Integrated Load and State Estimation Using Domestic Smart Meter Measurements

Ali Al-Wakeel

School of Engineering

Cardiff University

A thesis submitted for the degree of

Doctor of Philosophy

April 2016

Table of Contents

Abstract

The UK Government is promoting the decarbonisation of the power sector. The electrification of transport and heating, installation of distributed generators, development of smart grids and creation of an electricity and gas smart metering system are in progress.

Higher penetrations of distributed generation and low carbon loads may lead to operational difficulties in distribution networks. Therefore, increased real-time monitoring and control becomes a necessary requirement. Distribution network operators will have available to them smart meter measurements to facilitate safe and cost-effective operation of distribution networks. This thesis investigates the application of smart meter measurements to extend the observability of distribution networks.

Three main aspects were covered in this work:

1. The development of a cluster analysis algorithm to extract consumption patterns from smart meter measurements. The k-means based cluster analysis algorithm was demonstrated on measurements that were obtained from the Irish Smart Metering Trials. Daily and segmented load profiles of individual and aggregated domestic smart meters were clustered.

2. The development of a load estimation method to estimate missing and future measurements of smart meters. The load estimation algorithm uses the outputs of the clustering algorithm; and investigates the application of different distance functions to estimate any lost measurements. Different durations of lost measurements were simulated to assess the accuracy of the estimated load.

3. The development of an integrated load and state estimation algorithm to extend the observability of distribution networks. The developed load estimator provides pseudo measurements to an Iteratively Re-Weighted Least Squares (IRWLS) state estimator. The capability of the IRWLS state estimator was investigated using measurements from a low voltage microgrid. The IEEE 33 bus medium voltage distribution network was used to assess the performance of the integrated load and state estimation algorithm.

The developed load and state estimator was capable of estimating the voltage magnitude and phase angle, at each busbar of the distribution network, with high accuracy. For one hour of missing measurements, the Mean Absolute Percentage Error (MAPE) of the estimated voltage magnitude was less than 0.03%. For 24 hours of missing measurements, the estimated voltage magnitudes had a MAPE that was less than 0.5%.

Dedication

For the mother who sowed the seeds of knowledge and research since I was 6 or 7;

For the brother who kept standing by my side and taking the responsibility from me;

For the father who asked me to be patient till the end;

For the uncle who supported me and told me to overcome my fears …

For Susan, Nawaf, Sa’ad and Tahir …

Acknowledgement

I would like to express my sincere gratitude and appreciation to my supervisors, Professor Jianzhong Wu and Professor Nick Jenkins, for their never-ending guidance, support and patience. I'm extremely honoured to have this research conducted under their supervision.

I'm indebted to Dr Lee Thomas for the help he provided and keeps providing, not to mention the significant amount of time he spent discussing and purifying my ideas, as well as reviewing this thesis.

I'm also thankful to Catherine Roderick for her support proofreading the thesis, and the time she spent helping me to improve my technical writing in English.

I will always be grateful to Mrs Chris Lee and Mrs Jeanette Whyte for their motherly care and kindness. I will never forget the assistance, guidance and patience they showed throughout my four years in Cardiff.

Special thanks to Babis Marmaras and Kriton Xydas for the productive discussions, for their remarks and feedback.

Finally, thanks to my family members and friends for their support.

List of Figures

List of Tables

List of symbols

List of symbols of Chapter 3

$\mathbf{cc}_j$	j^{th} cluster centre
i	Index of training load profile
j	Cluster index
J	Objective function
k	Number of clusters
l	Index of test load profile
$\mathbf{lp}_{\text{agg. daily}}$	aggregated daily load profile
$\mathbf{lp}_i$	i^{th} load profile
$\text{lp}_i(t)$	the measurement of the i^{th} load profile at the t^{th} half hour
L	Number of test load profiles
m	Number of profiles per cluster
n	Number of training load profiles
r	segmentation time window
rnd_m	randomisation number
t	the half-hour index
T	half-hourly measurements per a load profile
x	Percentage of the mean active power consumption during the training period

List of symbols of Chapter 4

$\mathbf{cc}_j$	j^{th} cluster centre
i	Index of test load profile
j	index of training cluster
k	Index of test load profile
$\mathrm{lp}_{act}(t)$	actual half-hourly measurement of an individual smart meter or the aggregated load
$\mathrm{lp}_{est}(t)$	estimated half-hourly measurement that was obtained using the k-means based load estimation algorithm
$\mathbf{lp}_i$	i^{th} test load profile
$\mathbf{lp}_k$	k^{th} test load profile
$\mathrm{lp}_i(t)$	the measurement of the i^{th} load profile at the t^{th} half hour
N	Overall number of samples (considered in the brute-force approach)
r	segmentation time window
t	the half-hour index
T	half-hourly measurements per a load profile

List of symbols of Chapter 5

b_{kj}	Susceptance of the feeder connecting busbar k to busbar j
B_{kj}	Sum of the susceptances of the feeders connecting busbar k to busbar j
$\mathbf{e}$	Vector of measurement errors
g_{kj}	Conductance of the feeder connecting busbar k to busbar j
G_{kj}	Sum of the conductances of the feeders connecting busbar k to busbar j
$\mathbf{h}$	Vector of non-linear functions that relate the measurements vector to the system state vector
$\mathrm{h}_i(\mathbf{x})$	Non-linear function that relates the i^{th} measurement to the state vector $\mathbf{x}$
i	Measurement index
j	Busbar index
k	Busbar index
m	Number of measurements
n	Number of system states
P_k	Active power injection at busbar k
P_{kj}	Active power flow in the feeder connecting busbar k to busbar j
Q_k	Reactive power injection at busbar k
Q_{kj}	Reactive power flow in the feeder connecting busbar k to busbar j
r_i	Residual of the i^{th} measurement

t	the half-hour index
T	half-hourly measurements per a load profile
$\mathbf{W}$	Weight matrix
W_{ii}	Weight of the i^{th} measurement
V_k	Voltage magnitude at busbar k
$\mathbf{x}$	System state vector
x_n	n^{th} system state
$\mathrm{x}_{estimated}$	Estimated system state
x_{exact}	Exact system state
$\mathbf{z}$	Measurements' vector
z_i	i^{th} measurement
σ_i^2	Variance of the i^{th} measurement
θ_k	Voltage phase angle at busbar k
θ_{kj}	Difference between voltage phase angle at busbar k and voltage phase angle at busbar j

Abbreviations

AC	Alternating Current
AD	Absolute Difference
AE	Absolute Error
AI	Artificial Intelligence
AMR	Automatic Meter Reading
ANN	Artificial Neural Networks
AvED	Average Euclidean Distance
BESS	Battery Energy Storage System
BT	British Telecommunications
CAD	Consumer Access Device
CBT	Customer Based Trials
CC	Cluster Centre
CGI	Consultant to Government and Industry
CH	Communication Hub
CHP	Combined Heat and Power
DC	Direct Current
DCC	Data Communications Company
DER	Distributed Energy Resources
DER-TF	Distributed Energy Resources - Test Facility
DG	Distributed Generation
DMS	Distribution Management System
DMSC	Distribution Management System Controller
DNO	Distribution Network Operator

EES	Electricity Export Supplier
EIS	Electricity Import Supplier
ENO	Electricity Network Operator
GHG	Greenhouse gas
GHz	Giga Hertz
GIS	Gas Import Supplier
GNO	Gas Network Operator
GPRS	General Packet Radio Service
HV	High Voltage
ICT	Information and Communication Technology
IHD	In-Home Display
IRWLS	Iteratively Re-Weighted Least Square
kVA	kilo Volt Ampere
kVAr	kilo Volt Ampere reactive
kW	kilo Watt
kWh	kilo Watt hour
LP	Load Profile
LV	Low Voltage
MAPE	Mean Absolute Percentage Error
MHz	Mega Hertz
min	minute
ms	milli-second
MV	Medium Voltage
Ofcom	Office of Communications

OU	Other User
PLC	Public Limited Company
PV	Photovoltaic
PVUR	Phase Voltage Unbalance Rate
RF	Radio Frequency
RMS	Root Mean Square
RMSE	Root Mean Square Error
RSA	Registered Supplier Agent
RSE	Ricerca sul Sistema Energetico
SCADA	Supervisory Control And Data Acquisition
sec	second
SHGM	Schweppe Huber Generalised M
SM HAN	Smart Metering Home Area Network
SM WAN	Smart Metering Wide Area Network
SME	Small and Medium Enterprises
SMETS	Smart Metering Equipment Technical Specifications
SMKI	Smart Metering Key Infrastructure
SOM	Self-Organising Maps
SOP	Soft Open Point
TLP	Typical Load Profiles
UKGDS	United Kingdom Generic Distribution System
WLAV	Weighted Least Absolute Value
WLS	Weighted Least Squares

Chapter 1 Introduction

1.1 Background

The UK Government is required to fulfil its commitments towards the Kyoto Protocol and its complementary 2020 Climate and Energy Package. The Kyoto Protocol is an international agreement which commits its parties to internationally binding emission reduction targets. The Protocol came into force in February 2005 [1]. The 2020 package is a set of binding legislations to ensure that the EU meets its climate and energy targets for the year 2020. The UK Government is committed to three targets. These are:

- At least 16% cut in the greenhouse gas (GHG) emissions (from the 2005 levels) [2];
- At least 15% of energy from renewables [2]; and
- At least 20% improvement in energy efficiency [3], [4].

To meet these commitments, the UK Government is promoting the decarbonisation of the power sector. The electrification of heating and transport and the development of Smart Grids are in progress [5].

The Smart Grid uses information and communications technology (ICT) to implement the following principles [6], [7]:

- Self-healing from power disturbance event;
- Enable active participation by consumers in demand response;
- Operate resiliently against physical and cyber attacks;
- Accommodate all generation and storage options;
- Enable new products, services, and markets;
- Optimise the transfer capability of the transmission and distribution networks and meet the demand for increased quality and reliable supply.

Grid integration of distributed energy resources (such as small-scale distributed generators, electric vehicles, heat pumps, demand response, and energy storage) brings

significant uncertainties and, at high penetrations, may lead to operational difficulties in the distribution network [8]. For example, high penetrations of distributed generation can lead to over-voltages at times of light load; the increase in demand resulting from the electrification of heating load and electric vehicles can give rise to local low voltages [5]. Therefore, knowledge of accurate system state information is critical for the network operator to guarantee safe, prompt and cost-effective operation of the network, while making the best use of the assets [8].

A distribution network usually comprises hundreds of busbars and feeders. Real-time monitoring and control of distribution networks is impractical and infeasible due to economic constraints. In a typical distribution system, real-time measurements can be obtained only at the primary substation (33kV level). Almost no real-time monitoring is carried out at the 11kV level or the 0.4kV LV level [9], [10]. A distribution network is therefore described as an under-determined system. This means that the installed real-time measurements are insufficient to make the system fully observable.

In over-determined power networks, e.g. transmission networks, state estimation [11] is used to clean up the errors in a set of redundant raw measurements. However, state estimation is applied in under-determined distribution networks to increase the observability of the network at the lower voltage levels.

A distribution network state estimator applies a minimum set of real-time measurements (such as the measurements of the primary (or if available, the secondary) substation, few real-time measurements) and large numbers of pseudo-measurements to identify the operating state(s) of the network. Pseudo measurements are obtained from load estimates using historical load profiles [1], transformer kVA ratings or near real-time measurements of smart meters [12]. Figure 1.1 shows the major components of state estimation in distribution networks.

[1] A load profile is the half-hourly pattern of electricity consumption during a day.

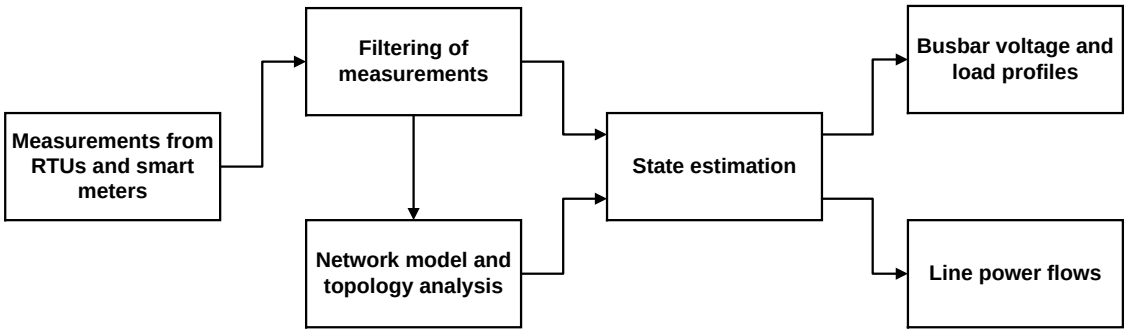


Figure 1.1 Major components of state estimation in distribution networks

In Figure 1.1, state estimation starts with measurements collected from Remote Terminal Units (RTUs) and smart meters. Measurements of smart meters can be either real-time or pseudo measurements derived from them. Firstly, the measurements are filtered by removing inconsistent (e.g., zero and negative) values from them. The remaining measurements are processed so that the configuration and topology of the network are determined. The available data are further processed to obtain the system state defined in terms of the voltage magnitudes and angles at each busbar of the network. The active and reactive demand of each busbar and the line power flows are then calculated based on the estimated voltage magnitudes and angles.

Smart metering is a basic component of Smart Grids; “All of the 2020 targets, the energy efficiency, the renewables, the reduction of CO_2, they all depend on the grid. The foundation of the grid is smart metering [13]”. Smart meters employ advanced metrology, control, data storage and communication technologies to offer a range of functions. Nearly 53 million gas and electricity smart meters will be deployed in all domestic and small non-domestic premises in the UK by the end of 2020 [14]–[16]. Customers will have near real-time (every 10 seconds for electricity smart meters and 30 minutes for gas smart meters [17], [18]) information on their energy consumption to help them control and manage their energy use, save money and reduce GHG emissions [19].

Smart meters and their associated ICT infrastructure can improve the observability of distribution networks. However, the communication systems are prone to time delays and temporary failures [12]. These problems will lead to erroneous or even missing real-time measurements. Imprecise and lost measurements will degrade the performance or even disable the state estimator. It is therefore required to develop a load estimation algorithm that estimates lost and future measurements of the smart meters. This research investigates the potential for using smart meter measurements to increase the observability of distribution networks through the development of an integrated load and state estimation algorithm.

1.2 Research Objectives

The availability of measurements from smart meters suggests that distribution networks will evolve gradually from being in an under-determined to over-determined state. However, due to technical problems, smart meter measurements may not be available (for example, to a network operator) in (near) real-time; they will typically be available within 30 seconds of request [20]. As a consequence, the question of

"**How can smart meter measurements extend the observability of a distribution network?**"

was raised. This thesis reports the research approach that has been adopted to answer the aforementioned question. The following objectives were set:

- To develop an algorithm based on the k-means cluster analysis method to group similar consumption patterns and build characteristic load profiles, defined in terms of the cluster centres, using past smart metering data.

- To investigate and validate the performance of the clustering algorithm using measurements of individual and aggregated smart meters.

- To create a load estimation algorithm that estimates missing and future smart meter measurements. The load estimator utilises the clusters centres and investigates the performance of different distance functions to produce its outputs.

- To assess the performance of an iteratively re-weighted least squares (IRWLS) state estimator using measurements collected from a practical test LV microgrid.

- To develop an integrated load and state estimation algorithm. The integrated estimation algorithm links the developed load estimator with the IRWLS estimator.

- To test and validate the integrated load and state estimator using real-time and pseudo measurements derived from smart metering data.

1.3 Thesis Outline

The research described in this thesis was an investigation of the applicability of smart meter measurements to extend the observability of distribution networks. Figure 1.2 illustrates the technical chapters of this work.

In this thesis, **Chapter 2** describes the proposed smart metering system in Great Britain. An overview of the availability of measurements according to the smart metering equipment technical specifications (SMETS 2) is presented. This chapter provides a concise introduction to cluster analysis, load estimation and state estimation methods. An up-to-date literature review of clustering based load estimation and state estimation methods is summarised.

In **Chapter 3**, a clustering algorithm based on the k-means cluster analysis method was developed to group similar consumption patterns and extract characteristic load profiles, defined in terms of the cluster centres, from measurements of individual and aggregated smart meters. The clustering of both daily and segmented load profiles was investigated. Segmented profiles cover a time window that is less than or equal to 24 hours. The best segmentation time window was determined.

In **Chapter 4**, a load estimation algorithm was developed to estimate missing and future measurements of smart meters. The load estimation algorithm investigated the application of daily and segmented cluster centres and different distances functions to produce the required load estimates. Several scenarios simulating different durations of lost measurements were carried out to assess the accuracy of the estimated measurements.

In **Chapter 5**, an integrated load and state estimation algorithm was developed using both the load estimation algorithm and an IRWLS state estimator. Initially, outputs of the IRWLS state estimator were validated using measurements collected from an LV microgrid. Finally, several simulations were carried out to examine the performance of the integrated load and state estimation algorithm using the IEEE 33 bus network.

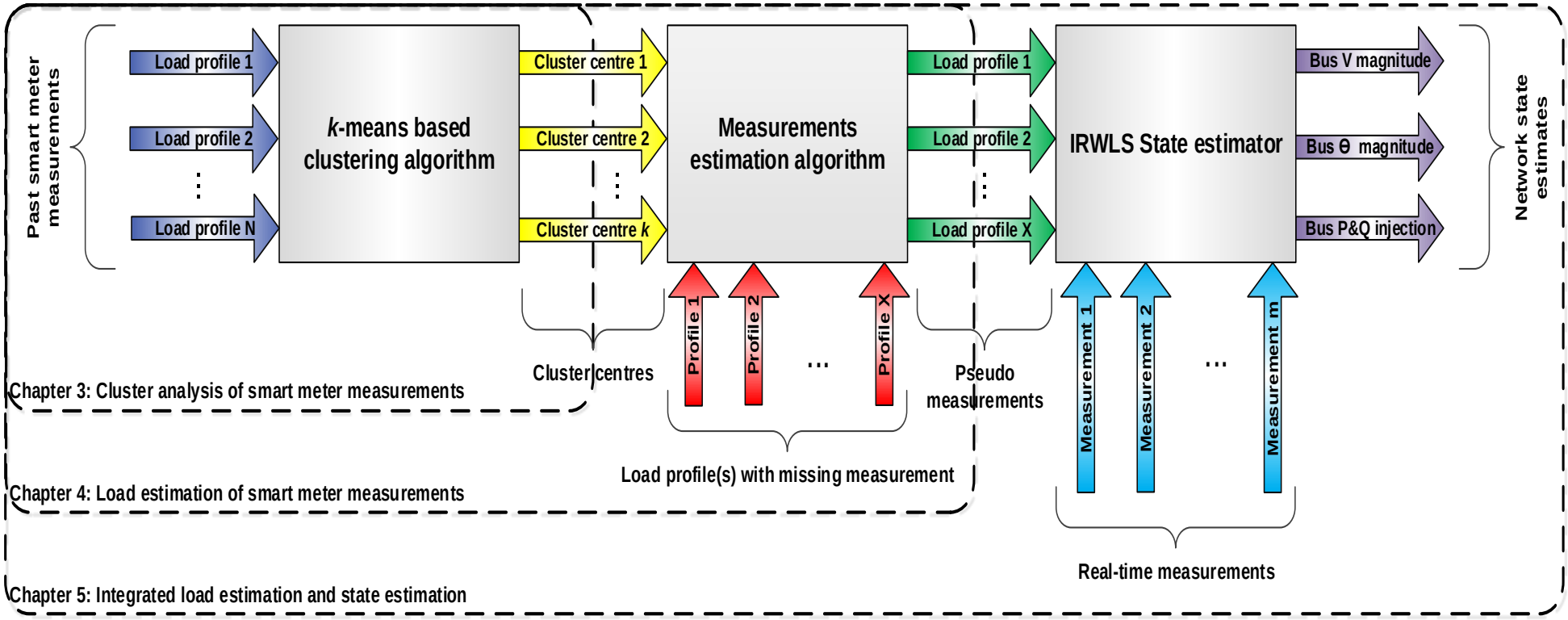


Figure 1.2 Outline of the presented research

1.4 Contributions of the Thesis

The main contributions of this work are:

- The development of a clustering algorithm based on the k-means cluster analysis method.
- The development of an adaptive approach to determine the required number of clusters.
- The development of a load estimation algorithm to estimate missing and future measurements of smart meters and provide the required pseudo measurements to state estimation algorithms.
- The development of an integrated load and state estimation algorithm to extend the observability of MV distribution networks.
- An investigation of the performance of an IRWLS state estimator using actual measurements from a test LV microgrid.
- A demonstration of the integrated load and state estimation algorithm using measurements of domestic smart meters.

Chapter 2 The Great Britain Smart Metering System

2.1 Introduction

Presently, the Government and industry in the United Kingdom are facing several energy challenges. The challenges defined in terms of the energy trilemma are:

- **Energy Security**: Includes the effective management of primary energy supplies from different resources, the reliability of energy infrastructure and the ability to meet both the present and future energy demands;

- **Energy Equity**: Encompasses the provision of energy supplies to all consumers at affordable prices; and

- **Environmental Sustainability:** Guarantees the achievement of supply and demand side energy efficiencies as well as the development of sustainable energy supply from renewable and other low carbon resources [21].

The installation of smart meters is one of the responses of the UK Government to overcome these challenges. The deployment of smart meters will bring advantages to end consumers (domestic and non-domestic), energy suppliers, and network operators. Providing near real-time consumption information to the end users will help them to manage their energy use, save money, and reduce GHG emissions. The installation of smart meters, at its peak, will also support 10,000 jobs [14]. At the same time, smart meters will benefit distribution network planning and operation, as well as demand management. In this regard, the measurements collected from smart meters will enable more accurate demand forecasts, allow improved asset utilisation in distribution networks, locate the outages and shorten the supply restoration times, and reduce the operational and maintenance costs of the networks [5], [14], [15].

This chapter presents the structure of Great Britain's Smart Metering System as proposed by the Department of Energy and Climate Change (DECC). This is complemented with an introduction to cluster analysis methods followed by the state-of-the-art of load and state estimation research.

2.2 An Overview of the GB Smart Metering System

Electricity and gas suppliers in Great Britain will install about 53 million smart meters in homes and microbusinesses by the end of 2020. The smart metering system, shown in Figure 2.1, comprises Energy Consumers, the Data Communications Company (DCC), and the DCC Service Users.

