/knn method. *Bioinformatics*, 17(12):1131–1142, 2001.

M. López, J. Ramírez, J. M. Górriz, D. Salas-Gonzalez, I. Álvarez, F. Segovia, and C.G. Puntonet. Automatic tool for Alzheimer's disease diagnosis using PCA and Bayesian classification rules. *Electronic Letters*, 45(8), 2009.

J. MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 1967.

F. Masulli and A. Schenone. A fuzzy clustering based segmentation system as support to diagnosis in medical imaging. *Artificial intelligence in medicine*, 16 (2):129–147, 1999.

Mathworks. Kurtosis. `http://www.mathworks.co.uk/help/stats/kurtosis.html`, 2013a.

Mathworks. Filtering and Smoothing Data. `http://www.mathworks.co.uk/help/curvefit/smoothing-data.html`, 2013b.

Mathworks. What Is the Genetic Algorithm? `http://uk.mathworks.com/help/gads/what-is-the-genetic-algorithm.html`, 2017.

P. E. McSharry, S. Bouwman, and G. Bloemhof. Probabilistic Forecasts of the Magnitude and Timing of Peak Electricity Demand. *IEEE Transactions on Power Systems*, 20(2), 2005.

P.H. Menold, R.K. Pearson, and F. Allgöwer. Online outlier detection and removal. *Proceedings of the 7th Mediterranean Conference on Control and Automation (MED99)*, 1999.

Z. Mohamed and P. Bodger. Forecasting Electricity Consumption in New Zealand Using Economic and Demographic Variables. *Energy*, 30(10), 2005.

National Grid. UK Future Energy Scenarios. `http://www2.nationalgrid.com/UK/Industry-information/Future-of-Energy/Future-Energy-Scenarios/`, 2013.

National Grid. Data explorer. `http://www2.nationalgrid.com/UK/Industry-information/Electricity-transmission-operational-data/Data-explorer/`, 2014.

National Grid Electricity Transmission plc. The grid code. `www2.nationalgrid.com/WorkArea/DownloadAsset.aspx?id=8589935310`, 2017.

NCAS British Atmospheric Data Centre. Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current), 2013.

Northern Powergrid. Long-term Development Statement. `https://www.northernpowergrid.com/long-term-development-statement`, 2015.

Northern Powergrid. Customer-Led Network Revolution. `http://www.networkrevolution.co.uk/`, 2016.

Office for National Statistics. Methodology Note for the 2011 Area Classification for Output Areas, 2015.

R. T. Ogden. *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhäuser Boston, 1997.

OSIsoft. PI System. `http://www.osisoft.com/pi-system/`, 2017.

D. C. Park, M. A. El-Sharkawi, R. J. Marks II, L. E. Atlas, and M. J. Damborg. Electric Load Forecasting Using An Artificial Neural Network. *IEEE Transactions on Power Systems*, 6(2), 1991.

B. Pitt and D. S. Kirschen. Application of Data Mining Techniques to Load Profiling. *Proceedings of the 21st 1999 IEEE International Conference*, pages 131–136, 1999.

D. M. W. Powers. Evaluation: from precision, recall and F-measure to ROC. *Journal of Machine Learning Technologies*, 2(1), 2007.

H. Quan, D. Srinivasan, and A. Khosravi. Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE transactions on neural networks and learning systems*, 25(2):303–315, 2014.

T. Räsänen, J. Ruuskanen, and M. Kolehmainen. Reducing energy consumption by using self-organising maps to create more personalized electricity use information. *Applied Energy*, 85, 2008.

T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen. Data-based method for creating electricity use load profiles using large amount of cusomter-specific hourly measured electricity use data. *Applied Energy*, 87(11), 2010.

J. D. M. Rennie, L. Smith, J. Teevan, and D. R. Karger. Tackling the Poor Assumptions of Naive Bayes Text Classifiers. *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, 2003.

P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

scikit-learn developers. Naive Bayes. `http://scikit-learn.org/stable/modules/naive_bayes.html`, 2016.

J. Scott. Electricity Distribution Network Planning - Engineering Recommendation P2/6. `https://www.ofgem.gov.uk/sites/default/files/docs/2007/08/p2_6-review-open-letter---1-august.pdf`, 2007.

A. Seppälä. *Load research and load estimation in electricity distribution*. PhD thesis, Technical Research Centre of Finland, 1996.

R. H. Shumway and D. S. Stoffer. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2010.

S. Sumathi and S. Paneerselvam. *Computational Intelligence Paradigms: Theory and Applications Using MATLAB*. CRC Press, Boca Raton, FL, 2010.

T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 1985.

J. W. Taylor. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operation Research*, 204(1), 2010.

J.W. Taylor and P.E. McSharry. Short-Term Load Forecasting Methods: An Evaluation Based on European Data. *IEEE Transactions on Power Systems*, 22(4), 2007.

J. L. Torres, A. Garcia, M. De Blas, and A. De Francisco. Forecast of hourly average wind speed with arma models in navarre (spain). *Solar Energy*, 79(1):65–77, 2005.

A. Vlahou, J. O. Schorge, B. W. Gregory, and R. L. Coleman. Diagnosis of Ovarian Cancer Using Decision Tree Classification of Mass Spectral Data. *Journal of Biomedicine and Biotechnology*, 2003(5):308–314, 2003.

W. J. Wang. New similarity measures on fuzzy sets and on elements. *Fuzzy Sets and Systems*, 85(3), 1997.

S. Weisberg. *Applied Linear Regression*. Wiley, 3rd edition, 2005.

Western Power Distribution. Demonstration of LV Network Templates through statistical analysis. `http://www.westernpowerinnovation.co.uk/Documents/LV-Network-Templates-Report-final.aspx`, 2013.

C. J. Willmott and K. Matsuura. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 2005.

C. J. Willmott, S. G. Ackleson, R. E. Davis, J. J. Feddema, K. M. Klink, D. R. Legates, J. O'Donnel, and C. M. Rowe. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research*, 90, 1985.

B. Wilson. The Electricity Safety, Quality and Continuity Regulations 2002. `http://www.legislation.gov.uk/uksi/2002/2665/contents/made`, 2002.

X. L. Xie and G. Beni. A Validity Measure for Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8), 1991.

R. Xu and D. Wunsch. Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, 16(3), 2005.

S. Y. Yerima, S. Sezer, G. McWilliams, and I. Muttik. A New Android Approach Using Bayesian Classification. *2013 IEEE 27th International Conference on Advanced Information Networking and Applications*, 2013.

K.Y. Yeung and W.L. Ruzzo. Principal component analysis for clustering gene expression data. *Bioinformatics*, 17(9):763–774, 2001.

Z. Yu, F. Haghighat, B. C.M. Fung, and H. Yoshino. A decision tree method for building energy demand modeling. *Energy and Buildings*, 42(10):1637–1646, 2010.

Z. Yun, Z. Quan, S. Caixin, L. Shaolan, L. Yuming, and S. Yang. Rbf neural network and anfis-based short-term load forecasting approach in real-time price environment. *IEEE Transactions on Power Systems*, 23(3):853–858, 2008.

D. Zhang and L. Zhou. Discovering Golden Nuggets: Data Mining in Financial Application. *IEEE Transactions on Systems, Man, and Cybernetics*, 34, 2004.

G. P. Zhang. Time Series Forecasting Using A Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 50, 2003.

H. Zhang. The optimality of naive bayes. *AA*, 1(2):3, 2004.

H. Zhang, A.C. Berg, M. Maire, and J. Malik. Svm-knn: Discriminative nearest neighbor classification for visual category recognition. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2126–2136. IEEE, 2006.