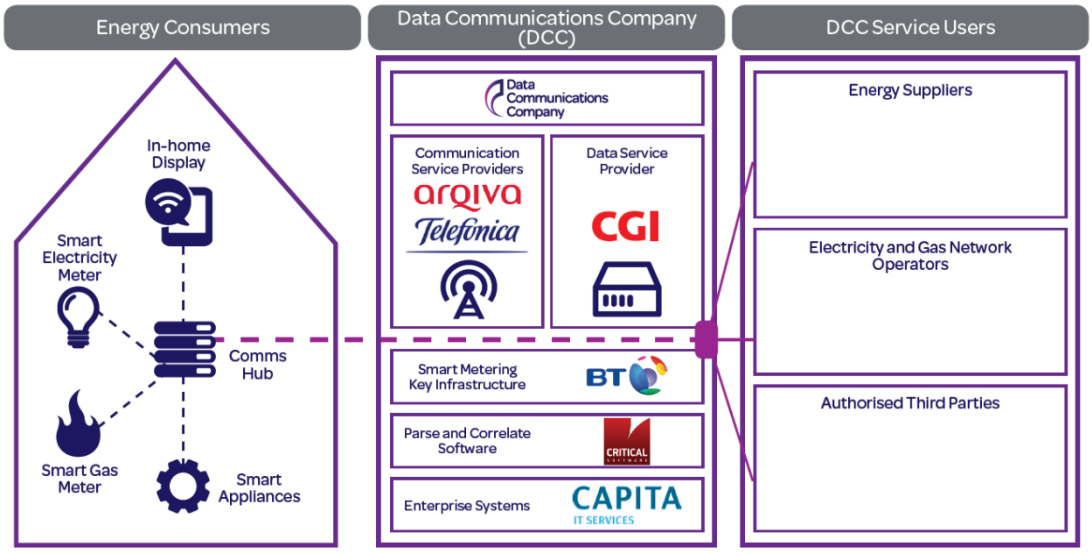


Figure 2.1 Great Britain's proposed smart metering system [15]

2.2.1 Energy consumers

These are the domestic premises and microbusinesses where the smart meters will be installed. The main elements of this block are:

a. The smart meters (Gas and Electricity);

b. The In-Home Display (IHD);

c. The Smart Appliances (if any exist); and

d. The Communication Hub (CH).

The smart meters will be capable of measuring and storing real-time consumption and tariff data. For recording energy, power, and voltage, an electricity smart meter possesses the following storage capabilities [16], [22]:

- The ability to store the daily electricity consumption in kWh for each of the previous 731 days;
- The ability to record half hourly measurements for:
    - 13 months of electricity consumption (in kWh);
    - 3 months of active energy (kWh) exported;
    - 3 months of reactive energy imported; and
    - 3 months of reactive energy exported;
- The ability to calculate the half-hourly mean values of active power imported and exported;
- The ability to record the maximum demand active power import and export over a half hour period (since last reset);
- The ability to calculate and record the mean RMS voltage over a configurable period.

A gas smart meter possesses the following storage capabilities [22]:

- Gas consumption (energy in kWh and cost in currency units) of 9 days (the current day and the prior 8 days);
- Gas consumption (energy in kWh and cost in currency units) of 6 weeks (the current week and the prior 5 weeks);
- Gas consumption (energy in kWh and cost in currency units) of 14 months (the current month and the prior 13 months);
- 13 months of half-hourly gas consumption data.

Smart meters will also have the capability to pair with Consumer Access Devices (CADs). A Consumer Access Device is capable of extracting real-time data from the smart meter. CADs will help the consumers to become more energy efficient and cost effective with regard to their consumption. The In-Home Display is one type of CADs [16], [23].

The IHD is an electronic device that will be linked to at least one smart meter. The IHD represents the first step of interaction between the consumers and their smart meters. IHDs will visualize the amount and the cost of the actual energy consumption within the premises. The IHD will be capable of displaying the following variables [22], [24], [25]:

- Cumulative daily, weekly, and monthly energy consumption (in kWh);
- The cost of the cumulative daily, weekly, and monthly energy consumption (in currency units);
- The historic energy consumption for the last 7 days, 4 weeks, and 12 months;
- The cost of the historic energy consumption for the last 8 days, 4 weeks, and 12 months;
- The active tariff price for consumption in currency units per kWh;
- The instantaneous active power import in kW; and
- The cost of maintaining the instantaneous active power import for one hour.

A smart appliance [26] is defined as “a product that uses electricity for its main power source which has the capability to receive, interpret and act on a signal received from a utility, third party energy service provider or home energy management device, and automatically adjust its operation depending on both the signal’s contents and settings from the consumer”. The list of smart appliances includes (but is not limited to) washing machines, tumble dryers, dishwashers, fridges and ovens. Smart appliances have been used to shift the peak load, provide balancing services, and to support the operation of distribution networks [27].

The Communication Hub (CH) is a device that allows the smart meter, IHD, and smart appliances to connect to each other [28]. In a consumer’s home, the smart meters, IHDs, smart appliances, and the CH will establish a wireless Smart Meter Home Area Network (SM HAN). The ZigBee® [29] protocol is used to empower the communication within the SM HAN. ZigBee is a wireless network protocol that uses a frequency of 2.4GHz (784MHz in China, 868MHz in Europe, and 915MHz in the USA and Australia [29]) to provide high rates of data transfer at a low cost and low power consumption. The CH acts as an interface that allows the connection of the SM HAN to the DCC and the DCC Service Users.

2.2.2 Data Communications Company

The Data Communications Company (DCC) is the second block in the smart metering system shown in Figure 2.1. The DCC will provide communication links between millions of smart meters and systems of energy suppliers, network operators and energy service companies [30].

The DCC license was awarded to Smart DCC, owned by Capita PLC, which is responsible for the operation, management, and maintenance of the data and communication network to all consumers regardless of their energy supplier. The DCC will connect the smart meters with the DCC Users across the Smart Metering Wide Area Network (SM WAN).

Two service providers were selected to provide the communications over the SM WAN. Telefonica will provide communication services for the Central and South Regions of Great Britain. Communication services provided by Telefonica consist of general packet radio service (GPRS) in the regions of existing cellular (also known as mobile) networks. This will be complemented with small scale radio frequency (RF) mesh technology in the areas that are hard to cover using cellular networks [31], [32]. A comparison between the two different technologies is provided in Table 2.1.

Cellular network technologies are the main communication links between the smart meters and the DCC Users. High rates of data transfer, lower latency, and ease of deployment and operation make cellular networks the preferred option for communications across the WAN. The entries in Table 2.1 represent the actual average values of the data rate and the message delivery latency measured by Ofcom in the UK.

Table 2.1 A comparison between different WAN technologies [31], [33], [34]

	RF Mesh	***Mobile 2G***	***Mobile 3G***	***Mobile 4G***
Average Data Rates	9.6 – 100+ Kbps	236.8 Kb	6.1 Mbps	15.1 Mbps
Message Delivery Latency	1 – 60 sec	~ 1 sec	66.7 ms	55 ms
Ease of network deployment and operation	• Responsibility of utilities to deploy and maintain the networks • Requires new communication network infrastructure	• Deployed and maintained by cellular network operators • Strong ecosystem driven by a lower cost	• Same as Mobile 2G networks	• Same as Mobile 2G networks

Arqiva is the communication service provider for the DCC in Scotland and Northern England [32], [35]. Arqiva provides communication services using Sensus FlexNet™ communication platform. The FlexNet™ technology uses exclusively licensed high power RF to create communication links between millions of smart meters and the DCC [36], [37].

British Telecommunications (BT) is providing the Smart Metering Key Infrastructure (SMKI). The SMKI service is used to issue digital certificates to ensure the authenticity, integrity, and approval of communications and data between the smart meters and the DCC Users. The SMKI represents one of the principal mechanisms by which the communications between the smart meters and the DCC Service Users are secured [38].

CGI Group is the Data Service Provider for the DCC. CGI is developing and operating DCC Data Services. The Data Services will link the smart meters with the business systems of the DCC Service Users. These services will enable the secure exchange of messages between the smart meters and the systems of the DCC Service Users [30], [39].

Another company, Critical Software, provides the Parse & Correlate software to help the DCC Service Users communicate with millions of smart meters. The parse process implements the translation of smart meter responses and alerts into a technical

format suitable for integration with the systems of the DCC Users. The correlate process checks the semantic accuracy of a service request originating from an energy supplier with its equivalent HAN interface command.

2.2.3 DCC Service Users

The third block of Figure 2.1 shows the DCC Service Users. These include energy suppliers, electricity and gas network operators, and other parties eligible to become DCC Service Users [40].

The smart metering system was designed to conserve the privacy of consumers to a maximum. Principally, the energy suppliers can only access the monthly energy consumption for billing purposes. Moreover, for the DCC Service Users to access the detailed consumption information stored in smart meters, consumer consent must be obtained.

Figure 2.2 demonstrates the flow of information between DCC Users and consumers so that smart meter measurements are retrieved via the DCC.

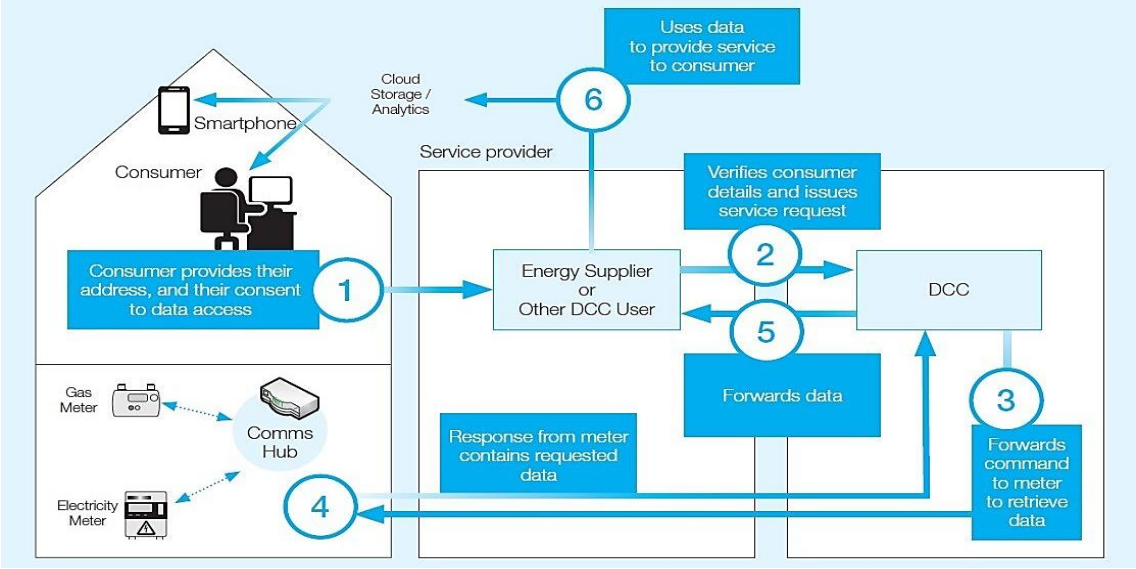


Figure 2.2 Remote access via WAN using DCC [16]

Figure 2.2 shows that once the consumer consent is granted to a DCC User, the DCC User will be able to issue an ad-hoc request to retrieve data including half-hourly consumption measurements and tariff information from the meter of that consumer. All

requests and responses between the DCC Users and smart meters must pass through the DCC [16], [40].

A summary of the smart metering data that the DCC Users can request is illustrated in Table 2.2. The response times are tabulated accordingly. The acronyms used in Table 2.2 are.

- EIS: Electricity Import Supplier;
- GIS: Gas Import Supplier;
- ENO: Electricity Network Operator;
- GNO: Gas Network Operator;
- EES: Electricity Export Supplier;
- OU: Other User – such as businesses that offer switching services to customers; and
- RSA: Registered Supplier Agent.

The Import Suppliers (e.g. British Gas) are the suppliers of gas and electricity to domestic premises and microbusinesses. The Network Operators are the Gas and Electricity Distribution Network Operators (DNOs). A total number of six electricity and four gas DNOs cover Great Britain [41], [42]. The Electricity Export Supplier [43] is the electricity supplier to which a consumer (with an installed DER) is selling electricity. Other Users are all eligible users of the DCC Services with a role other than an Import (or Export) Supplier, Network Operator, or a Registered Supplier Agent. The Registered Supplier Agent (RSA) defines the User that can be appointed as a Smart Meter Operator (for Electricity Smart Meters) or a Meter Asset Manager (for Gas Smart Meters).

Table 2.2 Requests and Response Times of DCC Service Users [20], [44]

Service Description	*Eligible DCC Users*	*Target Response Time*
Update Supplier Name	EIS GIS	24hr
Restrict access for Change of Tenancy	EIS GIS	30 sec
Read Instantaneous Import Registers	EIS GIS ENO GNO	30 sec
Read Instantaneous Export Registers	EES ENO	30 sec
Retrieve Import Daily Read Log [2]	EIS GIS	30 sec
Retrieve Export Daily Read Log [3]	EES	30 sec
Retrieve Daily Consumption Log	EIS GIS ENO GNO OU	30 sec
Read Active Import Profile Data	EIS GIS ENO GNO OU	30 sec
Read Reactive Import Profile Data	EIS ENO OU	30 sec
Read Active Power Import	EIS ENO	30 sec
Read Export Profile Data	EES ENO OU	30 sec
Communication Hub Status Update	EIS GIS	30 sec
Read Maximum Demand Import Registers	EIS ENO	24hr
Read Maximum Demand Export Registers	EES ENO	24hr
Read Network Data	EIS GIS ENO GNO	30 sec
Read Device Configuration (Voltage)	EIS ENO RSA	30 sec
Read Device Configuration (Gas)	GIS GNO RSA	30 sec
Update HAN Device Log [4]	EIS GIS OU	30 sec
Read Supply Status	EIS GIS ENO GNO RSA	30 sec

The GB smart metering system faces significant technical and operational challenges. The technical challenges include:

- The lack of sufficient signal strength (for both mobile and RF networks);
- The shortage of tools to detect mobile network failure; and
- The indoor/outdoor placement of meters.

[2] Where a change of supplier occurs on any day, both the new supplier and the old supplier will be eligible to retrieve the daily read log for that day.

[3] This is the same as above.

[4] Other Users may only add (or remove) Type 2 Devices that are not IHDs to (or from) a HAN Device Log. A Type 2 CAD means the CAD that does not store or use the Security Credentials of other Devices to communicate with them via its HAN Interface.

The operational challenges can be categorised as:

- Planned or unplanned maintenance of the system;
- Software and hardware faults or malfunction of the smart meters; and
- Customers unwilling to communicate their energy consumption data.

As a result of these challenges, the measurements of smart meters will be prone to time delays or even temporary loss when requested by the DCC Users [31], [45]–[49]. It is therefore necessary to develop a load estimation algorithm to replace missing and future smart meter measurements.

2.3 An Overview of Load and State Estimation

The rollout of smart meters in Great Britain will provide information about the consumption patterns of LV customers. The majority of these smart meters will be installed at domestic premises [50], [51]. Volumes of data ranging from several hundreds of gigabytes to tens of petabytes or even exabytes (One Petabyte $= 10^{15}$ Bytes = 1000 Terabytes, One Exabyte $= 10^{18}$ Bytes = 1000 Petabytes) will be available for the energy suppliers and the network operators to exploit [18], [52]–[54]. The volume of the data will vary according to several factors. These include

- The number of installed smart meters;
- The number of received smart meter messages that transfer the meter reading into the DCC Service User; and store the corresponding reading in an appropriate format (e.g. database or text file format);
- The message size (in bytes per message) and;
- The frequency of recording the measurements – e.g., every 15 or 30 minutes etc.

To make the best use of these massive volumes of data, it is necessary to transform the smart meter measurements into practical information such as the typical daily profile of a customer class. This information is utilised to improve the operation, planning, and control of distribution networks.

Statistical, regression, engineering and time-series methods have been developed [10], [55], [56] to analyse domestic profiles and extract the required

information, for example the class load profile, from the collected domestic measurements in power networks. However, these methods face many challenges and have several limitations when smart meter measurements become widely available. For example, regression methods (such as segmented polynomial regression) can be costly and complex to implement when large volumes of consumption measurements become available through the smart metering system. Sometimes, these methods suffer from multi-collinearity between variables, and the extracted load profiles tend to be based on the average profiles of a customer class (e.g. residential, industrial or commercial). At the same time, AI methods (such as neural networks) can also suffer from multi-collinearity problems and that may not represent the actual temporal and magnitude components of an electricity load profile.

One efficient approach to overcome the drawbacks of traditional methods is the employment of cluster analysis methods to extract the necessary information from the raw measurements of smart meters. Clustering methods do not suffer from many of the problems highlighted above especially when these methods are applied prior to the implementation of any statistical analysis.

Clustering aims to partition a set of load profiles into groups (or clusters) provided that the number of clusters is less than the number of profiles. The load profiles are clustered based on the principle of "maximising the intra cluster similarity and minimising the inter cluster similarity [5]". In this manner, a cluster is defined as a group of load profiles having similar patterns (usually in proximity to one another). Each cluster is characterised by a cluster centre which is a summary profile representing the consumption pattern of that cluster. Clustering analyses the profiles without consulting a known class label, unlike classification and prediction, which analyse class-labelled data points.

2.3.1 k-means cluster analysis

The k-means is a classic partitional cluster analysis method. The method groups a number of load profiles into k discrete clusters by minimising the intra-cluster sum of squares.

The simplicity, fast convergence, and the ability to handle big datasets mark the k-means as one of the most widely used data clustering methods [61]–[63]. However,

[5] This is the same as maximising both the intra cluster similarity and the inter cluster dissimilarity.

the k-means method is sensitive to the random selection of cluster centres that is the initial step towards the clustering of a dataset [63], [64]. To overcome the impact of the initial random selection of cluster centres, the k-means is repeatedly run for n different times [64] with different initial choice of cluster centres. One other disadvantage of the k-means method is that the number of clusters must be defined prior to the clustering of any dataset. The selected number of clusters (k) might not reflect the optimal number of the clusters. Therefore (k) should be chosen according to some cluster selection criterion [55], [65]–[67].

2.3.2 Hierarchical cluster analysis

Hierarchical clustering methods group the load profiles into the required number of clusters using a series of nested partitions. The structure of these partitions, called a dendogram, is similar to a clustering tree. Hierarchical cluster analysis methods can be either agglomerative or divisive.

One of the advantages of hierarchical clustering is that prior knowledge of the number of clusters is not necessary. Additionally, hierarchical clustering can partition a dataset even when no information about the composition of the dataset exists [64], [68], [69]. Nevertheless, these advantages are hampered by the low computational efficiency of hierarchical clustering methods. The computational (or time) complexity of hierarchical clustering is $O(n^2)$, i.e., quadratic with the size of the dataset. The researchers in [64] report the use of hierarchical clustering with a 5% portion of their original dataset so as to overcome the time complexity and the need to employ large computational resources. Hierarchical clustering is sensitive to noise and outliers in the input dataset. These methods provide no guide to the choice of the optimal number of clusters [70], [71].

2.3.3 Self-Organising Maps

Self-Organising Maps (SOM) is one type of unsupervised neural networks [72]. The SOM reduces an input high-dimensional dataset to a bi-dimensional output dataset. In other words, the SOM maps a set of data vectors from an original n-dimensional space to a bi-dimensional output space. At the same time, any relationships between the original input vectors are preserved in the output space [73].

Self-Organising Maps are mainly recognised for their robustness, computational efficiency, and tolerance to noise within the input dataset. On the other hand, several

factors affect the performance of these techniques. The requirement to adjust large numbers of parameters, such as the weights of network connections and the selection of the neighbourhood function, is one of the drawbacks of SOM. The overall quadratic time complexity of SOM renders the application of methods infeasible for the clustering of big datasets [74], [75].

A summary of the aforementioned cluster analysis methods and the corresponding research papers is illustrated in Table 2.3.

Table 2.3 A summary of load profile clustering methods [58], [76]–[78]

Cluster analysis method	***Advantages***	***Disadvantages***	***References***
k-means	• Simple • Efficient • Scalable • Ability to handle big data • Linear complexity with the size of dataset	• Sensitive to the initial selection of cluster centres • Number of clusters must be defined in advance • Sensitive to noise and outliers • Provides local (not global) optimum solution	[55], [61]–[65], [67], [79]–[99]
Hierarchical clustering	• Easy to implement • Prior knowledge of number of clusters is not required	• Time consuming • Sensitive to noise and outliers • Sensitive to the selected distance function • Difficult to identify the output number of clusters by the dendogram.	[64], [69], [80], [86], [87], [90], [97], [100]–[103]
Self-Organising Maps (SOM)	• Computationally simple • Highly efficient • Maps high-dimensional data into low-dimensional data • Fault- and noise-tolerant capabilities	• Impacted by the weights of network connections • Impacted by the learning technique • Sensitive to the selected neighbourhood function • Time consuming as the dimensions of the data increase	[55], [62]–[64], [67], [73], [79], [85], [86], [88], [90], [91], [96], [97], [104]–[108]

based on the literature reviewed and presented in Table 2.3, cluster analysis methods were the principle technique used to construct the typical load profiles (TLPs) of different types of customers in power networks. The TLPs, were employed for pricing and settlement purposes. Less attention has been paid to the use of cluster analysis methods for load estimation and load forecast.

This research applies the k-means cluster analysis method to group similar consumption patterns, of domestic smart meters, and build the cluster centres. The cluster centres, which represent characteristic load profiles, are employed by a load estimation algorithm to estimate missing and/or future measurements of smart meters and therefore provide pseudo measurements to a state estimation algorithm.

2.3.4 Load estimation

In power networks, load estimation is used to calculate the load of a customer, or groups of customers. Load estimation is also used to replace bad (or missing) measurements of the load [109], and provide pseudo-measurements to state estimation.

Load estimation methods use historical and real-time measurements to estimate the required missing or future measurements. Historical measurements include the data collected from load surveys, transformer kVA ratings, and customer energy meter readings. Real-time load measurements include smart meter measurements and measurements of Automatic Meter Reading (AMR) systems.

Statistical analysis and artificial intelligence (AI) methods have been extensively used to estimate and forecast the load in power network networks [110]–[116]. [117]. The application of cluster analysis methods to estimate the load in power networks has been limited [65], [99], [108], [118].

Wan and Miu [119] investigated the application of state estimation to solve the problem of load estimation in unbalanced distribution networks using two different approaches. The first approach defined load estimation as a constrained optimisation problem and attempted to estimate the network states and load parameters simultaneously. The second approach estimated the voltage magnitudes and angles which were used later to estimate the loads at each busbar of the distribution networks. When noisy voltage measurements were applied, the second approach produced load estimates of higher accuracy than the first approach which implied that the second approach was less sensitive to errors in voltage measurements.

A real-time load-modelling algorithm based on the measurements of AMR systems was proposed in [120]. The algorithm initially estimates the customer load by generating the load profile of a customer class. The load of a transformer is then estimated using the customer load estimates. Loads of different types of customers were estimated using a day-before, and the mean measurements of the past four (or more) days. Wang and Schulz concluded that the mean measurements produce more accurate load estimates than the day-before measurements.