Appendix A

Cluster Load Profiles

Table A.1: Reordering of K-Means Clusters

Original Cluster	Cluster in Thesis
1	1
2	3
3	7
4	8
5	2
6	5
7	9
8	4
9	6

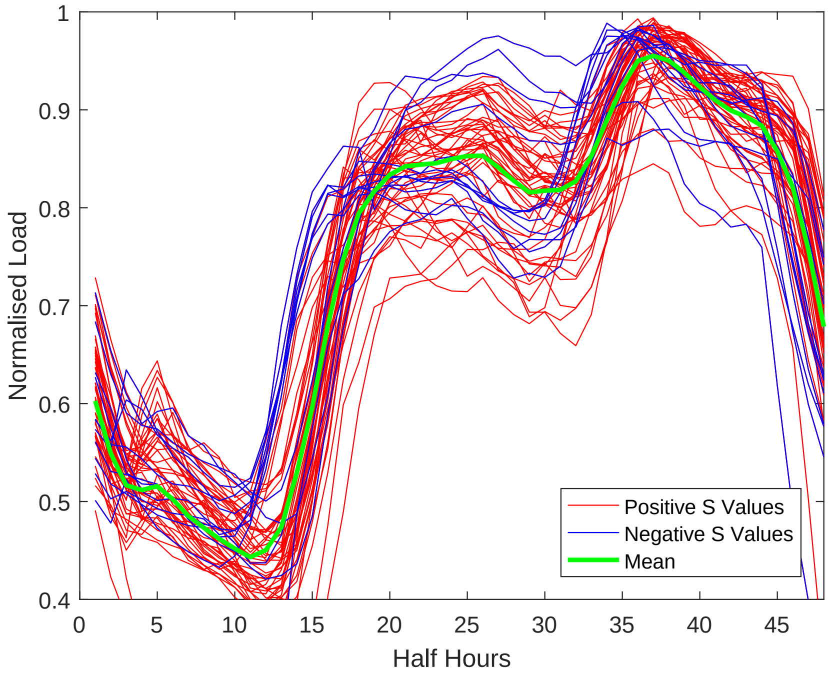


Figure A.1: Cluster 1 Substation Daily Load Profiles with Postive and Negative Silhouette Values Highlighted

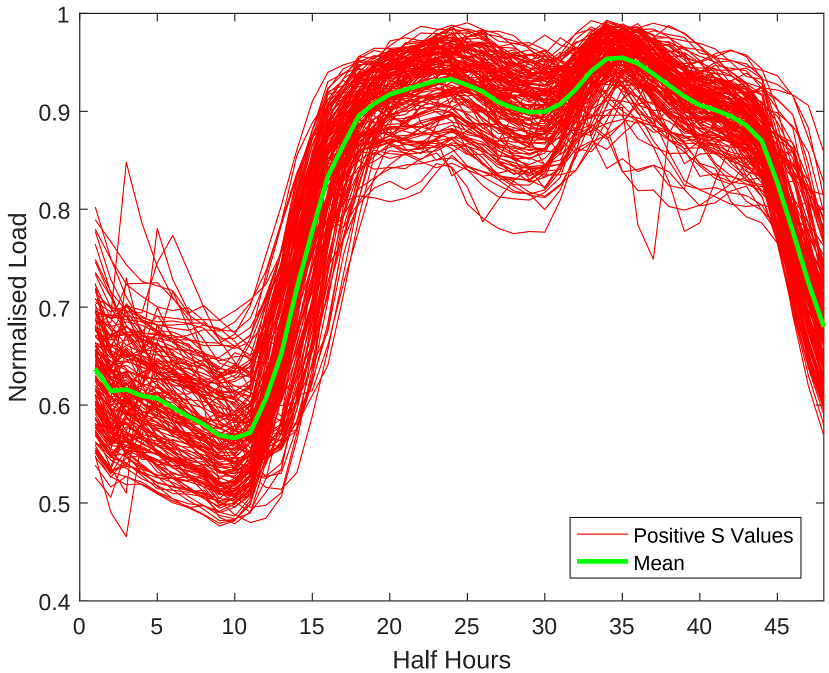


Figure A.2: Cluster 2 Substation Daily Load Profiles with Postive and Negative Silhouette Values Highlighted

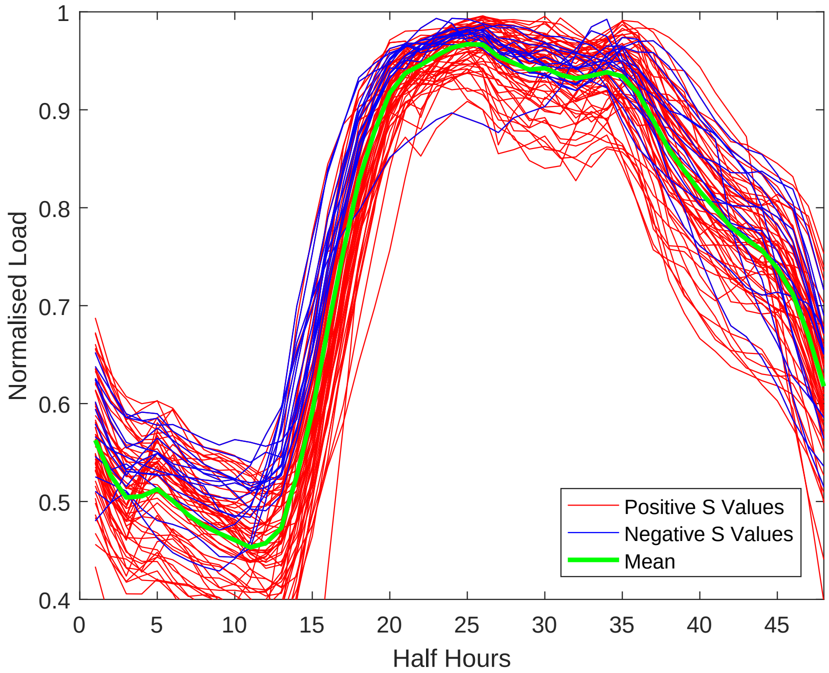


Figure A.3: Cluster 3 Substation Daily Load Profiles with Postive and Negative Silhouette Values Highlighted

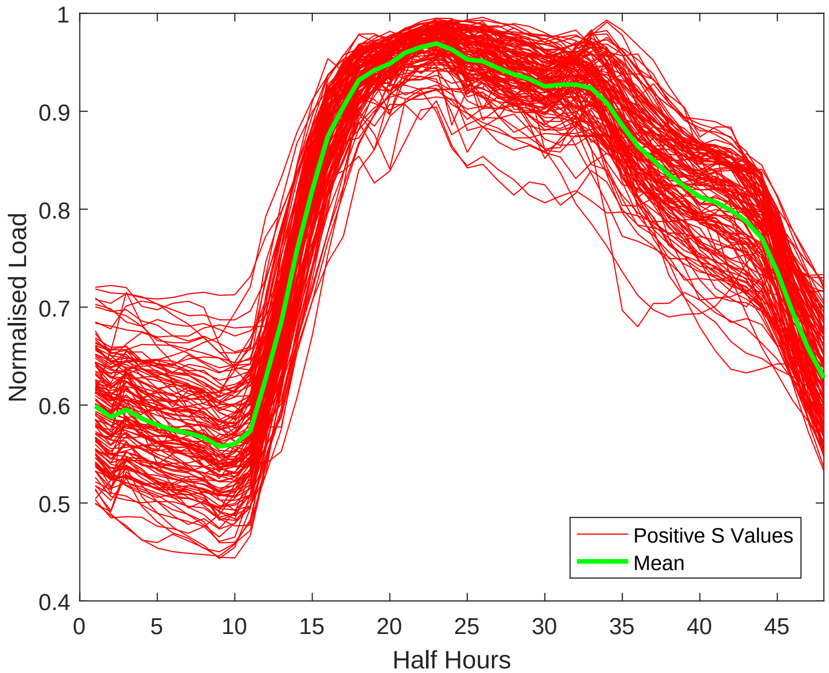


Figure A.4: Cluster 4 Substation Daily Load Profiles with Postive and Negative Silhouette Values Highlighted

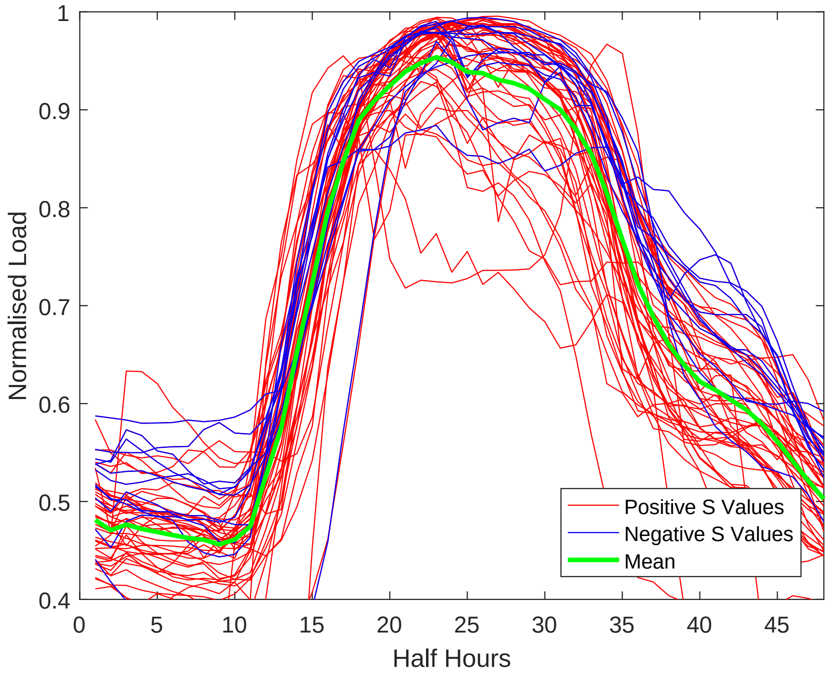


Figure A.5: Cluster 5 Substation Daily Load Profiles with Postive and Negative Silhouette Values Highlighted