A regression based prediction method to estimate missing active power measurements of the load was developed in [109]. The method makes use of the strong correlation between loads within the same geographical area to address the low measurement redundancy and provide reliable load estimates. Baran et al. demonstrated that the mean error of load estimates depends on the type (domestic, commercial or industrial) and location of the lost measurements, and the presence of updated historical data.

Wu et al. [12], [121] developed a machine learning based load estimation algorithm that was applied to provide pseudo measurements to a state estimator. The machine-learning algorithm was designed to identify, store, update and utilise repeating patterns of historical load data. The algorithm was applied to provide load estimates by identifying similar patterns hidden inside the load time series of MV nodes. The load estimator was equipped with incremental self-adaptive learning ability which resulted in high quality load estimates.

An artificial neural networks (ANN) approach for modelling active and reactive pseudo measurements in distribution network state estimation was presented in [122]. Load profiles obtained from the UK Generic Distribution System (UKGDS) and a small number of power flow measurements derived from load flow simulations were applied to train the ANN. The ANNs were capable of generating higher quality pseudo measurements than the average load profiles of the UKGDS. Additionally, the ANNs could be easily retrained to take into consideration the annual load growth and system expansion, and the availability of additional real-time power flow measurements.

2.3.5 State estimation

State estimation exploits pseudo measurements and a minimum set of real-time measurements to extend the observability and define the operating state of a distribution network. Weighted least squares (WLS) estimators have been widely applied and tested in radial distribution networks [12], [122]–[135] and meshed transmission networks

[136]–[139]. In these techniques, each measurement is assigned a weight which is inversely related to the error variance of that measurement. WLS estimators can either be node voltage or branch current estimators. Node voltage estimators calculate the voltage magnitude and voltage phase angle at each busbar of the network.

Baran and Kelley [128] proposed a three-phase WLS state estimator. The state estimator aimed to improve the accuracy of load forecasts used for real-time monitoring and control of distribution feeders. Real-time measurements comprising the magnitudes of busbar voltages, branch currents, and line power flows were applied to the state estimator. Historical load injections were used as pseudo measurements. This approach omitted branch current measurements from the first three iterations to reduce the complexity of the WLS algorithm. Power flow measurements were more effective than current measurements in the identification of erroneous data. Baran and Kelley stated that for a limited number of real-time measurements, the effectiveness of the state estimator was strongly dependent on the accuracy of pseudo measurements.

Li [140] proposed a state estimator based on the WLS method, and three-phase modelling of a distribution network. A stochastic load model was developed to provide pseudo measurements to the state estimator. Customers' load was synthesized using customer monthly billings, kW rating, customer type, and load patterns to establish the load model. Li concluded that the load error correlation, availability and placement of real-time measurements, and accuracy of load estimates (i.e., pseudo measurements) affected the accuracy of state estimates.

Lu et al. [127] adopted a different approach to solve the WLS state estimation problem. All real-time and pseudo measurements were transformed into their current equivalents in the rectangular coordinate system. However, the operating state was defined in terms of the voltage phasors at each busbar of the network. The transformation was carried out as a necessary step to assure the convergence of the algorithm. The WLS algorithm investigated the applicability of busbar voltage and line current magnitudes, and line power flows. As compared to line current measurements, line power flows produced better estimation results since the former measurements led to divergence in some cases. A special procedure, which simply involves using small weights for the first three iterations of state estimation, was followed once current measurements were applied. The weights were later restored back to their original values. This resulted in longer computational times than if no current measurements were included.

A weighted least squares estimator based on transmission network state estimation was proposed in [141]. The estimator was part of a Distribution Management System Controller (DMSC) applied to estimate voltages of a distribution network; and take the required actions to guarantee that voltage limits were not violated as a result of certain DG penetrations. The state estimator utilised real-time measurements of voltage magnitude and line power flows; and pseudo measurements based on estimated load consumption. Cobelo et al. concluded that the number, location, and accuracy of real-time voltage measurements affected the accuracy of estimated busbar voltages. The authors also reported that accurately estimated voltage magnitudes do not guarantee accurate active line power flow estimates. The placement of real-time active line power flows resulted in a marginal improvement in the local voltage estimates (of the busbars close to the measurement). However, this observation was true only for measurements placed along a main feeder. If the measurement was placed along a feeder connecting an isolated load to the main feeder, no impact was observed. The authors also reported that the accuracy of the load estimates was a key factor only for very high network loading (~ 200% of the original load). The estimated line power flows were significantly impacted by the accuracy of the estimated load.

Korres et al. [138] proposed a multi-microgrid state estimator based on the WLS method to estimate the network state that was defined in terms of the busbar voltage phasors and the status of switching devices in a distribution network. The state estimator was assessed in both grid-connected and islanded modes of operation. The accuracy of busbar voltage estimates was significantly impacted by the accuracy of load estimates and real-time measurements especially DG voltage magnitudes. The results showed that erroneous measurements affected the estimated local, near the location of the bad measurement, busbar voltages and line power flows. Voltage angles were more sensitive than voltage magnitudes to errors in load estimates. Results also revealed that high redundancy of real-time measurements results in correct identification of errors in circuit breaker statuses.

The WLS estimators that were reported in [128], [138], [140], [141] defined the operating state in terms of the voltage magnitudes and voltage phase angles at each busbars of a distribution network. The WLS state estimation method was formulated to define the state of a distribution network in terms of branch current magnitudes.

Baran and Kelley [142] developed a branch current based WLS approach to solve the state estimation problem in three phase distribution networks. All real-time and pseudo measurements were converted into their equivalent current flows and injections. All voltage measurements except the substation voltage measurements were ignored.

Following the same approach in [127], the current measurements were excluded for the first three iterations to guarantee convergence of the proposed algorithm and estimation of more accurate current phase angles. Baran and Kelley compared their branch current state estimator with the one introduced in [128] and concluded that the latter state estimation algorithm provided better performance in correcting the voltage measurement errors.

Lin et al. [143] developed a revised linear WLS estimator using the original branch current estimator presented in [142]. The linear WLS estimator decoupled the real and imaginary parts of current magnitude resulting in a linear estimator with a constant gain matrix. The constant gain matrix led to significantly faster convergence of the linear WLS estimator than the original estimator. The mutual coupling, implemented in the original estimator, between different phases in weakly meshed networks was discarded. The requirement to exclude current measurements in the first three iterations was not reported in the linear WLS estimator.

Wang and Schulz [144] developed a polar version of the branch current estimator, that was introduced by Lin et al. [143], to find the operating state of a distribution network in terms of the magnitude and phase angle of the branch currents. The authors decoupled the original three-phase estimator to increase the computational efficiency and decrease the calculation time accordingly. Real-time measurements of line power flows, current magnitudes, and voltage magnitudes were used while load estimates based on AMR systems were applied as pseudo measurements. The impact of measurement placement was considered. The authors concluded that the installation of branch power flow and/or current magnitude measurements significantly improved the quality of state estimates. This was true only for measurements installed near the MV/LV substation or along the main feeder.

Singh et al. [131] investigated the capabilities of the WLS, weighted least absolute value (WLAV), and Schweppe Huber Generalised M (SHGM) estimators. The WLS estimator was the best solver for the distribution state estimation problem. This was true only when the measurement errors and noise characteristics were known and normally distributed. WLAV and SHGM estimators were not applicable in distribution networks. The WLAV estimator failed to work because it treated every pseudo measurement as bad data. In distribution networks, since there are no redundant measurements, pseudo measurements cannot be eliminated. The SHGM estimator produced inconsistent state estimates when erroneous real-time measurements were applied. An estimator is consistent if it converges to its estimand as the sample size

tends to infinity [145]. The increase in errors in real-time and pseudo measurements produced less accurate network state estimates.

The results and conclusions that were reported in [127], [128], [131], [138], [140]–[144] indicate the impacts of the accuracy of real time and pseudo measurements upon the outputs of WLS estimators. Abur and Exposito [136] reported the impacts of measurement error distributions and outliers upon the performance of WLS based state estimators. Robust estimators have been developed to find the accurate system states; and detect, identify, and eliminate the inherent errors in the measurements, network model, or system parameters [136]. IRWLS estimators are robust state estimation methods [146], [147]. In contrast to WLS estimation, which assigns the same weight to a measurement throughout all iterations, IRWLS estimators iteratively change the measurement weights. Measurements with large residuals will have their weights reduced iteratively [148]–[151].

Iterative reweighting of the measurements within a conventional WLS state estimator was proposed [148]. The algorithm combined the computational efficiency of WLS algorithm and the robustness of a linear programming estimator [149]. Irving and Sterling applied the IRWLS estimator to estimate the operating state of the IEEE 118-bus test system. Low, medium and high measurement redundancies, and measurements with gross errors were investigated. The algorithm was capable of handling large numbers of erroneous measurements. When the range of modified weights was restricted, observability problems were not reported during the bad data rejection process.

Smith et al. [150] proposed an IRWLS estimator capable of rejecting bad measurements in the presence of noise. Measurement weights were modified based on the values of measurement residuals. Measurements with large residuals were given small weights whereas large weights were assigned to measurements with small residuals. The algorithm was applied for both power and water system state estimation. In both systems, the final state vector was not impacted by Gaussian noise and gross errors in the measurements.

Wu et al. [12] developed a closed-loop state estimator for under-determined (with limited real-time measurements) or over-determined (but with delayed smart meter measurements) MV distribution networks using the principles of IRWLS. Several case studies were implemented taking into consideration the impacts of measurement errors; type, location and accuracy of the measurements; and the temporary failure of smart metering communication systems. The authors concluded that their estimator

outperforms the classical WLS [136] estimator in the presence of gross errors. The addition of gross errors to multiple leverage measurements resulted in the failure of WLS estimator. However, the estimator they developed was capable of providing reliable busbar voltage magnitudes and angles.

The work carried out in this thesis proposes an integrated load and state estimation algorithm based on the k-means cluster analysis method for load estimation and IRWLS method for state estimation.

Chapter 3 Cluster analysis of smart meter measurements

3.1 Introduction

A load profile clustering algorithm based on the k-means method was developed and used to extract characteristic load profiles from smart meter measurements. A cluster encompassed load profiles with similar consumption patterns and was defined in terms of the cluster centre which is a characteristic load profile. A cluster centre is the mean (average) value of the profiles comprising the cluster, calculated at each half-hourly time step of the profile. As a result, both the input load profiles and their cluster centres have the same duration defined in hours (or alternatively, the number of half-hourly measurements).

Clustering the load profiles of individual (at the 0.433kV LV level) and aggregated (at the 11kV MV level) domestic smart meters was investigated. Figure 3.1 illustrates the clustering of the load profiles of individual and aggregated smart meters.

Daily and segmented load profiles of individual and aggregated customers were clustered. Segmented profiles extend over different time windows that are less than or equal to 24 hours.

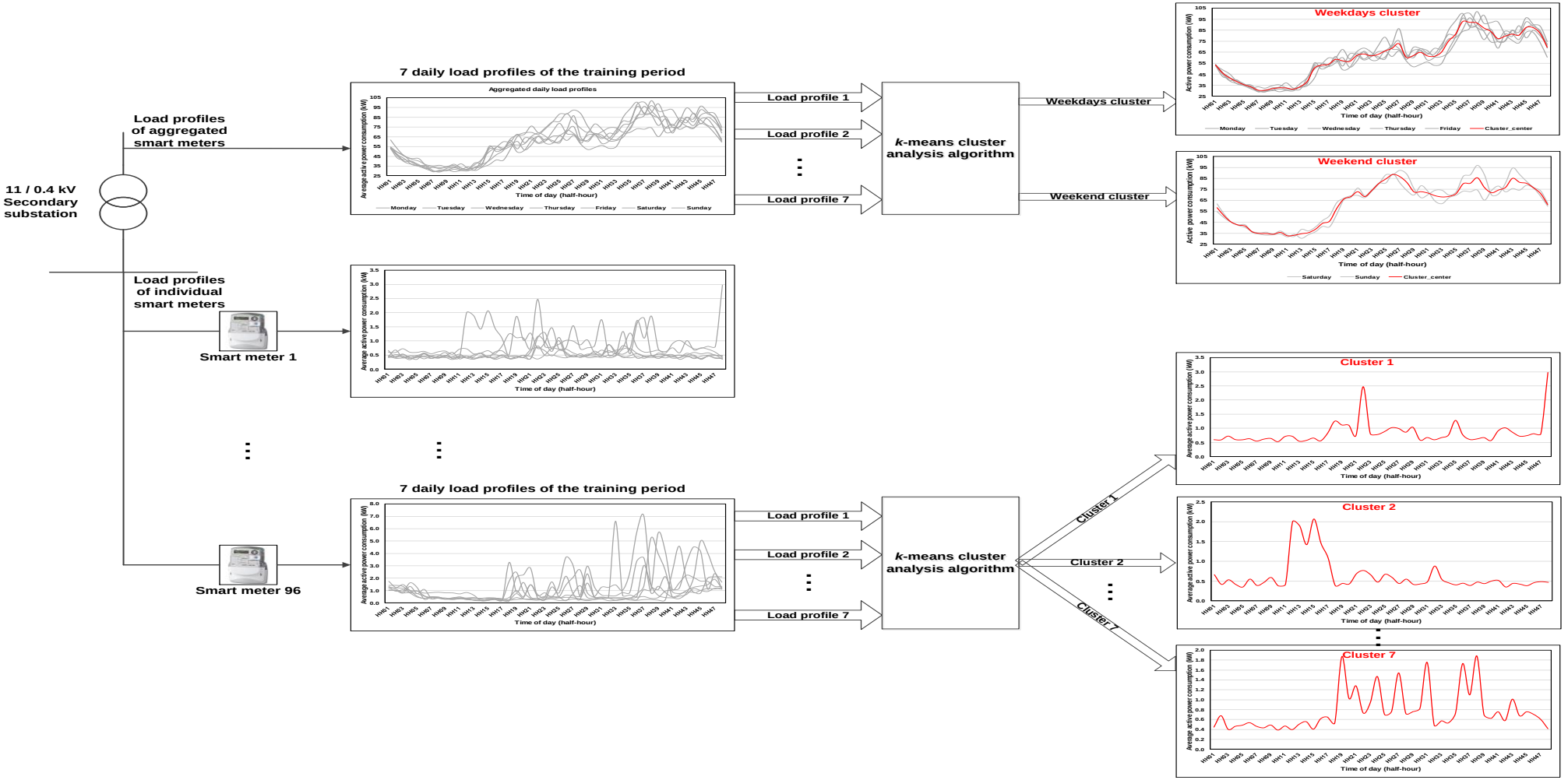


Figure 3.1 Cluster analysis of individual and aggregated smart meter measurements

3.2 Cluster Analysis Methods

Clustering is the grouping of similar load profiles into a number of clusters such that profiles within the same cluster are similar to each other, whereas profiles belonging to different clusters are as dissimilar as possible [78], [152]. Similarity (and hence dissimilarity) is defined in terms of the distance(s) between the profiles within the same cluster and profiles among different clusters.

Cluster analysis methods explore and discover the groups that might reside within a set of profiles without having any background knowledge of the characteristics of the profiles [153]. The applications of cluster analysis methods include load classification, segmentation, and load forecast [114], [154]. Figure 3.2 illustrates that clustering methods can be categorised into hard and fuzzy cluster analysis methods. Hard clustering, which is subdivided into hierarchical and partitional (also called centre-based) methods, assigns a load profile to a single cluster, whereas fuzzy clustering implies that a profile belongs to one or more clusters with certain degrees of membership [155].

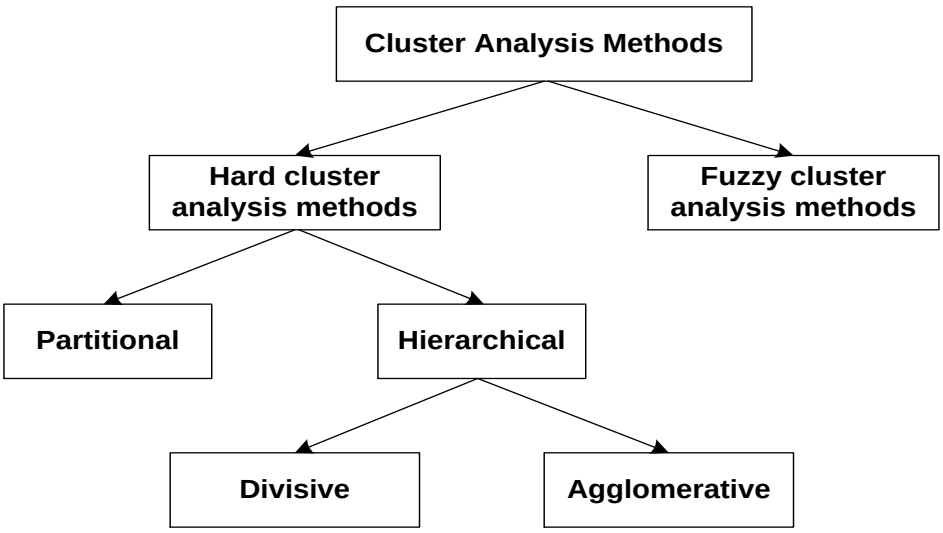


Figure 3.2 Classification of cluster analysis methods according to [6]

Alternatively, cluster analysis methods can be categorised according to the clustering principles into hierarchical, partitional, search-based, grid-based, density-based, model-based, fuzzy, and subspace clustering methods [78], [153], [155]–[157]. Figure 3.3 illustrates the alternative categorisation of cluster analysis methods.

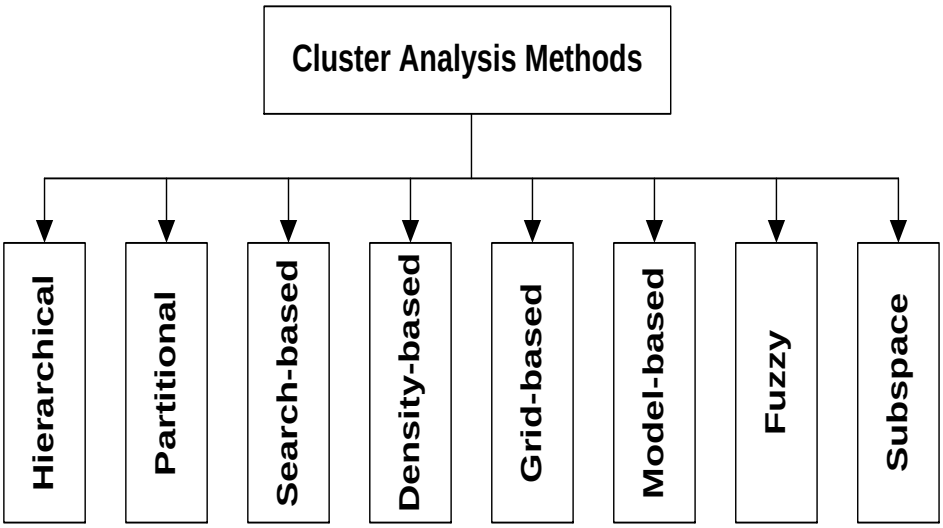


Figure 3.3 Alternative classification of cluster analysis methods

3.2.1 Hierarchical cluster analysis methods

Hierarchical cluster analysis methods group a set of profiles into the required number of clusters through a series of nested partitions. This results in a hierarchy of partitions leading to the final cluster. Hierarchical clustering methods are either agglomerative or divisive. Agglomerative (or bottom-up) clustering starts with each profile in its own cluster. Closest pairs of clusters are iteratively merged, to form new clusters, until all profiles are assigned to one final cluster. On the other hand, divisive (or top-down) clustering operates in the opposite direction to agglomerative methods. Initially, all load profiles are in one big cluster, which is repeatedly split into smaller clusters as the hierarchy advances.

The main advantages of hierarchical clustering methods are easy implementation and lack of necessity for prior knowledge of the required number of clusters. However, the drawbacks include irreversible false groupings, sensitivity to noise and outliers, and sensitivity to the cluster similarity measures that are used.

3.2.2 Partitional cluster analysis methods

Partitional (or centre-based) cluster analysis methods represent each cluster by a centre. The cluster centre is a summary description of all load profiles contained within the cluster. Cluster centres can be virtual points since these points do not usually exist within the set of inputs.

Partitional clustering divides a given set of profiles into a specific number of clusters by minimising the intra cluster distances between the profiles and their cluster centres. In these methods, each profile is assigned to one cluster such that the number of clusters, k, must be predefined or known in advance. Alternatively, k can be calculated by taking into consideration all possible values of cluster numbers and selecting the one that satisfies a specific criterion.

Advantages of partitional clustering methods include easy implementation, scalability, and ability to handle big datasets. However, prior knowledge of cluster numbers and provision of sub-optimal final cluster assignment are drawbacks of partitional methods [78], [155], [158].

3.2.3 Search-based cluster analysis methods

As opposed to partitional clustering, search-based cluster analysis methods explore the solution space beyond the local optimality to find the global optimal solution that will group a given set of input profiles [155].

3.2.4 Density-based cluster analysis methods

Density-based cluster analysis methods [71], [78], [155] characterise the clusters as high density, separated by low density, regions in the data space. These methods assume that the profiles within each cluster are drawn from a specific probability distribution. A mixture of several distributions defines the overall distribution of profiles. Density-based methods are capable of automatically detecting the clusters along with their numbers.

3.2.5 Grid-based cluster analysis methods

Grid-based [71], [78], [155] and density-based cluster analysis methods are similar in their capability of finding and defining the clusters in a large multidimensional space. Density-based clustering methods deal with the load profiles whereas grid-based methods are concerned with the space that surrounds these profiles. Grid-based methods divide the space that surrounds a dataset into a finite number of cells forming a grid structure, and group the patterns that are observed within this structure. The capability of handling large datasets, reduction of the computational complexity, and fast processing time are advantages of the grid-based clustering.

3.2.6 Model-based cluster analysis methods

Model-based clustering [71], [155], [159] assumes that the dataset is generated by a finite mixture of underlying probability distributions. Therefore, these methods attempt to optimize the fit between the given dataset and some probabilistic models. In model-based clustering, the prior knowledge of number of clusters is unnecessary. Furthermore, the availability of a variety of models that fit a given dataset are considered key advantages of these methods.

3.2.7 Fuzzy cluster analysis methods

Fuzzy cluster analysis methods [155], [160] are used to cluster complex datasets of possibly overlapping profiles when the use of hard clustering becomes impractical. Using fuzzy clustering, a load profile is assigned to more than one cluster with a degree of membership to define the likelihood of allocation to each cluster.

3.2.8 Subspace cluster analysis methods

Subspace clustering methods [155], [161]–[164] identify the clusters that reside in multiple overlapping subspaces of a given high dimensional dataset. Most traditional clustering methods fail to detect these clusters because the distance between any two points (or profiles) become almost the same, and different clusters can be embedded within different subspaces.