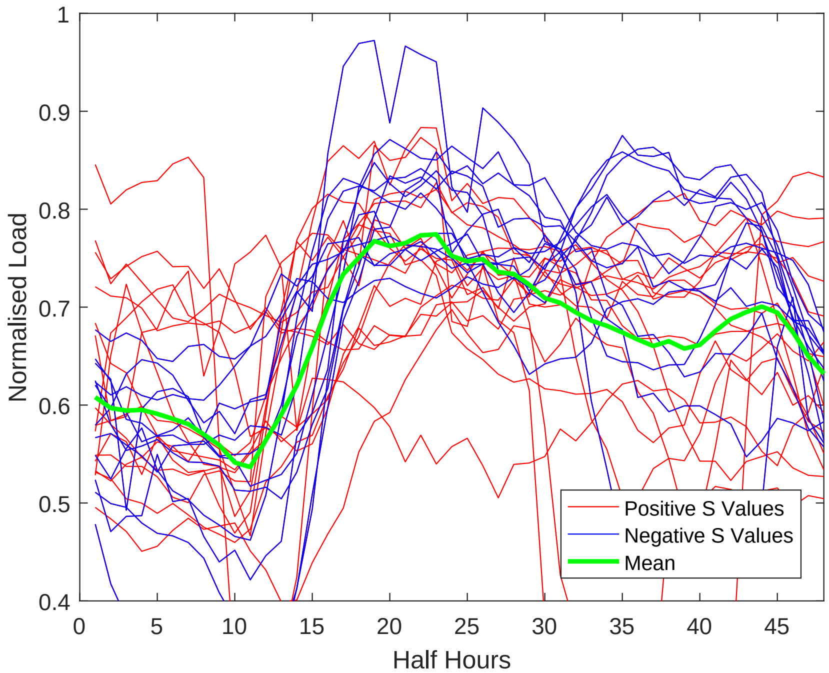


Figure A.6: Cluster 6 Substation Daily Load Profiles with Postive and Negative Silhouette Values Highlighted

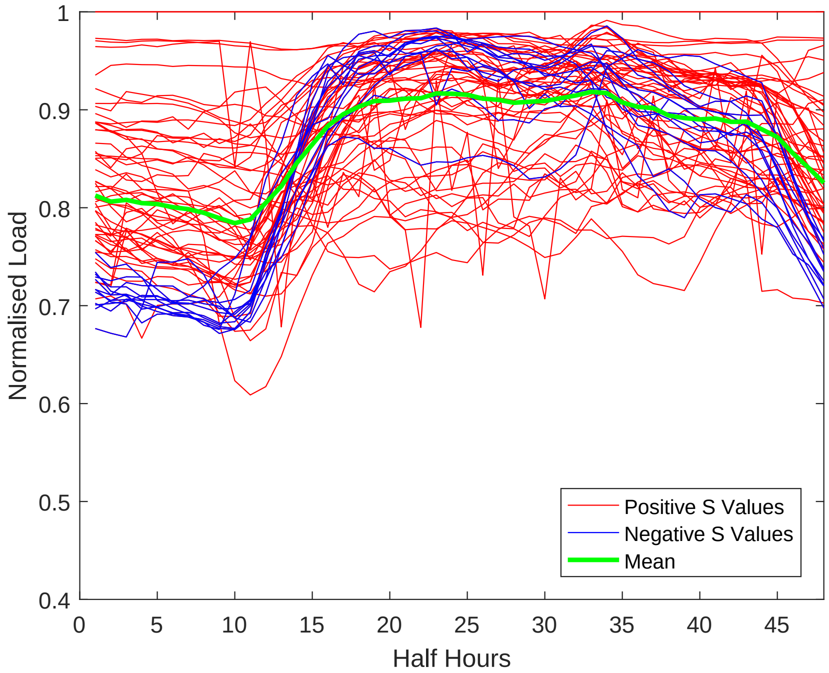


Figure A.7: Cluster 7 Substation Daily Load Profiles with Postive and Negative Silhouette Values Highlighted