3.3 Smart Metering Data Structure

Domestic load profiles based on smart meter measurements were used in this research. The load profiles were obtained from the Irish Smart Metering [165] Customer Behaviour Trials (CBT). The Irish CBT is one of the largest and most statistically robust smart metering trials. These trials were implemented to investigate the impact of smart meter technology upon power consumption behaviour for different types of customers, and to identify a tipping point, at which the price of electricity stimulates a significant change in consumption. Customers' behaviour in terms of peak demand and overall electricity consumption was analysed combining smart meter technology with time-of-use tariffs and demand side management stimuli. The trials were carried out in the

course of 18 months (from 1st July 2009 until 31st December 2010), and covered more than 4200 domestic customers and 485 small-and-medium enterprises (SMEs).

The smart meters installed at the customers' premises recorded the daily load profiles which consist of 48 half-hourly (mean active power consumption) measurements a day. The first measurement (recorded at hour 00:30) is the mean power consumed between hours 00:00:00 and 00:29:59, whereas the last measurement (recorded at hour 00:00) is the mean power consumption between hours 23:30:00 and 23:59:59.

Three weeks of smart meter measurements collected from 96 randomly selected domestic smart meters were used in this study. The number of smart meters was selected in accordance with the generic urban model presented in [166], where 96 domestic customers were connected to an LV feeder. One week was used as a training period, while the other two weeks were a test period. An aggregated daily load profile was created through summing the measurements of the 96 smart meter at each half hour time step. Equation (3.1) illustrates the aggregation of smart meter measurements at the 11kV MV level.

$$\mathbf{lp}_{\text{agg. daily}} = \left(\left(\sum_{i=1}^{96} \mathrm{lp}_i(t) \right)_{t=1}, \left(\sum_{i=1}^{96} \mathrm{lp}_i(t) \right)_{t=2}, \dots, \left(\sum_{i=1}^{96} \mathrm{lp}_i(t) \right)_{t=48} \right) \quad \textbf{(3.1)}$$

given $\mathbf{lp}_{\text{agg. daily}}$ is the aggregated daily load profile, $\mathrm{lp}_i(t)$ is the measurement of the i^{th} smart meter at the t^{th} half hour, i is the smart meter index, i = 1, 2, 3, …, 96 and t is the half-hour index, t = 1, 2, 3, …, 48.

3.4 k-means Cluster Analysis Method

The k-means [58], [78], [155] is a classic partitional cluster analysis method that iteratively groups n load profiles, each comprised of T half-hourly measurements, into k clusters, by minimising the intra cluster sum of squared distances between the load profiles and cluster centres. Equation (3.2) illustrates the objective function of the k-means method

$$\min \sum_{j=1}^{k} \sum_{i=1, i \in j}^{n} \left\| \mathbf{lp}_i - \mathbf{cc}_j \right\|^2 \quad \textbf{(3.2)}$$

$\mathbf{lp}_i$: is the i^{th} load profile, $i = 1, 2, 3, \dots, n$, and $\mathbf{cc}_j$: is the j^{th} cluster centre, $j = 1, 2, 3, \dots, k$. The i^{th} load profile is described as $\mathbf{lp_i} = (\mathrm{lp}_i(t), t = 1, 2, 3, \dots, T)$. Similarly, the j^{th}

cluster centre is defined as $\mathbf{cc_j} = \left(\mathrm{cc}_j\,(t), t = 1, 2, 3, \ldots, T\right)$. The centre of a cluster is determined in terms of the mean values of all load profiles that are assigned to this specific cluster calculated at each half-hourly time step. Equation (3.3) defines the centre of a cluster.

$$\mathbf{cc}_{j,final} = \left(\left(\frac{\sum_{i=1}^{m} \mathrm{lp}_i(t)}{m}\right)_{t=1}, \left(\frac{\sum_{i=1}^{m} \mathrm{lp}_i(t)}{m}\right)_{t=2}, \ldots, \left(\frac{\sum_{i=1}^{m} \mathrm{lp}_i(t)}{m}\right)_{t=T}\right) \quad \textbf{(3.3)}$$

The inputs of the k-means clustering algorithm that was developed in this research include the domestic load profiles and the maximum number of clusters. The inputs also include the randomisation number (rnd_m) which is a variable that overcomes the impact of the initial random selection of cluster centres upon the outputs. For a number of (k) clusters, the randomisation number runs the k-means method rnd_m different times; each time with a different set of initial cluster centres. The best results, those with the smallest intra cluster distances, are produced as outputs.

At each iteration of the k-means, the Average Euclidean Distance (AvED) is calculated between the input load profiles and their cluster centres according to Equation (3.4). As a result, each load profile is assigned to the cluster that has the nearest centre.

$$AvED\left(\mathbf{lp}_i, \mathbf{cc}_j\right) = \sqrt{\frac{\sum_{t=1}^{T}\left(\mathrm{lp}_i(t) - \mathrm{cc}_j(t)\right)^2}{T}} \quad \textbf{(3.4)}$$

The k-means algorithm increments the number of clusters until the mean value of AvED falls below a threshold that was equal to 1-10% of the mean active power consumption of the training period. Equations (3.5) and (3.6) define the criterion used to calculate the required number of clusters.

$$Mean\ AvED = \underset{n}{\mathrm{mean}}\left(AvED(\mathbf{lp}, \mathbf{cc})\right) = \frac{\sum_{i=1}^{n} AvED\left(\mathbf{lp}_{i,i\in j}, \mathbf{cc}_j\right)}{n} \quad \textbf{(3.5)}$$

$$AvED\ threshold = \frac{x}{100} \times \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} \mathrm{lp}_i(t)}{n \times T} \quad \textbf{(3.6)}$$

given x is a percentage of the mean active power consumption during the training period. Since the percentage of mean consumption, x, of the training period determines the AvED threshold which in turns defines the number of output clusters, x was changed between 1-10 to assess the impact of AvED threshold upon the compactness of the

resulting clusters. The outputs of the clustering algorithm include the number of clusters, cluster centres, and assignment of load profiles to their respective clusters.

In this research, the productivity ratio [169] of the clustering algorithm (i.e., Cluster productivity ratio), defined as the ratio between the numbers of output clusters to the numbers of input profiles, quantifies the compactness of the clusters. Equation (3.7) defines the cluster productivity ratio.

$$Cluster\ productivity\ ratio = \frac{Number\ of\ output\ clusters}{Number\ of\ input\ profiles} \quad \textbf{(3.7)}$$

Small values of the clustering productivity ratio indicate the presence of few output clusters, which is a standard feature of compact clustering, whereas a productivity ratio of unity means that each load profile is its own cluster.

A flow chart of the k-means cluster analysis algorithm is shown in Figure 3.4. Pycluster [167], an open source cluster analysis software, was used to develop the clustering module in Python 2.7.

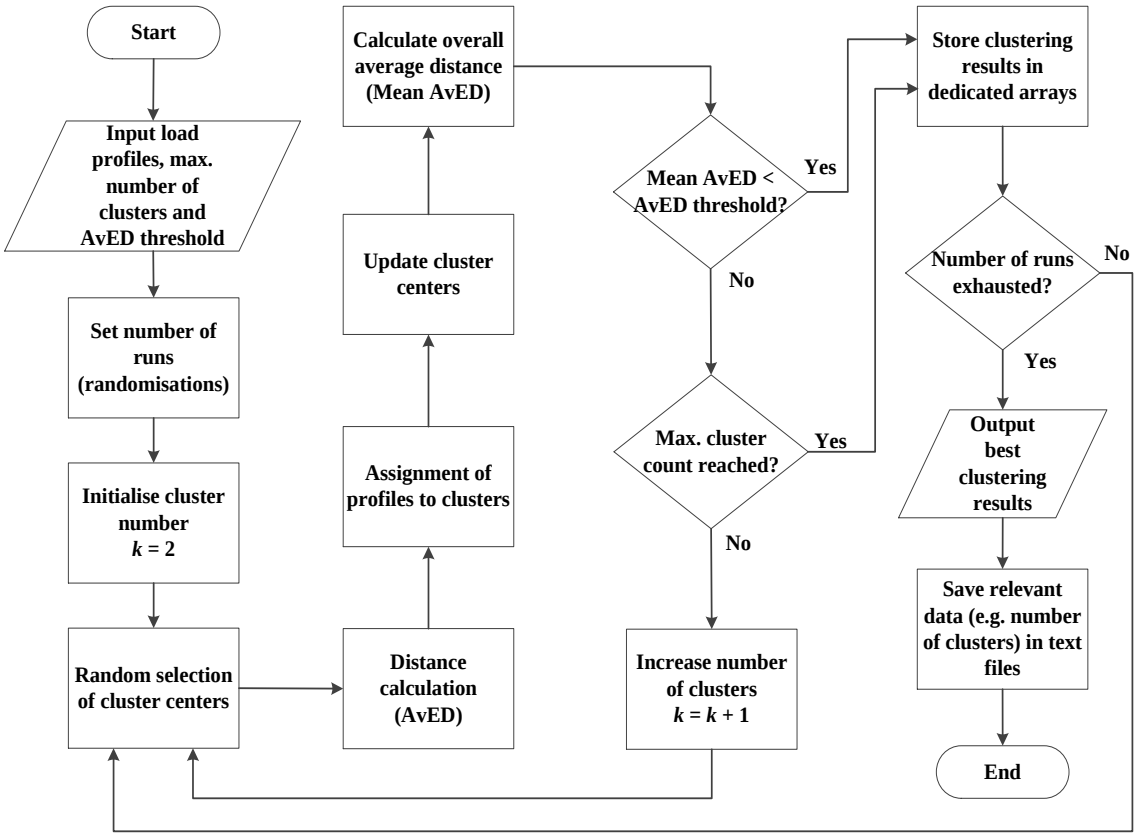
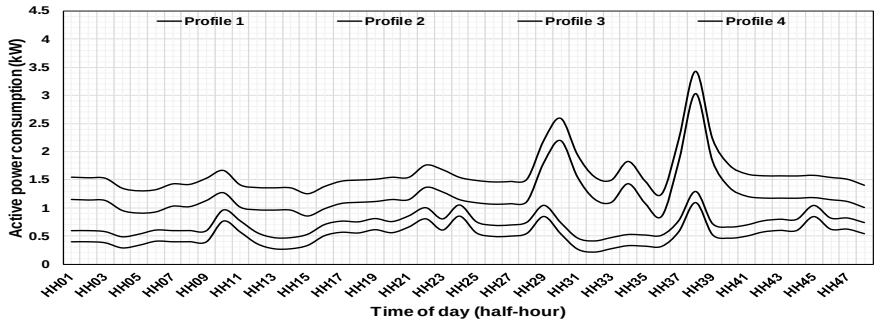
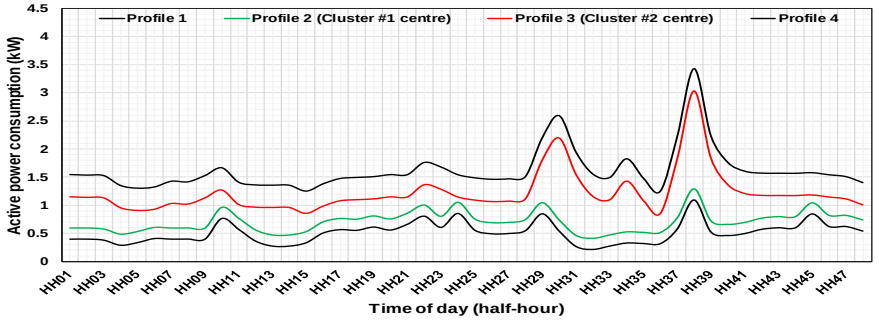


Figure 3.4 Flow chart of the developed k-means cluster analysis algorithm

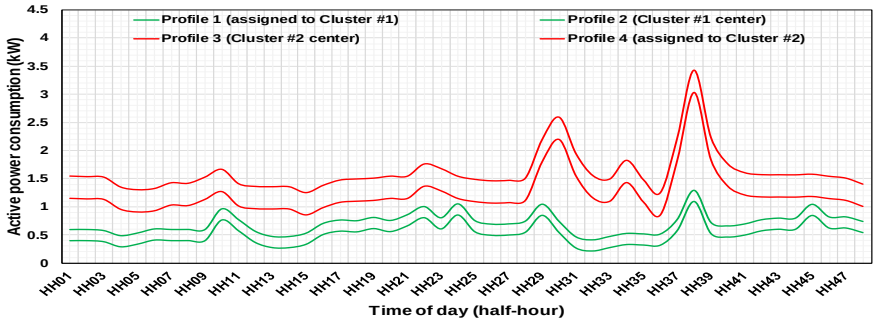
An example illustrating the basic principles of the developed k-means cluster analysis algorithm is shown in Figure 3.5.



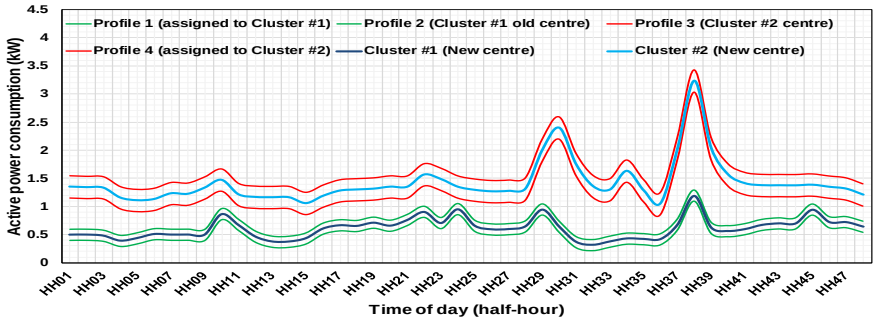
(a)



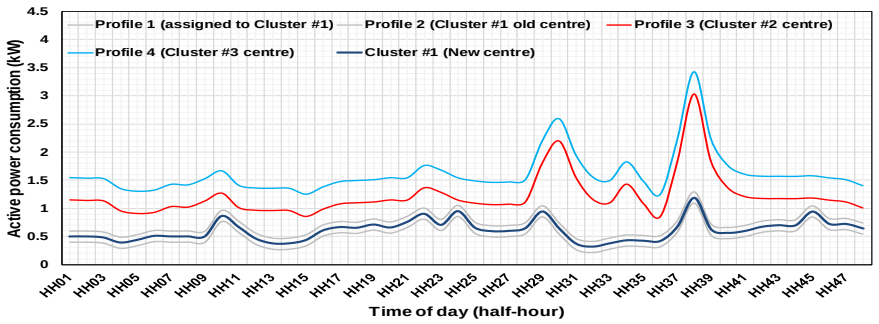
(b)



(c)



(d)



(e)

Figure 3.5 Example of *k*-means cluster analysis method
(a) load profiles, (b) initial random selection of cluster centres, (c) assignment of profiles to clusters, (d) updated cluster centres, and (e) final results

In Figure 3.5(a), the k-means clustering algorithm attempts to group 4 daily load profiles into a number of clusters that ranges between 2 (where the profiles are divided between two clusters) and 4 (where each profile is its own cluster). The mean active power consumption of the four profiles was 1.01471kW. Using equation (3.6) and choosing x = 10% (of the mean active power consumption), the AvED threshold was equal to 0.101471kW. Steps 1-5 provide a detailed step-by-step explanation of how the load profiles are grouped into a small number of clusters.

1. Clustering the load profiles starts with k = 2 clusters. Therefore, two of the four profiles are randomly selected to be the centres of the two clusters. Figure 3.5(b) shows that Profiles 2 (green) and 3 (red) are the randomly selected cluster centres in this case.

2. The average Euclidean distance is measured between the other two (black) profiles and the random (green and red) cluster centres according to equation (3.4). Each other profile is then assigned to the cluster that has the nearest centre. Figure 3.5(c) shows that Profile 1 was assigned to Cluster #1 characterised with the green colour, while Profile 4 was allocated to Cluster #2 designated with the red profile.

3. Once all load profiles are assigned to their nearest clusters, the new cluster centres calculated according to equation (3.3) and are updated accordingly. Figure 3.5(d) shows the new cluster centres in dark blue (Cluster #1) and light blue (Cluster #2).

4. The mean average Euclidean distance (Mean AvED) between the profiles and their corresponding cluster centres is calculated according to equation (3.5) and are compared to the AvED threshold. In this example, Mean AvED was equal to 0.15kW which is greater than 0.101471kW (AvED threshold). Therefore, the number of clusters, k, is incremented.

5. Using the new number of clusters, the k-means cluster analysis algorithm iterates through steps 1-4 until the mean AvED falls below AvED threshold. In the case of the given example, this is achieved when k was equal to 4. Figure 3.5(e) shows the final clusters centres and profiles assigned to each cluster.

3.5 Load Profile Clustering Methodology

Several case studies were implemented to investigate the performance of the developed algorithm using different sets of load profiles of individual and aggregated domestic smart meters.

3.5.1 Clustering of training period profiles

One week, between 14th July and the end of 26th July 2009, of load profiles was applied to train the k-means cluster analysis algorithm and obtain the cluster centres. The daily and segmented load profiles were separately clustered. Whereas a daily load profile consists of 48 half-hourly measurements, a segmented load profile extends over a time window that is less than or equal to 24 hours. The 2, 4, 6, …, 24 hours' time windows were used, on a rolling basis, to create the segmented load profiles. Figure 3.6 illustrates the concept of segmented load profiles.

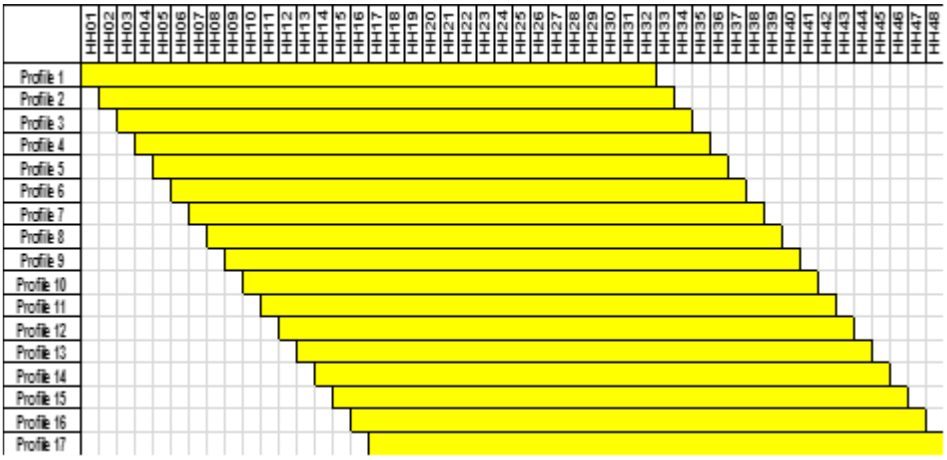


Figure 3.6 Segmented load profiles

Figure 3.6 shows that a daily load profile is segmented into 17 segments each with a time window equal to 16 hours. For any segmentation time window (r) in hours, provided that the segmented profiles are rolled one half-hourly step at a time, then the number of segmented profiles is determined according to Equation (3.8) [168]

$$Number\ of\ segmented\ profiles = (n \times T) - 2r + 1 \quad \textbf{(3.8)}$$

given n is the number of the daily load profiles, and T is the number of (half-hourly) measurements per daily load profile.

Segmented load profiles provide better presentation of the temporal power consumption both within the premises and at the MV/LV transformer. The main objectives of using segmented load profiles were to

1. increase the number of input profiles available for the k-means based cluster analysis algorithm, and
2. determine the best segmentation time window for classification of test load profiles and for load estimation of missing and future measurements **(Chapter 4)**.

Application of the k-means cluster analysis algorithm groups the training daily and segmented load profiles of the individual and aggregated smart meters into k different clusters. The output of this step is expressed in terms of number of clusters and their corresponding centres. The training load profiles (daily or segmented) and their corresponding cluster centres have the same duration expressed in hours (or alternatively, the number of half-hourly measurements).

3.5.2 Classification of test period profiles

The load profiles between 27th July and the end of 9th August 2009 were used to test the performance of the k-means clustering algorithm. The two weeks' test period was divided into 10 sub-periods, each consisting of three consecutive days. Table 3.1 illustrates the details of the sub-periods.

Table 3.1 Details of test sub-periods

Sub-period number	*Day included (Year 2009)*
1	27/07, 28/07, 29/07
2	28/07, 29/07, 30/07
3	29/07, 30/07, 31/07
4	30/07, 31/07, 01/08
5	31/07, 01/08, 02/08
6	03/08, 04/08, 05/08
7	04/08, 05/08, 06/08
8	05/08, 06/08, 07/08
9	06/08, 07/08, 08/08
10	07/08, 08/08, 09/08

The load profiles of the test period were allocated to the nearest cluster centres obtained from the training period. The AvED, equation (3.4), was applied to calculate the distances between the test profiles and training cluster centres. The Mean AvED and the maximum value of the maximum absolute difference (Max.Max.AD) were used to quantify the differences between the test profiles and their corresponding training cluster centres. Equation (3.9) illustrates the Mean AvED, that was presented earlier as equation (3.5) but has been re-written to reflect the test period profiles, whereas the overall maximum absolute difference (Max. Max. AD) is demonstrated in equation (3.10).

$$Mean\ AvED\ = \underset{L}{\text{mean}}\big(AvED(\mathbf{lp}, \mathbf{cc})\big) = \frac{\sum_{l=1}^{L} AvED\ \left(\mathbf{lp}_{l,l\in j}, \mathbf{cc}_j\right)}{L} \quad \textbf{(3.9)}$$

$$Max.Max.AD\ = \max_{L}\left(\left(\max_{T}\left(\left|\mathrm{lp}_{l,l\in j}\,(t) - \ \mathrm{cc}_j(t)\right|\right)_{t\,\in\,T}\right)\right) \quad \textbf{(3.10)}$$

given l is the l^{th} load profile of the test period, $l\ =\ 1, 2, \dots, L$.

3.6 Results and Discussion

3.6.1 Clustering and classification of daily load profiles of individual smart meters

The results of clustering 672 daily load profiles of individual smart meters (i.e., 96 smart meters × 7 daily load profiles per smart meter) confirm the dynamic and unpredictable behaviour of the individual domestic demand. The increase in the AvED threshold, from 1 to 10% of the mean consumption of the training period, results in marginal reduction in the number of output clusters per single smart meter. Figure 3.7 shows the cluster productivity ratio of the 96 smart meters for different percentages of AvED thresholds.

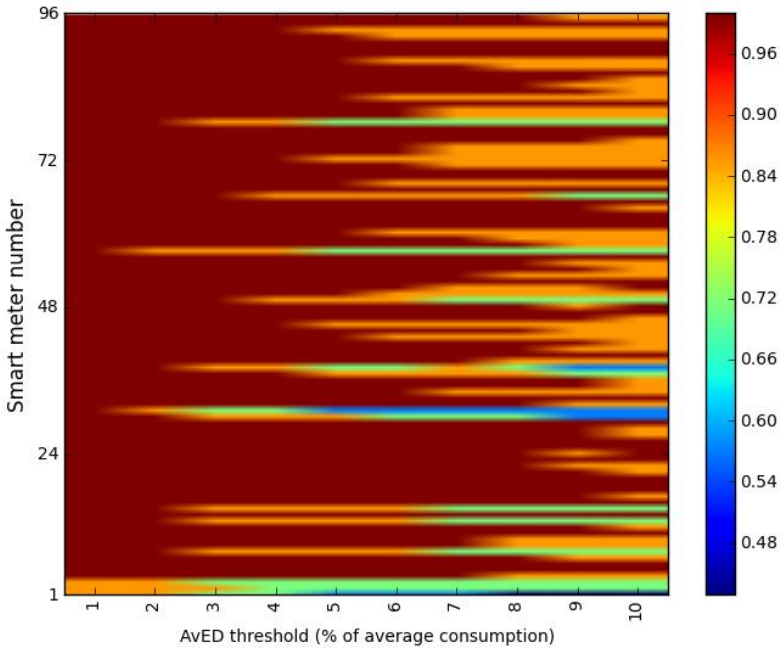


Figure 3.7 Cluster productivity ratio for different AvED thresholds – Daily profiles of Individual smart meters

In Figure 3.7, the brown colour which starts at an AvED threshold of 1% for the majority of smart meters indicates a productivity ratio of unity. This implies that the number of output clusters was equal to the number of input profiles. As the AvED threshold increases from 1-10% of the mean consumption, the majority of smart meters retained high values of the cluster productivity ratio. Detailed analysis of Figure 3.7 reveals that increasing the AvED threshold from 1-10% was found to halve the productivity ratio for one smart meter only.

The classification differences of allocating the test period daily load profiles (of each of the 96 smart meters) to their nearest training cluster centres were calculated according to equations (3.9) and (3.10). Figure 3.8 shows a box-whisker plot representation of the distribution of overall maximum absolute differences and the mean average Euclidean difference for different values of Mean AvED threshold.

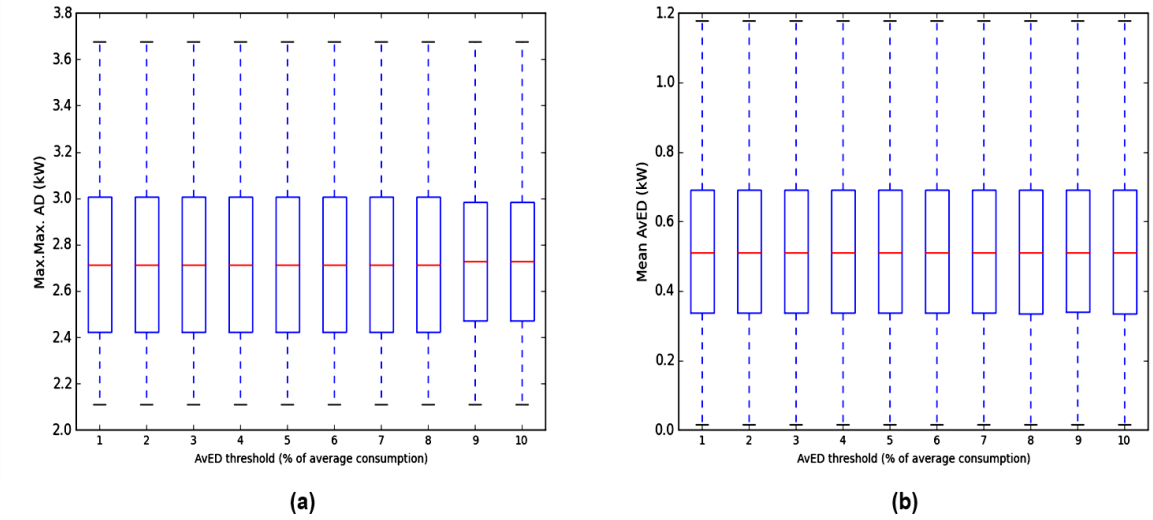


Figure 3.8 Distribution of classification differences – Daily profiles of Individual smart meters
(a) Overall maximum absolute difference, and (b) Mean Average Euclidean Difference

Figure 3.8 shows that (a) the overall maximum absolute differences and (b) the mean average Euclidean differences attained the same values (and hence distributions) for AvED threshold ranging between 1-10% of the mean consumption of the training period. The shapes of these distributions were similar because the number and structure of clusters were not impacted by the increase in AvED threshold.

The red horizontal lines in Figure 3.8 represent (a) the median of Max.Max.AD that was equal to 2.7kW and (b) the median of Mean AvED that was about 0.511kW. The median of Max.Max.AD is approximately equal to 440% while the median of Mean AvED is approximately 84%, both of the mean consumption of all individual smart meters during the training period which was approximately 0.611kW [6]. The boxes below the red lines are the values of (a) Max.Max.AD and (b) mean AvED between the first quartile and the median of the respective difference, whereas those boxes over the red lines are the (a) Max.Max.AD and (b) mean AvED values between the median and the third quartile of the corresponding difference. The whiskers are the minimum and maximum values of the overall maximum absolute and mean average Euclidean differences. Appendix (A.1) illustrates the details of a box-whisker plot.

[6] The mean consumption of all individual smart meters during the test period was approximately 0.620kW

3.6.2 Clustering and classification of segmented load profiles of individual smart meters

The results of clustering the segmented load profiles of 96 individual smart meters reveal that the smallest value of the cluster productivity ratio was obtained at the (shortest) 2-hour segmentation time window. Figure 3.9 demonstrates the cluster productivity ratio for AvED thresholds ranging between 1-10% of the mean consumption of the training period profiles.

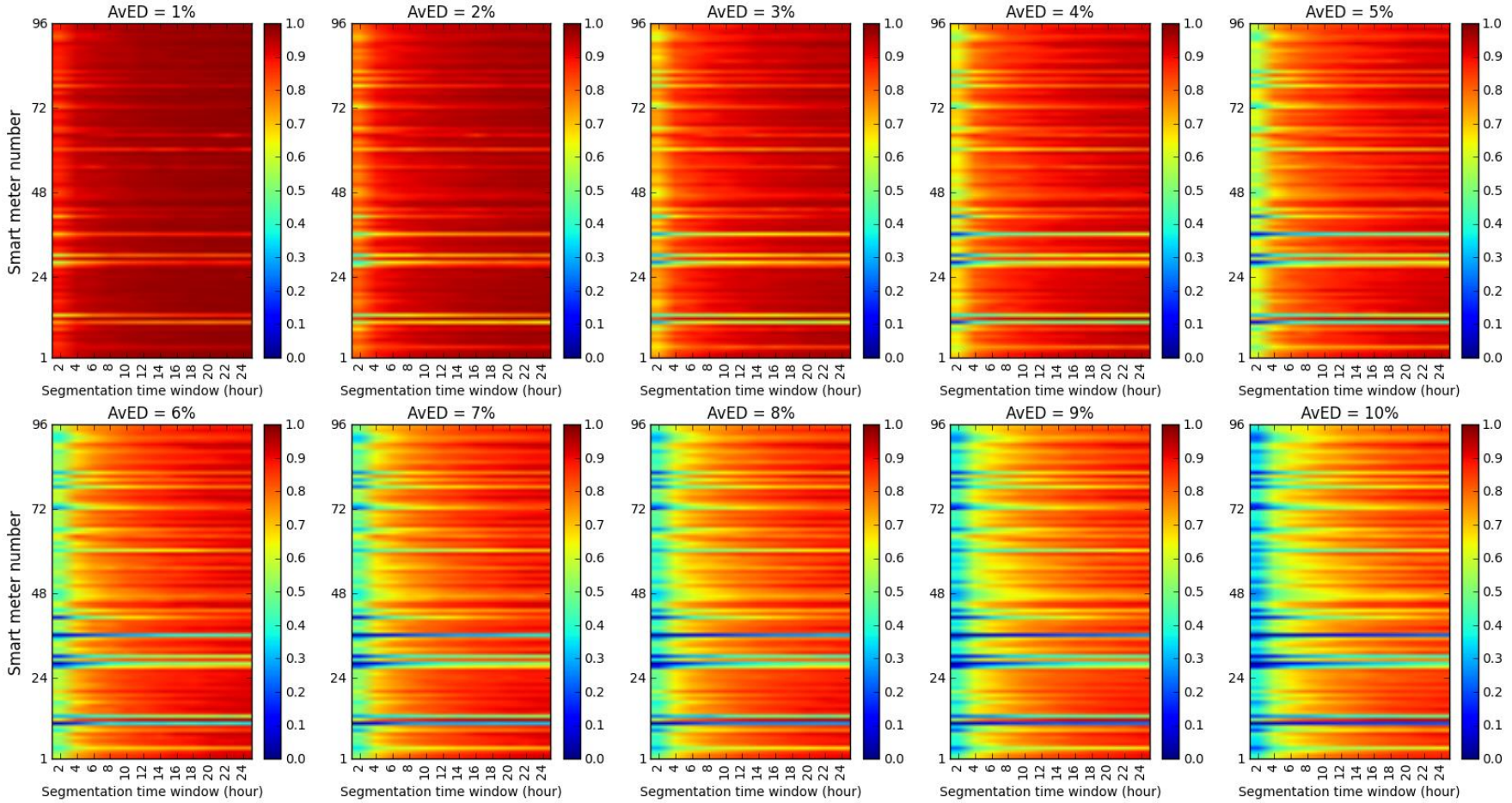


Figure 3.9 Cluster productivity ratio for different AvED thresholds – Segmented load profiles of individual smart meters

Figure 3.9 shows that that for each value of the AvED threshold, the smallest productivity ratio was observed at the shortest time window. The productivity ratio increases until the largest values were obtained at the (longest) 24-hour segmentation time window. Figure 3.10 shows the cluster productivity ratio for the even segmentation time windows between 2-24 hours.

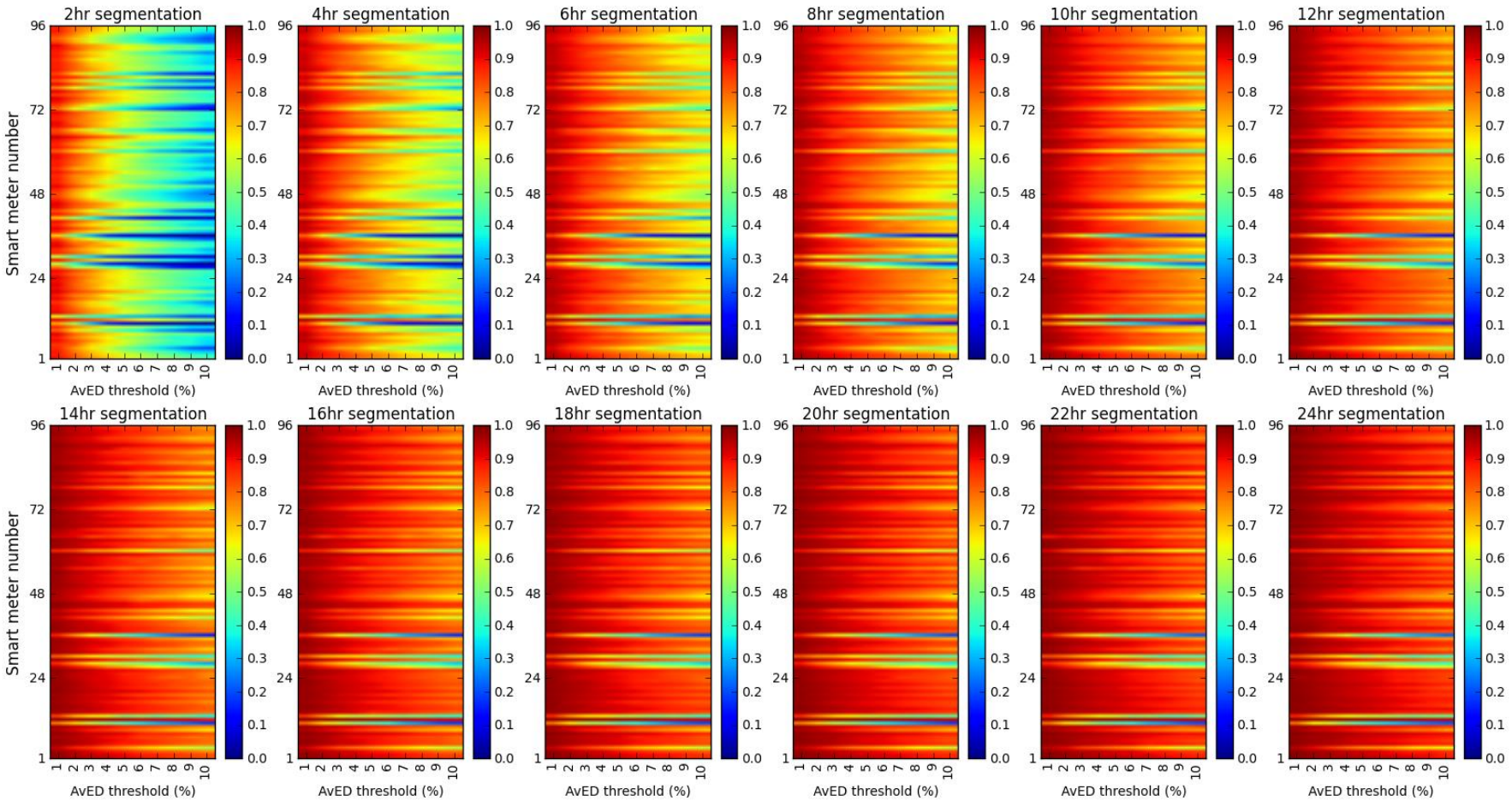


Figure 3.10 Cluster productivity ratio for different segmentation time windows – Segmented load profiles of individual smart meters

Figure 3.10 indicates that the cluster productivity ratio decreased with the increase in the AvED threshold for each segmentation time window. The largest productivity ratio was observed at an AvED threshold of 1%, whereas an AvED threshold of 10% of the mean power consumption during the training period resulted in the smallest productivity ratio. Figure 3.10 confirms that for long (14-24 hours) segmentation time windows, the productivity ratio had significantly higher values than at the short (2-6 hours) time windows.

Figure 3.11 shows the distribution of the overall maximum absolute difference between the segmented test profiles of 96 individual smart meters and their nearest training cluster centres for different values of AvED threshold. The differences between the segmented profiles of the individual smart meters during test period and their nearest training cluster centres shows that the smallest values were obtained at the 2-hour segmentation time windows. The distribution of the Mean AvED between the segmented profiles of the individual smart meters of the test period and the cluster centres of the training period are illustrated in Figure 3.12. The median of the overall maximum absolute difference was approximately between 1-2kW, whereas the median of the Average Euclidean distance between the test profiles and the training centres attained values in the range of 0.15-0.4kW.

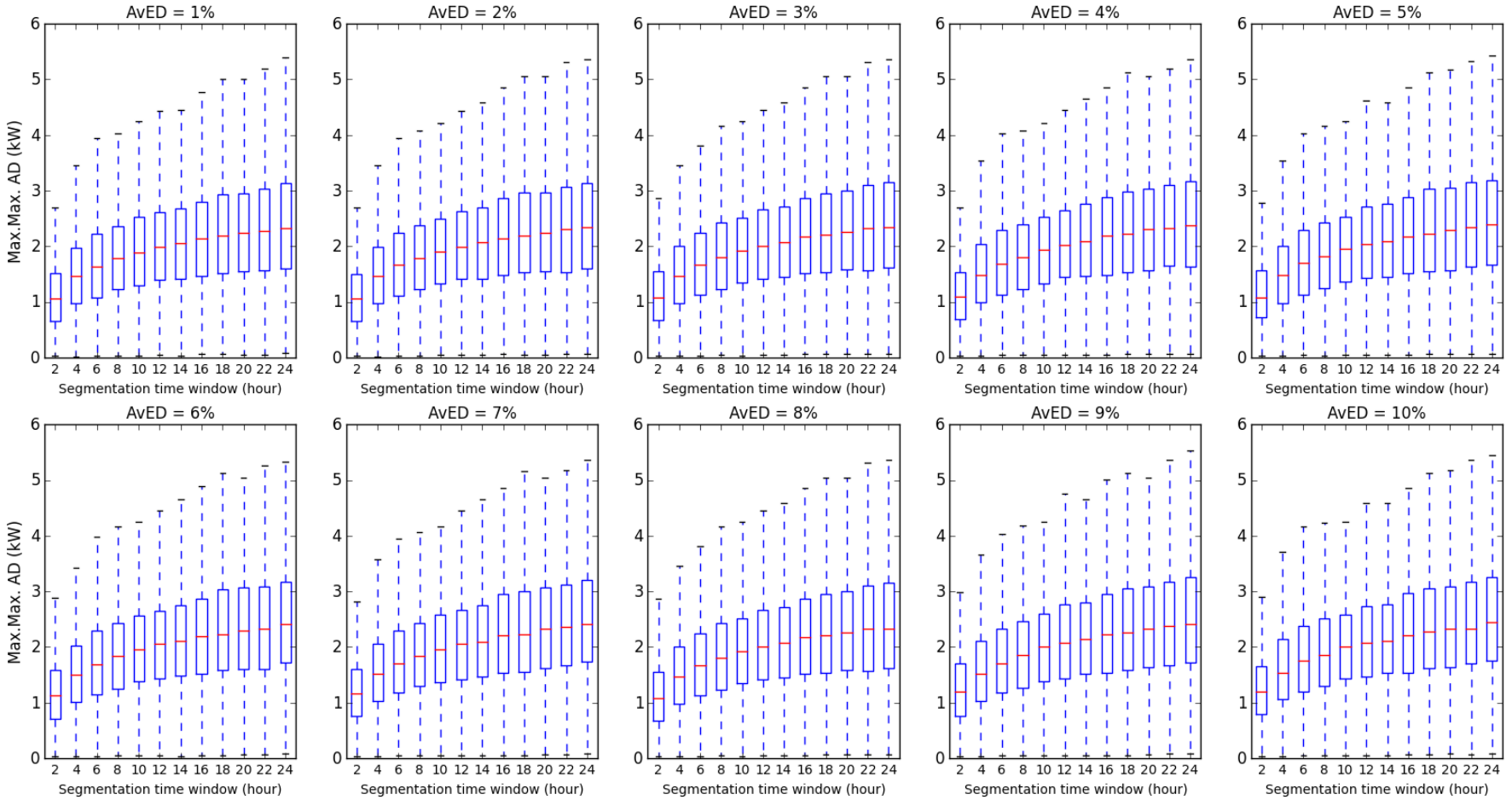


Figure 3.11 Classification of segmented profiles of individual smart meters - Max.Max.AD distributions

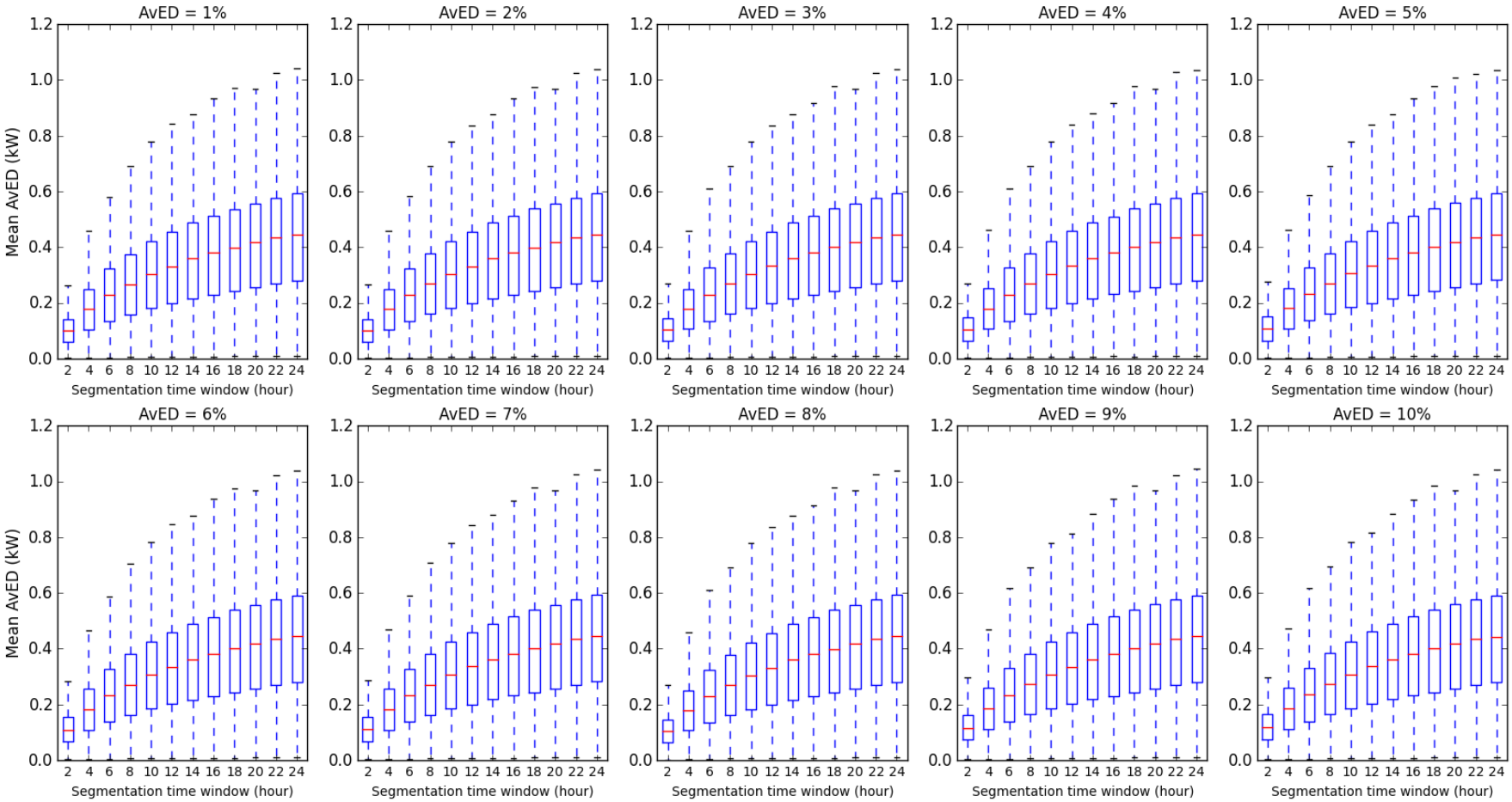


Figure 3.12 Classification of segmented profiles of individual smart meters – Mean AvED distributions

3.6.3 Clustering and classification of daily load profiles of aggregated smart meters

Results of clustering the seven aggregated daily load profiles (at the MV level) of 96 smart meters indicate that increasing the AvED threshold from 1 to 10% of the mean consumption of the training period reduces the number of output clusters from seven clusters to two clusters. Figure 3.13 shows the cluster productivity ratio of the aggregated measurements for different percentages of AvED thresholds.