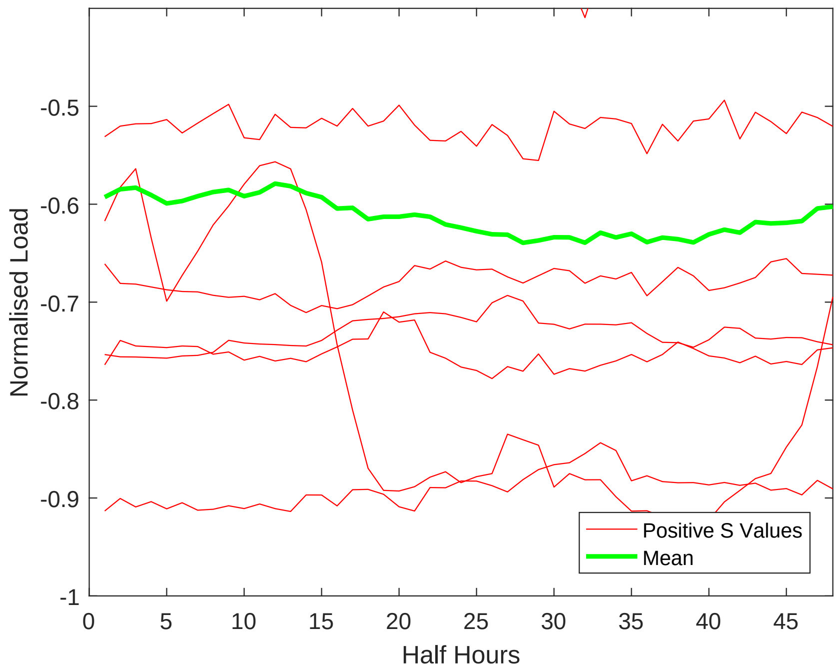


Figure A.8: Cluster 8 Substation Daily Load Profiles with Postive and Negative Silhouette Values Highlighted

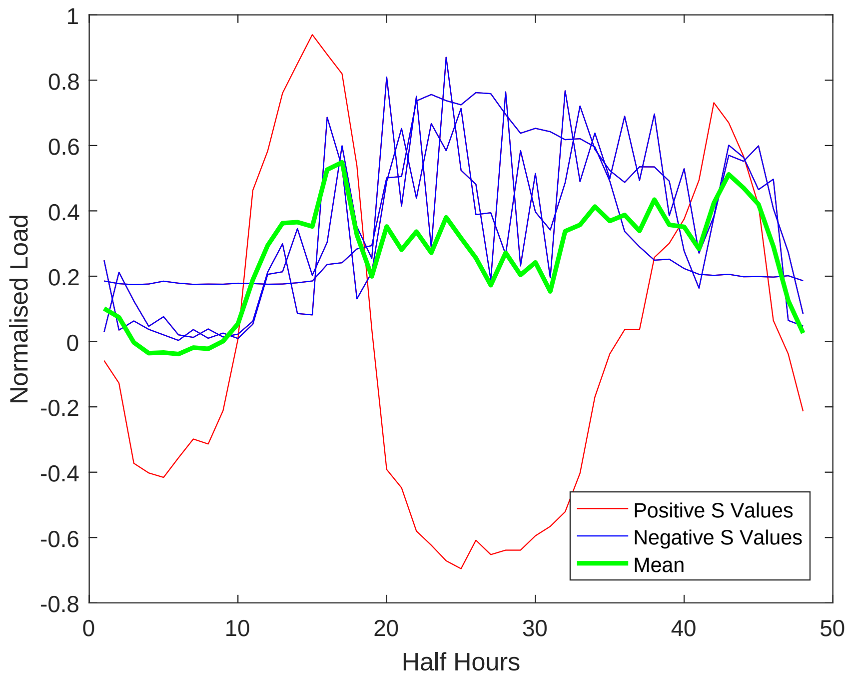


Figure A.9: Cluster 9 Substation Daily Load Profiles with Postive and Negative Silhouette Values Highlighted