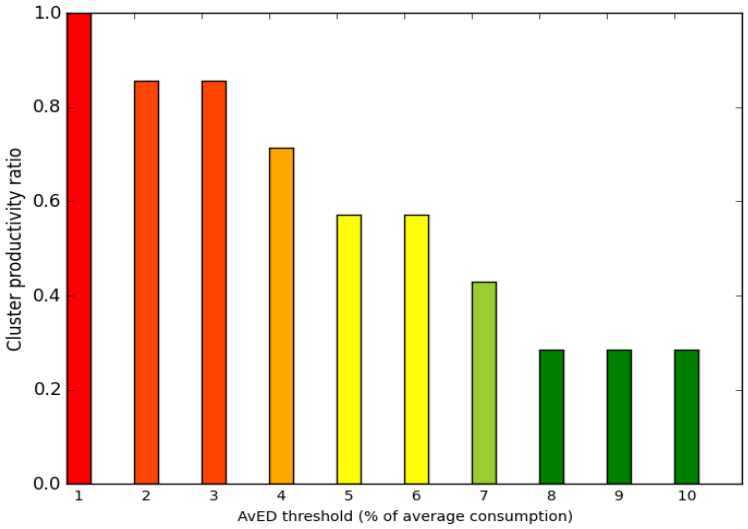


Figure 3.13 Cluster productivity ratio for daily profiles of aggregated smart meters

Figure 3.13 shows that an AvED threshold of 8-10% of the mean consumption of the training period grouped the aggregated daily profiles into two groups. The first group was the weekdays, Monday to Friday, whereas the second included weekend, Saturday and Sunday. Figures 3.14 and 3.15 show the patterns of these two clusters.

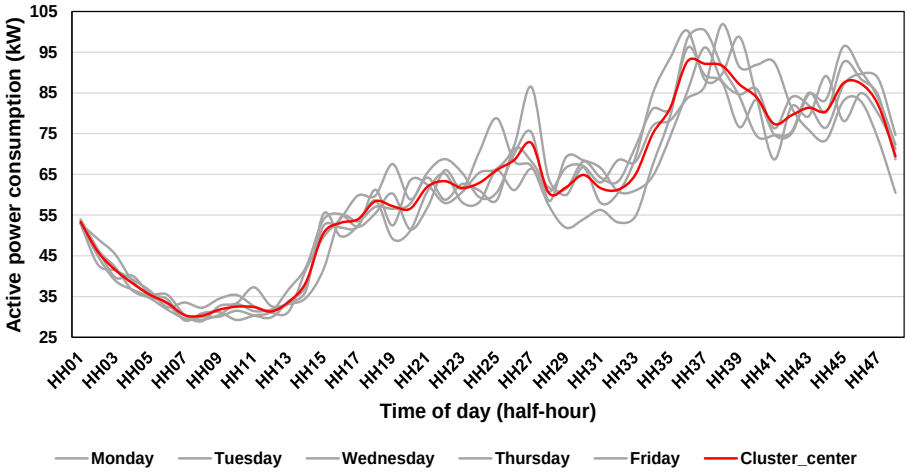


Figure 3.14 Weekdays cluster of aggregated daily profiles

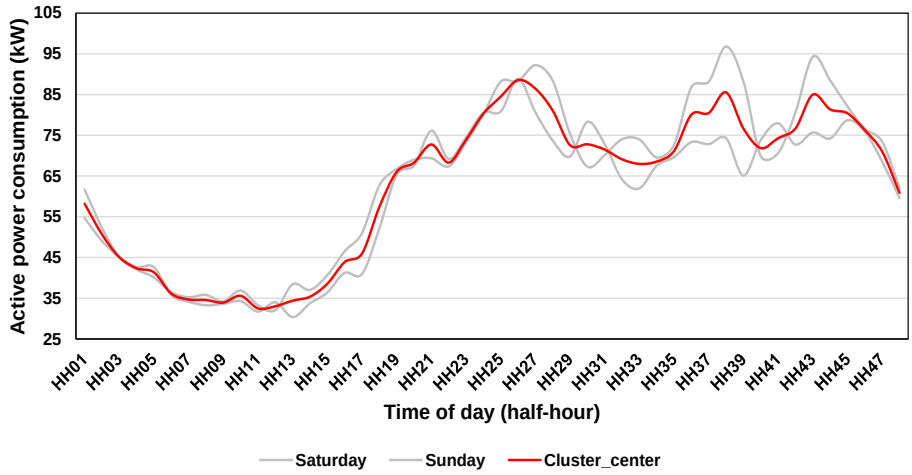


Figure 3.15 Weekend cluster of aggregated daily profiles

The overall maximum absolute differences between the aggregated daily profiles of the test period and their nearest training cluster centres exhibit different distributions for various values of AvED threshold as shown in Figure 3.16.

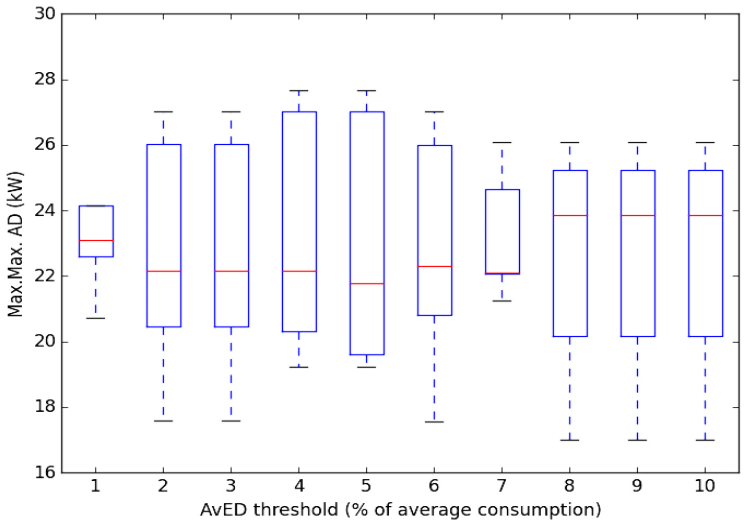


Figure 3.16 Classification of daily profiles of aggregated smart meters – Max. Max. AD distributions

For AvED threshold in the range of 1-7% (of the mean consumption), Figure 3.16 shows that the median of the maximum absolute differences was closer to the first quartile than the third quartile. At the same time, an 8-10% range of AvED threshold resulted in the median value of Max.Max.AD being closer to the third quartile.

The distribution of the mean value of the Average Euclidean Distance between the aggregated daily profiles of the test profiles and their nearest training cluster centres is shown in Figure 3.17. The distribution of Mean AvED attained the smallest values at the 8-10% range of AvED thresholds when the cluster productivity ratio was the smallest.

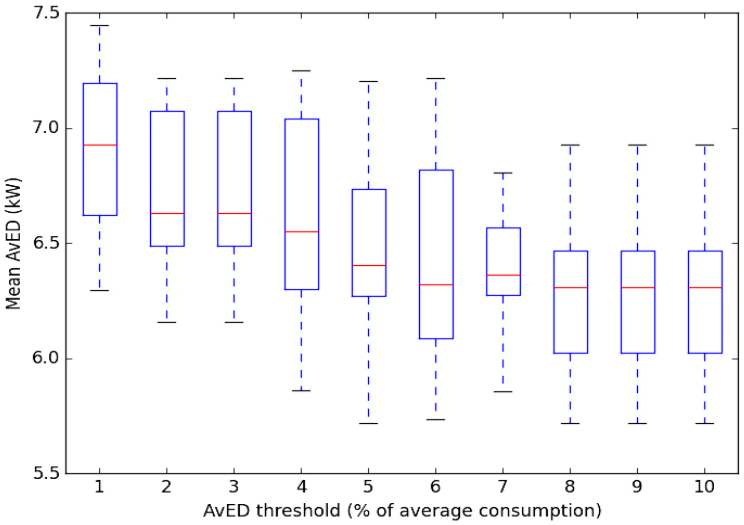


Figure 3.17 Classification of daily profiles of aggregated smart meters – Mean AvED distributions

Based on the results of allocating the aggregated daily load profiles of the test period, a conclusion was made that higher values of AvED threshold are more desired when the daily load profiles are highly correlated. High AvED thresholds lead to a smaller number of clusters which in turn, results in more accurate allocation of the test profiles to their nearest training clusters.

3.6.4 Clustering and classification of segmented load profiles of aggregated smart meters

Results of clustering the segmented load profiles of aggregated (96 individual) smart meters reveal that the smallest value of the cluster productivity ratio was obtained at the (shortest) 2-hour segmentation time window. Figure 3.18 shows the cluster productivity ratio for different segmentation time windows and different AvED thresholds.

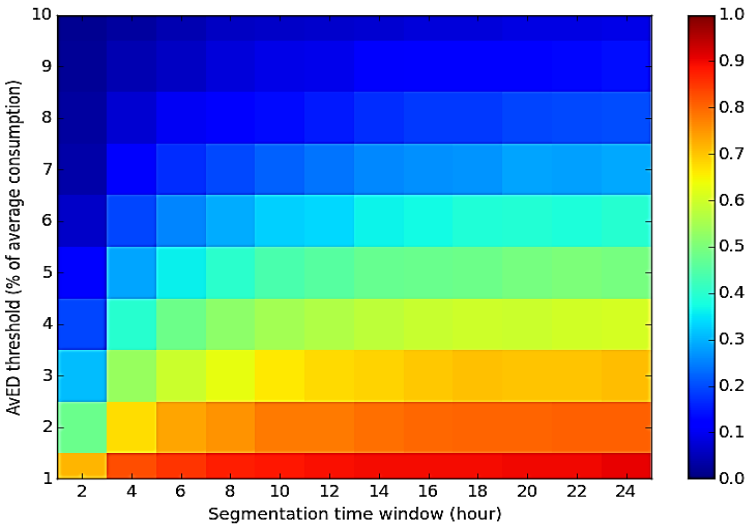


Figure 3.18 Cluster productivity ratio for different segmentation time windows and AvED thresholds – Aggregated smart meters

Figure 3.18 shows that for different values of AvED threshold, increasing the segmentation time window from 2-24 hours results in a corresponding increase in the cluster productivity ratio. Additionally, Figure 3.18 shows that the minimum cluster productivity ratio was obtained at an AvED threshold of 10% of the mean consumption of the training period.

Results of allocating the segmented test period profiles to the nearest training cluster centres reveals that the smallest differences between the test profiles and training cluster centres were observed at the shortest (2-hours) segmentation time window. Figure 3.19 shows the distribution of the overall maximum absolute difference for different values of AvED threshold applied to obtain the training cluster centres.

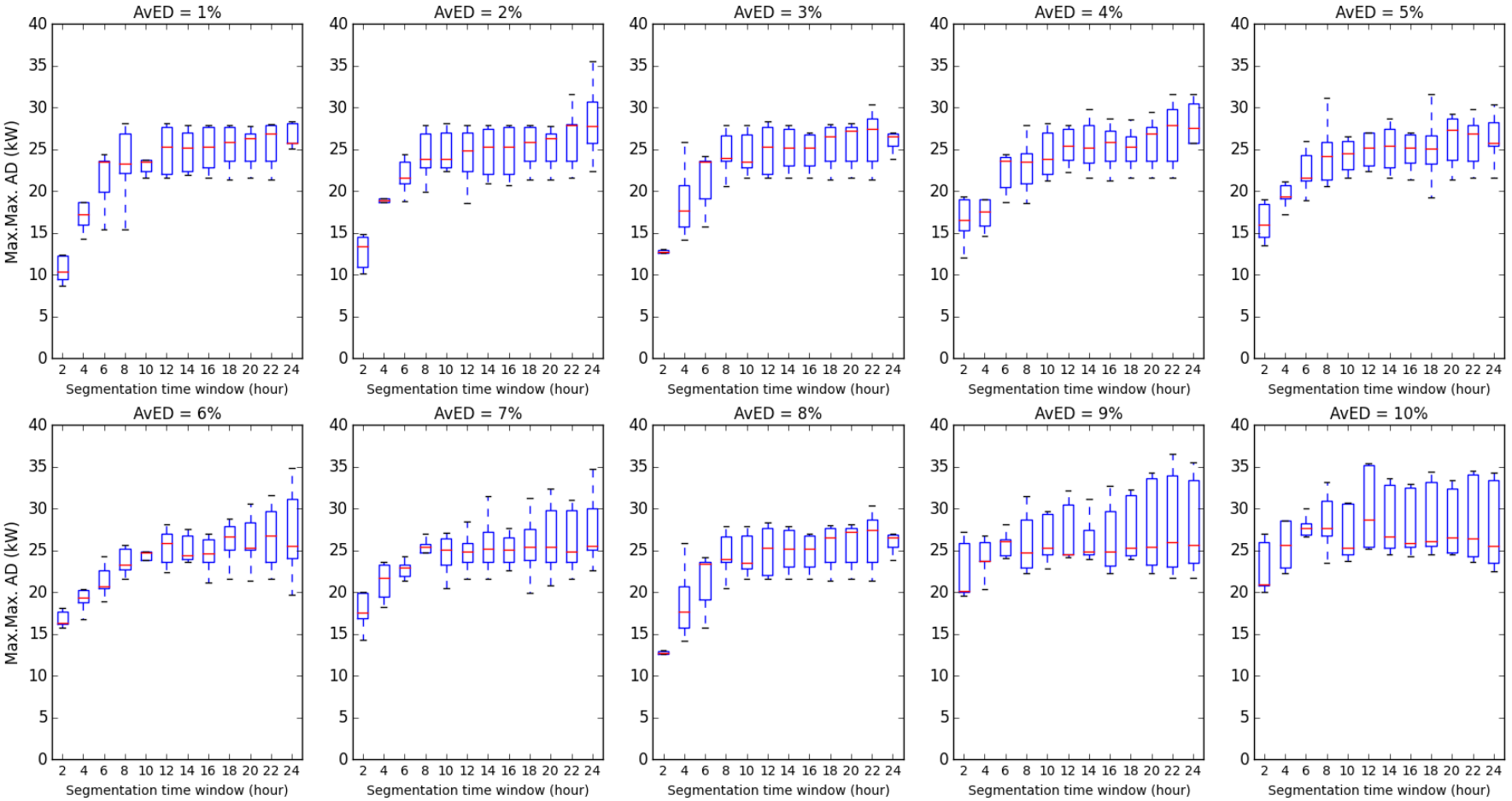


Figure 3.19 Classification of segmented profiles of aggregated smart meters – Max. Max. AD distribution

Figure 3.19 indicates that the distribution of the overall maximum absolute difference, between the segmented profiles of the test period and the cluster centres of the training period, had the smallest values at the shortest segmentation time windows. Figure 3.20 illustrates the distribution of mean AvED for allocating segmented test period profiles for different values of AvED threshold (that were applied to obtain the training cluster centres).

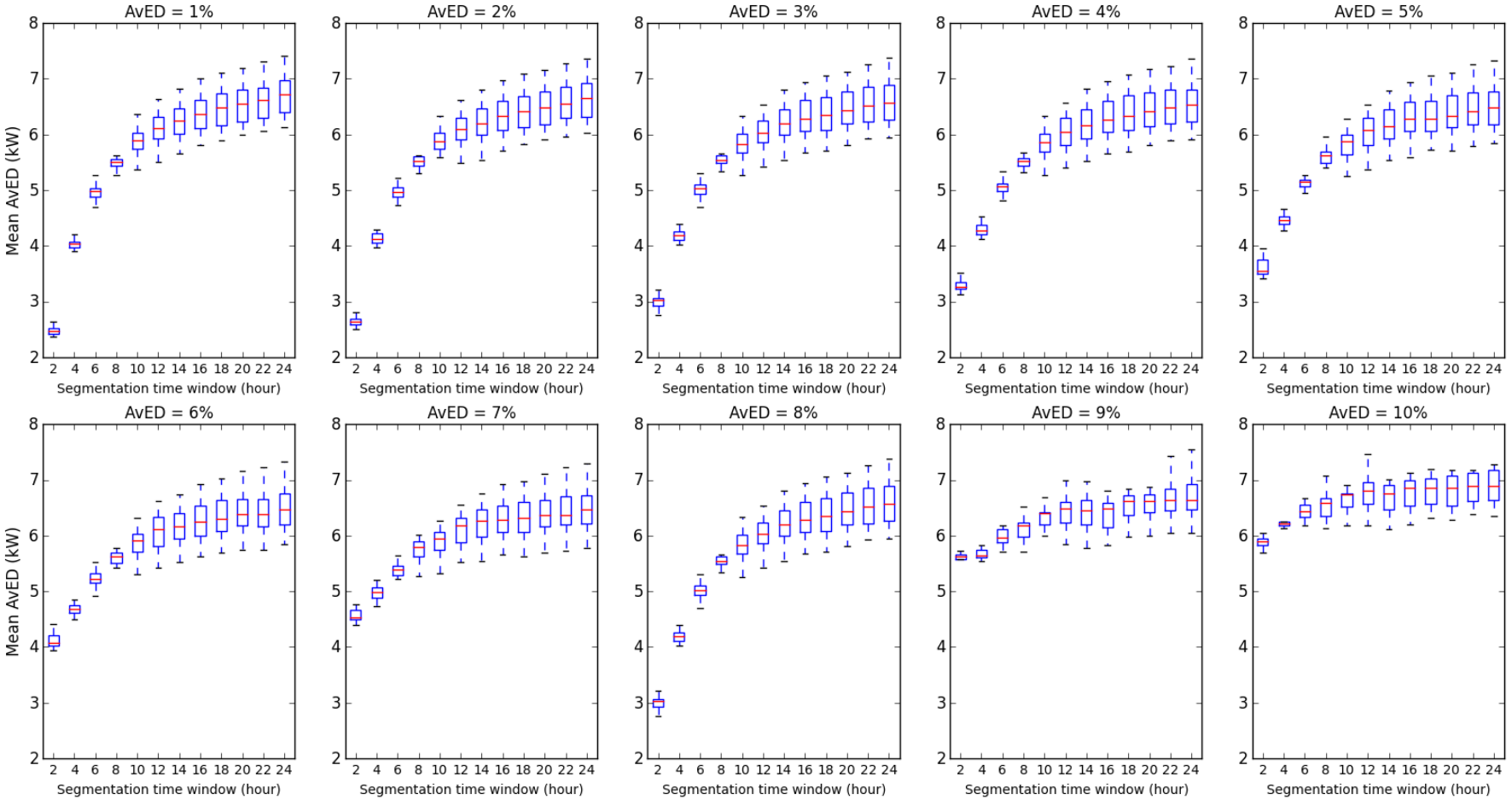


Figure 3.20 Classification of segmented profiles of aggregated smart meters – Mean AvED distributions

Figure 3.20 indicates that high values of the Mean AvED were obtained as a result of increasing the AvED threshold from 1-10% of the mean consumption of the training period. This means that a smaller number of clusters, and hence cluster centres, results in higher differences (or errors) when the test profiles were allocated to their nearest cluster centres.

3.7 Conclusions

In this chapter, a clustering algorithm based on the k-means cluster analysis method was developed and used to extract characteristic load profiles from smart meter measurements from measurements of individual and aggregated residential smart meters. Cluster centres are the characteristic load profiles that were extracted.

Performance of the clustering algorithm was investigated using daily and segmented load profiles. Training and test load profiles obtained from the Irish smart metering trials were used in this chapter.

The training load profiles were explicitly clustered using the cluster analysis algorithm. A criterion based on the mean AvED and the AvED threshold was applied to determine the required number of clusters. The mean AvED was calculated between the load profiles and their corresponding cluster centres and the AvED threshold was defined as 1-10% of the mean consumption of the training load profiles. To measure compactness of the clusters, the cluster productivity ratio defined the ratio of the number of output clusters to the number of input profiles.

The test load profiles were allocated to their nearest cluster centres (that were obtained as a result of grouping the training load profiles). The overall maximum absolute difference and mean AvED were used to quantify the differences between test load profiles and training cluster centres.

The results showed that for the majority of individual smart meters, the numbers of output clusters and input load profiles were approximately the same. A small reduction in the cluster productivity ratio was observed despite relaxing the AvED threshold from 1 to 10% of the mean consumption. This implies that clustering the load profiles at the LV level was ineffective, which can be referred to the dynamic and irregular power consumption patterns of individual domestic customers.

The results also revealed that clustering the load profiles of aggregated smart meters into small numbers of clusters (e.g. weekdays and weekends) was successfully

achieved. The diversity in appliance usage by different customers renders the uncertainty and abrupt changes in the aggregated load profile very low.

Clustering the segmented load profiles of individual and aggregated smart meters showed that the smallest cluster productivity ratio was obtained at the 2-hour segmentation time window. This indicates that the segmented profiles exhibit similar patterns at the shortest segmentation time window. The similarity decreases as the segmentation time window increases.

The smallest differences of allocating the segmented profiles of individual and aggregated smart meters, to their nearest clusters, were observed at the 2-hour segmentation time window. This confirms that the segmented profiles exhibit the highest similarity at the shortest segmentation time window.

Comprehensive analysis of the results showed that the smallest classification differences were observed when an AvED threshold of 1% (of the mean consumption) was applied to obtain the training cluster centres. A conclusion was made to use the 1% AvED threshold to produce the cluster centres (that are applied to estimate the missing measurements in **Chapter 4**).

Chapter 4 Load Estimation of Smart Meter Measurements

4.1 Introduction

A load estimation algorithm based on the k-means cluster analysis method was developed to estimate missing and future measurements of smart meters. The load estimation algorithm applied both the daily (consisting of 48 half-hourly measurements) and segmented (with 4, 8, 12, …, 48 half-hourly measurements) cluster centres, that were obtained from **Chapter 3**, and different distance functions to estimate the measurements of individual and aggregated smart meter measurements. Segmented cluster centres are the results of clustering the segmented load profiles.

The load estimation algorithm investigated the impacts of daily and segmented cluster centres and segmentation time windows upon the accuracy of the estimated measurements.

The application of Canberra, Manhattan, Euclidean and Pearson correlation distance functions in order to provide high accuracy load estimates was investigated.

Durations of measurement loss ranging from 1-24 consecutive hours () were simulated. The missing measurements were estimated iteratively, i.e. only one half-hourly measurement was estimated at a time.

The estimation errors between the actual and estimated smart meter measurements were quantified using the overall maximum absolute error and the mean value of root-mean square error.

4.2 k-means based load estimation algorithm

A load estimation algorithm based on the k-means cluster analysis method and principles of pattern recognition was developed to estimate missing and future smart meter measurements. A block diagram representation of the developed load estimation algorithm is shown in Figure 4.1.

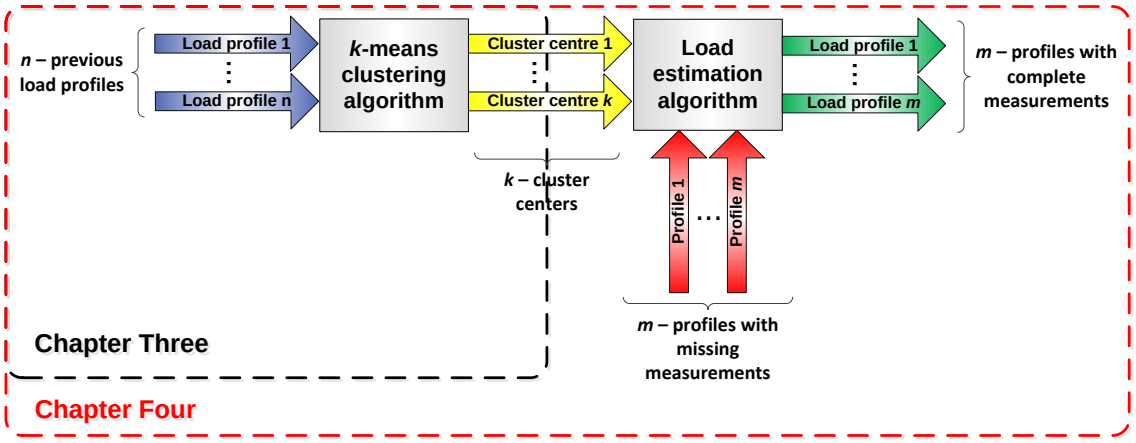


Figure 4.1 Block diagram of the developed load estimation algorithm

The load estimation algorithm initially groups similar historical load profiles, using the k-means cluster analysis algorithm developed in **Chapter 3**, and applies the cluster centres to estimate any missing measurements of the smart meters. Distance functions and similarity (dissimilarity) measures, which are basic principles of pattern recognition, were applied to replace missing and estimate future smart meter measurements. In this manner, the load profiles that have missing measurements were compared against cluster centres using different distance functions and matched to the nearest cluster. As a result, the half-hourly measurements of cluster centres were used to replace the missing of smart meter measurements.

4.3 Load estimation methodology

The load estimation algorithm tested the application of two approaches to estimate the measurements of individual and aggregated smart meters. While daily cluster centres (of the 20th – 26th July 2009 training period) were applied in the first approach, segmented cluster centres were used in the second approach to estimate the measurements. For each day of the test period (27th July – 3rd August 2009), 24 scenarios of lost measurements were simulated using a brute-force approach. The scenarios consider different durations of lost measurements from 1 to 24 consecutive hours. The measurements were estimated iteratively, i.e. only one half-hourly measurement was estimated at a time.

4.3.1 Load estimation using daily cluster centres

Regardless of the number of measurements to be estimated, 48 half-hourly measurements (the measurement to be estimated plus 47 half-hourly measurements that precede it) were matched to the nearest daily cluster centre (obtained from clustering the daily load profiles of the training period). Figure 4.2 shows that the last half-hourly measurement of the nearest daily cluster centre iteratively replaces the missing measurement(s).

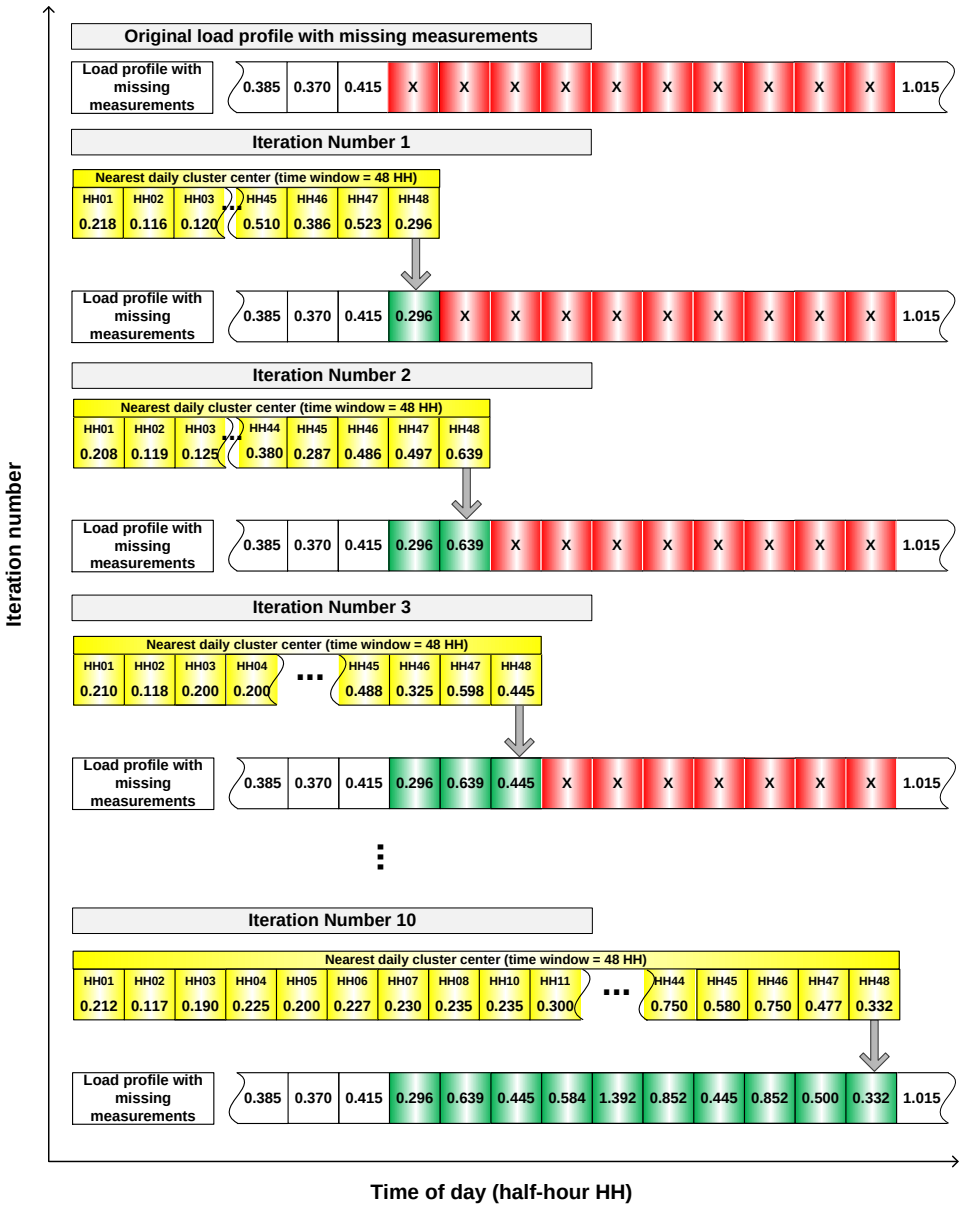


Figure 4.2 Load estimation using daily cluster centres

4.3.2 Load estimation using segmented cluster centres

Rather than the application of a time window of 24 hours, $2r$ half-hourly measurements (one measurement to be estimated plus the $2r - 1$ measurements that precede it) were paired to the nearest segmented cluster whose length is $2r$ half-hourly measurements (obtained from clustering the segmented load profiles of the training period). Figure 4.3 illustrates this approach. Segmented cluster centres with a segmentation time window of 2, 4, 6, …, 24 hours were used to estimate smart meter measurements, r is the segmentation time window (as presented in **Chapter 3**).

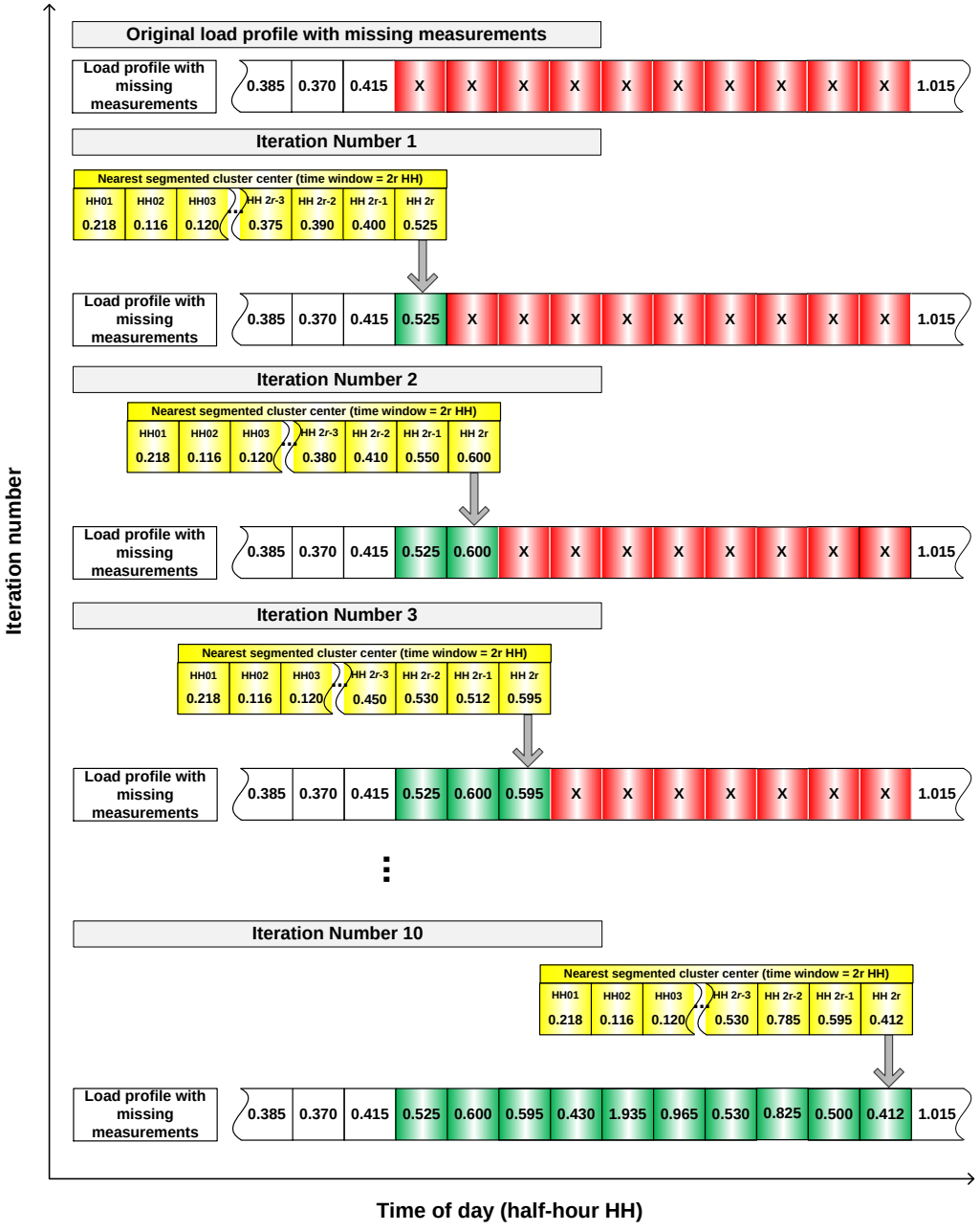


Figure 4.3 Load estimation using segmented cluster centres

4.3.3 Dissimilarity and distance measures

Dissimilarity and distance measures are essential in all applications of pattern recognition including clustering and classification. These measures quantify the similarity of two (or more) profiles. A dissimilarity measure $d(\mathbf{lp}_i, \mathbf{cc}_j)$ is defined to be a distance metric if it fulfils the following conditions:

a. $d(\mathbf{lp}_i, \mathbf{cc}_j) \geq 0$ for all $\mathbf{lp}_i$ and $\mathbf{cc}_j$,

b. $d(\mathbf{lp}_i, \mathbf{cc}_j) = 0$ if and only if $\mathbf{lp}_i = \mathbf{cc}_j$,

c. $d(\mathbf{lp}_i, \mathbf{cc}_j) = d(\mathbf{cc}_j, \mathbf{lp}_i)$

d. $d(\mathbf{lp}_i, \mathbf{cc}_j) \leq d(\mathbf{lp}_i, \mathbf{lp}_k) + d(\mathbf{lp}_k, \mathbf{cc}_j)$,

where $\mathbf{lp}_i$, and $\mathbf{lp}_k$ are any two load profiles and $\mathbf{cc}_j$ is the j^{th} cluster centre [78], [157], [170]. Table 4.1 demonstrates the distance functions that were investigated.

Table 4.1 Summary of distance functions

Measure	*Formula*
D1: Average Euclidean distance	$d(\mathbf{lp}_i, \mathbf{cc}_j) = \left[\frac{\sum_{t=1}^{T}(\mathrm{lp}_i(t) - \mathrm{cc}_j(t))^2}{T}\right]^{1/2}$
D2: Average Manhattan (city block) distance	$d(\mathbf{lp}_i, \mathbf{cc}_j) = \frac{\sum_{t=1}^{T}\lvert\mathrm{lp}_i(t) - \mathrm{cc}_j(t)\rvert}{T}$
D3: Average Canberra distance	$d(\mathbf{lp}_i, \mathbf{cc}_j) = \begin{cases} 0 & \text{for } \mathrm{lp}_i(t) = \mathrm{cc}_j(t) = 0 \\ \frac{\sum_{t=1}^{T}\frac{\lvert\mathrm{lp}_i(t) - \mathrm{cc}_j(t)\rvert}{\lvert\mathrm{lp}_i(t)\rvert + \lvert\mathrm{cc}_j(t)\rvert}}{T} & \text{for } \mathrm{lp}_i(t) \neq 0 \text{ or } \mathrm{cc}_j(t) \neq 0 \end{cases}$
D4: Average Pearson correlation distance	$d(\mathbf{lp}_i, \mathbf{cc}_j) = \frac{1 - \frac{\sum_{t=1}^{T}(\mathrm{lp}_i(t) - \bar{\mathbf{lp}}_i)(\mathrm{cc}_j(t) - \overline{\mathbf{cc}}_j)}{\left[\sum_{t=1}^{T}(\mathrm{lp}_i(t) - \bar{\mathbf{lp}}_i)^2 \sum_{t=1}^{T}(\mathrm{cc}_j(t) - \overline{\mathbf{cc}}_j)^2\right]^{1/2}}}{T}$ where $\bar{\mathbf{lp}}_i$ is the mean value of the i^{th} load profile and $\overline{\mathbf{cc}}_j$ is the mean value of the j^{th} cluster centre.

given T is the number of half-hourly measurements in a load profile and $t \in T$ (as presented in **Section 3.4**).

Euclidean distance (D1) is the most commonly used distance function [78], [155]. Manhattan (also called City block, taxicab and rectilinear) distance (D2) measures distances on a rectilinear basis. Canberra distance (D3) is a special case of the city block measure. The only differentiation between these two measures is that in Canberra distance, absolute differences between the k^{th} instances of the load profiles are divided by the sum of the absolute values of these instances prior to summing all instances. Measure (D4) is a dissimilarity measure rather than an actual distance metric [78], [157], [171]. It is derived from the Pearson correlation coefficient applying equation (4.1)

$$d(\mathbf{lp}_i, \mathbf{cc}_j) = 1 - s(\mathbf{lp}_i, \mathbf{cc}_j) \quad \textbf{(4.1)}$$

given $s(\mathbf{lp}_i, \mathbf{cc}_j)$ is the Pearson correlation coefficient described by equation (4.2),

$$s(\mathbf{lp}_i, \mathbf{cc}_j) = \frac{\sum_{t=1}^{T}\left(\mathrm{lp}_i(t) - \overline{\mathbf{lp}}_i\right)\left(\mathrm{cc}_j(t) - \overline{\mathbf{cc}}_j\right)}{\left[\sum_{t=1}^{T}\left(\mathrm{lp}_i(t) - \overline{\mathbf{lp}}_i\right)^2 \sum_{t=1}^{T}\left(\mathrm{cc}_j(t) - \overline{\mathbf{cc}}_j\right)^2\right]^{1/2}} \quad \textbf{(4.2)}$$

4.3.4 Estimation error measures

The errors between the estimated and actual half-hourly measurements of individual and aggregated smart meters were quantified using the maximum absolute and root-mean-square error measures. Equations (4.3) and (4.4) define the estimation errors.

$$Max.AE = \max\{|\mathrm{lp}_{act}(t) - \mathrm{lp}_{est}(t)|\}_{t\in T} \quad \textbf{(4.3)}$$

$$RMSE = \left(\frac{\sum_{t=1}^{T}(\mathrm{lp}_{act}(t) - \mathrm{lp}_{est}(t))^2}{T}\right)^{1/2} \quad \textbf{(4.4)}$$

The overall maximum absolute error, and the mean RMSE of the estimated measurements were calculated according to equations (4.5) and (4.6) respectively.

$$Overall\_Max\_AE = \max\left\{\max\left\{\left|\mathrm{lp}_{act,n}(t) - \mathrm{lp}_{est,n}(t)\right|\right\}_{t\in T}\right\}_{n\in N} \quad \textbf{(4.5)}$$

$$Mean\ RMSE = \frac{1}{N}\sum_{n=1}^{N}\sqrt{\frac{\sum_{t=1}^{T}(\mathrm{lp}_{act}(t) - \mathrm{lp}_{est}(t))^2}{T}} \quad \textbf{(4.6)}$$

In equations (4.3) - (4.6), the variables are

$\mathrm{lp}_{act}(t)$ is the actual half-hourly measurement of an individual smart meter or the aggregated load of a number of smart meters;

$\mathrm{lp}_{est}(t)$ is the estimated half-hourly measurement that was obtained using the k-means based load estimation algorithm;

N is the overall number of samples considered in the brute-force approach, $n \in N$.

4.4 Load estimation of individual smart meters

The measurements of the test period for each of the 96 individual smart meters (**Section 3.3**) were estimated using the load estimation algorithm. The load estimator utilises the cluster centres obtained as a result of applying a Mean AvED threshold of 1% of the mean consumption of the training period.

4.4.1 Impact of the distance function

The performance of the four distance functions (**Section 4.3**) to estimate the half-hourly measurements of smart meters is compared in Figure 4.4. The results indicate that the application of Pearson correlation distance produces smaller values of the median of estimation errors than the other functions. Figure 4.4(a) shows the distribution of the overall maximum absolute estimation error, whereas the distribution of the mean RMS estimation error is shown in Figure 4.4(b).

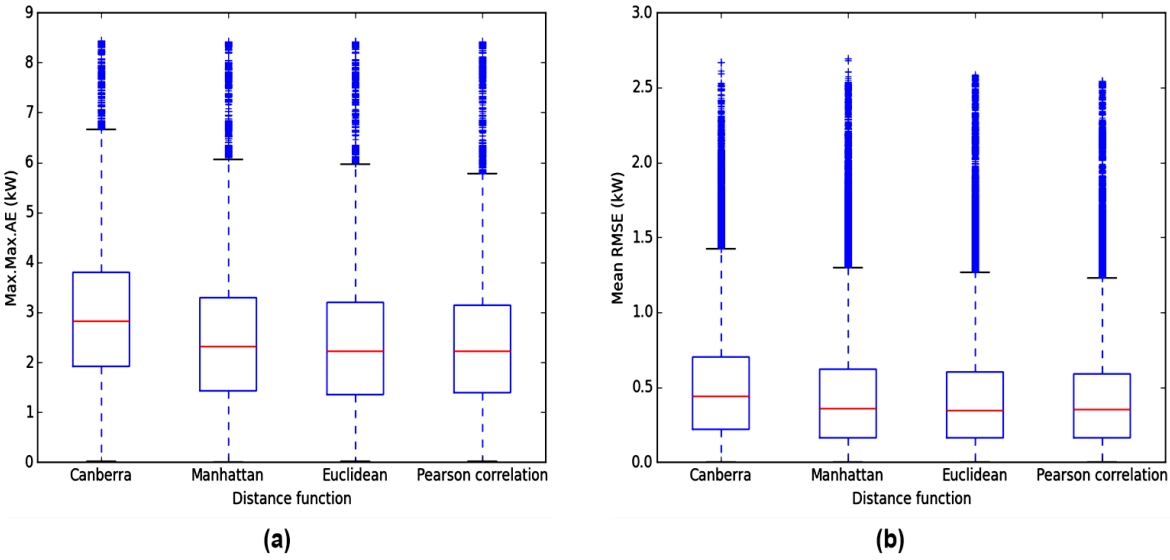


Figure 4.4 Load estimation errors for different distance functions – Individual smart meters
(a) Overall maximum absolute error and (b) Mean RMSE

Figure 4.4(a) and (b) shows that regardless of the segmentation time window and duration of missing measurements, the estimation errors had approximately similar distributions. However, as compared to the other functions, the maximum estimation error (whisker) was the smallest and the outliers ("+" signs) were relatively less when Pearson correlation distance was applied. The determination of maximum errors and outlier follows the box-whisker plot principles defined in Appendix A.1.

The results shown in Sections (**4.4.2**) – (**4.4.4**) are based on the application of Pearson Correlation distance to estimate the load of individual smart meters.

4.4.2 Impact of the daily and segmented cluster centres

The distributions of load estimation errors resulting from the application of daily and segmented cluster centres to estimate the load of the 96 individual smart meters are shown in Figure 4.5.

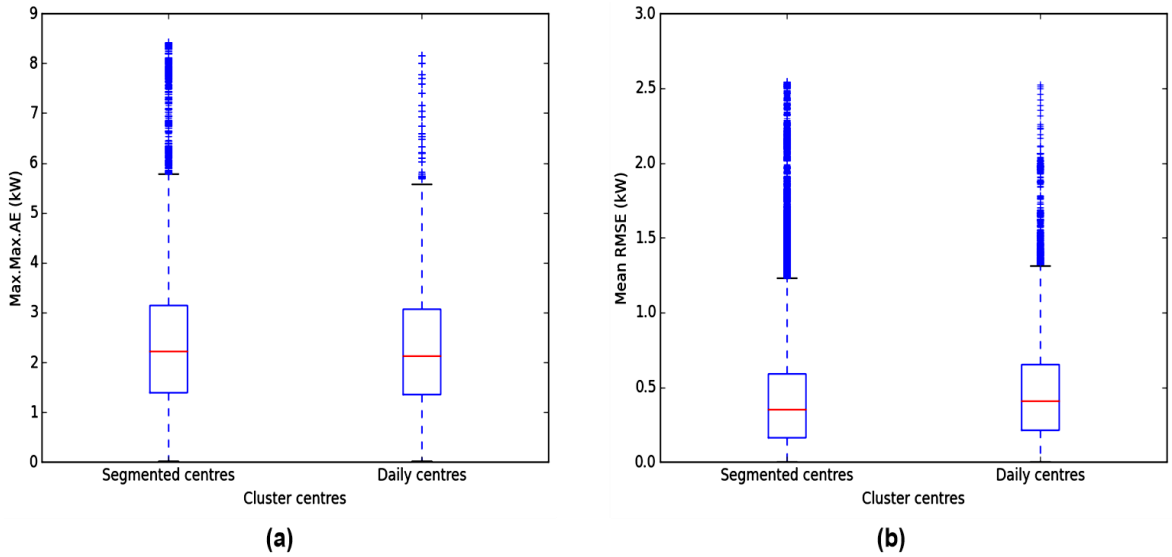


Figure 4.5 Load estimation errors for different types of cluster centres – Individual smart meters
(a) Overall maximum absolute error and (b) Mean RMSE

Figure 4.5(a) shows that the application of segmented and daily cluster centres results in similar distributions of the overall maximum load estimation errors. However, the distributions of mean RMS estimation errors, shown in Figure 4.5(b), indicate that the segmented cluster centres produce slightly smaller errors than the daily cluster centres.

4.4.3 Impact of the segmentation time window

The distributions of the load estimation errors for different time windows of segmented cluster centres are shown in Figure 4.6.

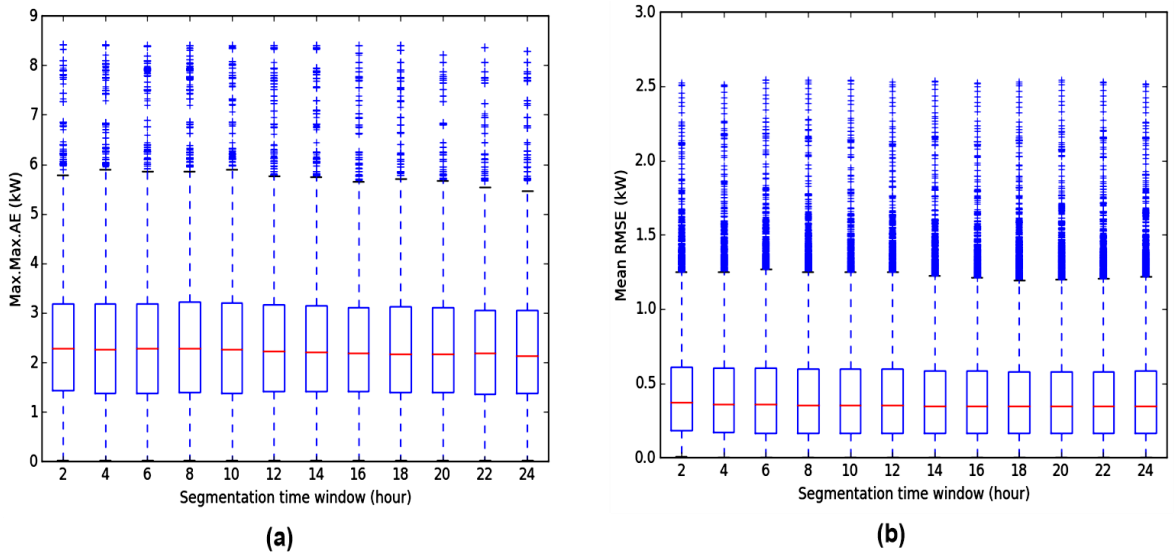


Figure 4.6 Load estimation errors for different segmentation time windows of cluster centres – Individual smart meters
(a) Overall maximum absolute error and (b) Mean RMSE

Figure 4.6 shows that (a) the overall maximum absolute error and (b) mean RMSE had similar distributions regardless of the segmentation time windows. However, detailed investigation of the load estimation errors reveals a small reduction in the maximum values of the errors (Max.Max.AE and Mean RMSE) when the segmented cluster centres were in the range of 16-24 hours.

4.4.4 Impact of duration of missing measurements

For different durations of lost measurements, distributions of the load estimation errors are shown in Figure 4.7.

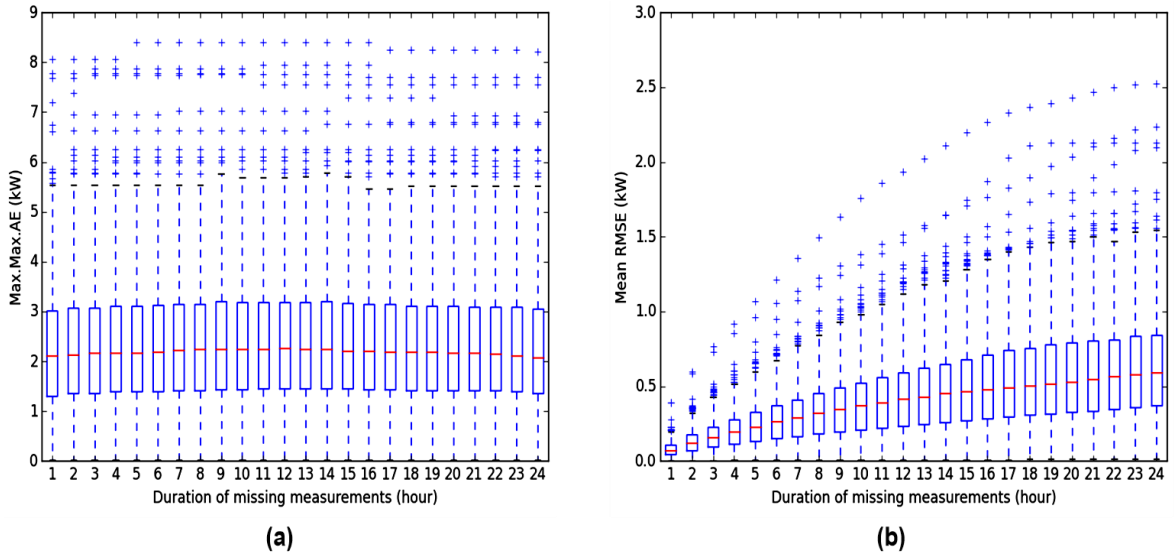


Figure 4.7 Load estimation errors for different durations of missing measurements – Individual smart meters
(a) Overall maximum absolute error and (b) Mean RMSE

Figure 4.7(a) shows that the overall maximum absolute load estimation errors had approximately the same distributions for different durations of missing measurement. However, the distribution of the mean RMS load estimation errors shown in Figure 4.7(b) clearly shows that the increase in the duration of missing measurements results in a similar increase in the estimation errors. In this case, the median of the mean RMS estimation error was in the range of 10-50% of the mean consumption of a smart meter (that was about 0.611kW during the training period and 0.620kW during the test period) up to 8 hours of measurement loss.

4.5 Load estimation of aggregated smart meters

The load estimation algorithm was applied to estimate the aggregated half-hourly measurements (of 96 individual smart meters) collected during the test period. Cluster centres obtained as a result of applying a Mean AvED threshold of 1% (of the mean consumption) to group the aggregated load profiles of the training period, were utilised to estimate missing measurements of the test period.

4.5.1 Impact of the distance function

Simulation results indicate that the application of Canberra distance yields more accurate load estimates than other distance functions. The distribution of the load estimation errors is shown in Figure 4.8.

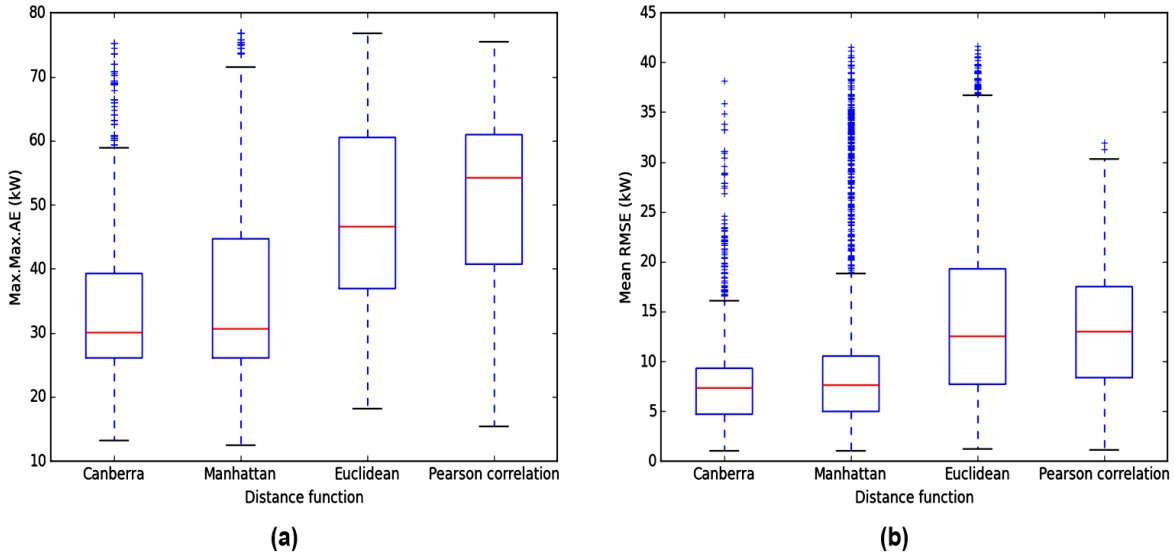


Figure 4.8 Load estimation errors for different distance functions – Aggregated smart meters
(a) Overall maximum absolute error and (b) Mean RMSE

Regardless of the duration of lost measurements and time window of segmented cluster centres, Figure 4.8 shows that the application of Canberra and Manhattan (city block) distance functions results in the smallest error distributions. The median of the (a) overall maximum absolute and (b) mean RMS estimation errors, due to the application of Canberra and Manhattan distances, had approximately the same values. However, Canberra distance results in slightly narrower error distributions and smaller whiskers of the estimation errors as compared to Manhattan distance function.

Figure 4.8(b) indicates that Canberra distance function produces more accurate load estimates; 75% of the errors were less than 9.5kW when the load was estimated using Canberra distance, whereas the same percentage of errors was less than 10.5kW when Manhattan distance was applied. The mean consumption of the aggregated smart meters was approximately 61.3kW during the training period and 62.7kW during the test period. The results presented in the Sections (**4.5.2**) – (**4.5.4**) are based on the application of Canberra distance function to estimate the aggregated load.

4.5.2 Impact of the daily and segmented cluster centres

Simulation results show that the application of Canberra distance, and segmented cluster centres for load estimation results in more accurate estimates than the application of daily cluster centres does. Figure 4.9 shows the distribution of load estimation errors when segmented and daily cluster centres were applied to estimate the missing measurements.

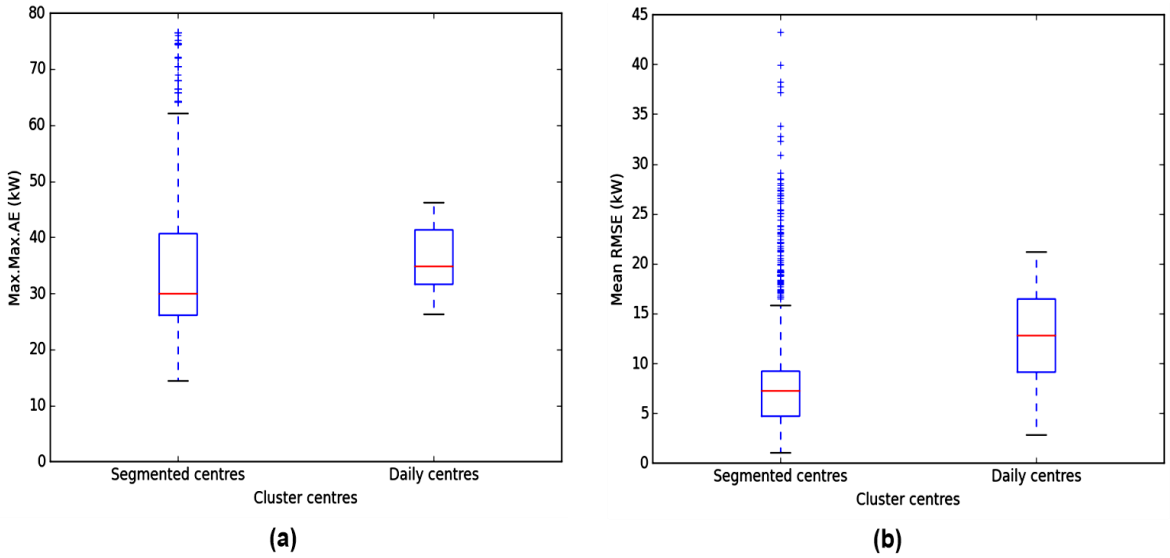


Figure 4.9 Load estimation errors for different types of cluster centres – Aggregated smart meters
(a) Overall maximum absolute error and (b) Mean RMSE

Figure 4.9(a) shows that the application of segmented cluster centres results in larger whiskers (maximum errors) and more outliers than the application of daily centres. These maximum errors and outliers were caused by the application of short (2-4 hours) segmented cluster centres to estimate the load; and can be ignored by using longer cluster centres to estimate the missing measurements. At the same time, Figure 4.9(b) shows that the application of segmented cluster centres results in a narrower distribution of the mean RMSE as compared to the application of daily cluster centres. However, it was found that the application of short segment cluster centres (with a time window between 2-6 hours) results in large numbers of outliers as compared to the application of long centres.

4.5.3 Impact of the segmentation time window

Simulation results revealed that the application of segmented cluster centres whose time window is in the range of 16-24 hours results in the smallest load estimation errors. Figure 4.10 shows the distributions of load estimation errors for different time windows of segmented cluster centres.

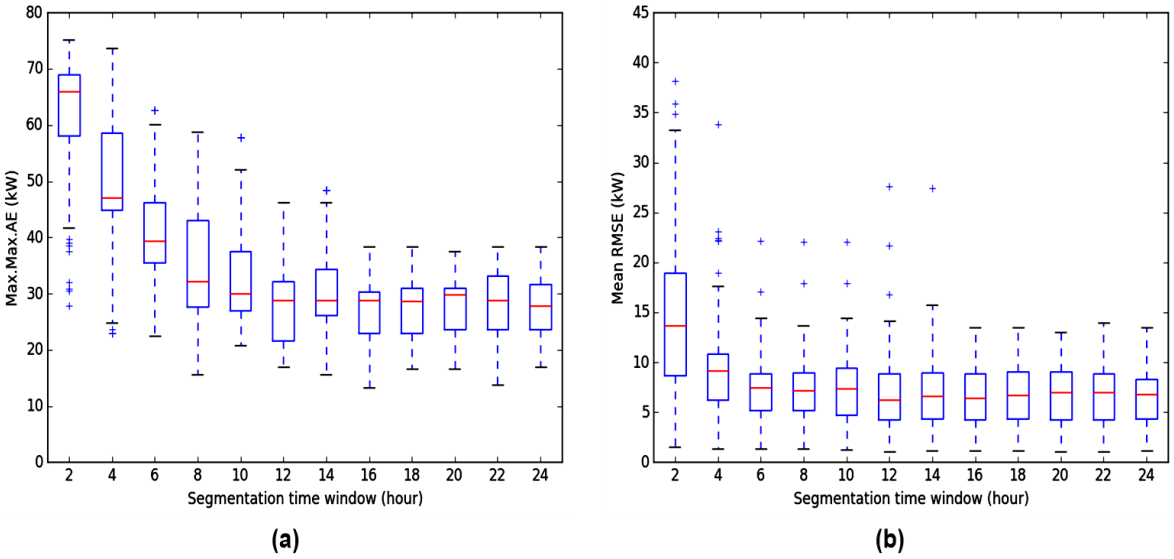


Figure 4.10 Load estimation errors for different segmentation time windows of cluster centres – Aggregated smart meters
(a) Overall maximum absolute error and (b) Mean RMSE

In Figure 4.10, (a) the overall maximum absolute errors and (b) mean RMSE of the estimated measurements show that the smallest values of minimum and maximum errors (error bars) were observed in the 16-24 hours' range of segmentation time windows.

4.5.4 Impact of duration of missing measurements

Simulation results indicate that the load estimator was capable of estimating the aggregated load for different durations of measurement loss. Figure 4.18 shows the distribution of load estimation errors up to 24 hours of measurement loss when a segmentation time window of 16 hours was selected.

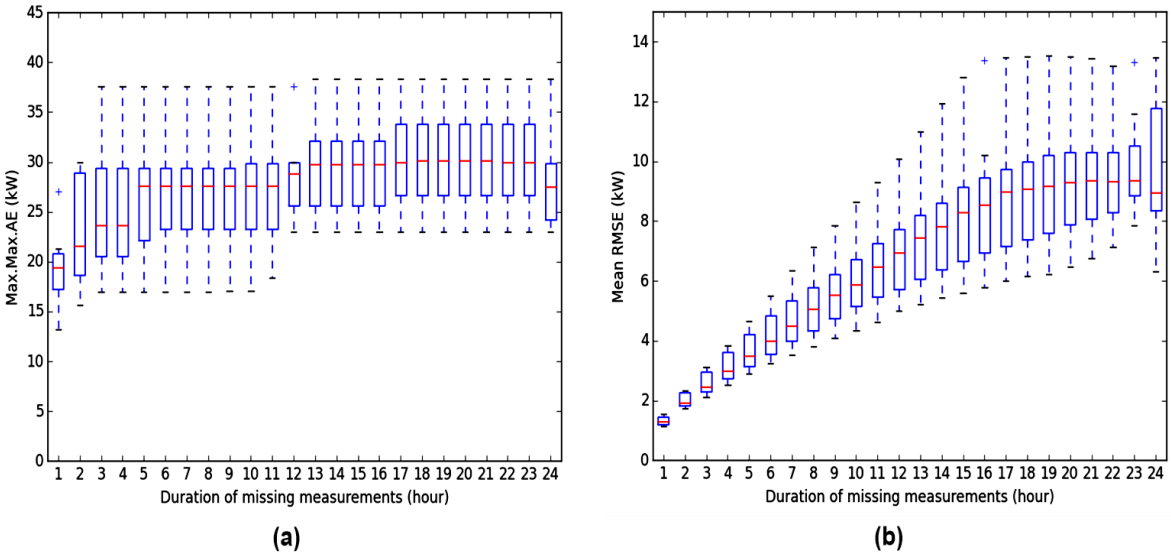


Figure 4.11 Load estimation errors for different durations of missing measurements – Aggregated smart meters
(a) Overall maximum absolute error and (b) Mean RMSE

The mean daily power consumption of the aggregated smart meter measurements during the test period was approximately 61kW for working days and 65kW for the weekends. Figure 4.11(b) shows that when the maximum values of the Mean RMSE are neglected, up to 10 hours of measurements are estimated with an approximate error that was equal to 10% of the mean daily consumption. When the maximum values of estimation errors are taken into consideration, the 10% estimation error covers only 7 hours of missing measurements.

4.6 Conclusions

A load estimation algorithm based on the k-means based cluster analysis method and principles of pattern recognition was developed. The load estimation algorithm utilises the cluster centres, obtained through the clustering of daily and segmented profiles of individual and aggregated residential smart meters, to estimate missing and future measurements of smart meters. The accuracy of the estimated measurements was investigated using different distance functions, different segmentation time windows and several scenarios of lost measurements.

Simulation results showed that the Pearson correlation distance function provided more accurate load estimates than other functions when the load of individual smart meters was estimated. At the same time, when the load of aggregated smart meters was estimated, the Canberra distance outperformed the other functions through the provision of higher accuracy load estimates.

A comparison of the estimation errors showed that the application of segmented cluster centres to estimate the load of individual smart meters produced approximately smaller errors than did the application of daily cluster centres. Different segmentation time windows resulted in similar distributions of the estimation errors. However, segmented cluster centres in the range of 16-24 hours yielded smaller maximum values of the mean RMS estimation errors than those centres in the range of 2-14 hours. The median (of mean RMS) error of estimating up to 8 hours of lost measurements was approximately 50% of the mean consumption of all individual smart meters (of the training period).

The application of the segmented cluster centres results in significantly higher accuracy load estimates, of the aggregated load. than did the application of daily cluster centres. Segmented cluster centres in the range of 16-24 hours resulted in less estimation errors as compared to the 2-14 hours of segmentation. Ignoring the maximum values of the Mean RMS estimation errors, up to 10 hours of measurements were estimated with an approximate error that was equal to 10% of the mean daily consumption. At the same time, taking the maximum values of estimation errors into consideration, the duration of lost measurements decreases from 10 to 7 hours for the same 10% error of the mean consumption.

Chapter 5 Integrated Load and State Estimation

Note: The Iteratively Re-Weighted Least Squares (IRWLS) state estimator developed by Professor Jianzhong Wu [172] was used in the study reported in this chapter.

5.1 Introduction

An integrated load and state estimation algorithm to extend the observability and estimate the voltage magnitudes and angle at each busbar of a distribution network is developed in this chapter. The integrated estimation algorithm links the k-means based load estimator with an IRWLS state estimator.

This chapter starts with investigating the capabilities of the IRWLS state estimator by using real measurements from an actual LV microgrid. The measurements, collected at 10-second intervals, encompassed busbar voltage magnitudes and (active and reactive) power demand as well as (active and reactive) line power flows. The state estimator outputs were compared to the measurements to assess the accuracy of the state estimator. The impact of the measurement type upon the estimated state of the network was also studied.

The IEEE 33 bus test network was used to assess the integrated load and state estimation algorithm. The inputs to the state estimator comprised a limited number of MV real-time measurements (only at the HV/MV transformer) and real-time and pseudo measurements derived from smart metering data (collected every 30 minutes). Pseudo measurements were obtained through the load estimation algorithm (**Chapter 4**).

5.2 Iteratively Re-Weighted Least Squares State Estimation

In the context of smart grid scenarios, state estimation is becoming a fundamental tool to monitor, control, and operate distribution networks. Distribution Management Systems (DMS) rely on state estimators to estimate the operating state of the distribution network. The state of a distribution network is defined as the voltage magnitude and angle of each busbar [140], [141], [173], [174].

State estimation is defined as the computation of an accurate operating state of a power network starting from a given set of measurements. In transmission networks, state estimation acts as a filter between a set of redundant raw real-time measurements and all applications that require the most reliable database of the present state of the

network. The set of real-time measurements usually comprises busbar voltage magnitudes, active and reactive power injections, active and reactive line power flows, and line currents. Real-time measurements are collected from meters installed at various network busbars and lines. In this manner, the state estimator minimises the error between the real-time measurements and the calculated values of these measurements [136], [141].

On the other hand, state estimation is used to extend the observability and to identify the operating state of a distribution network using a limited set of real-time measurements and large numbers of pseudo measurements. In this chapter, pseudo-measurements (that can usually be obtained from load estimates based on historical load profiles, transformer kVA ratings, and near real-time smart metering measurements [9], [12], [128], [132], [136]) are the active and reactive power demand at each busbar of the network.

Distribution state estimators were developed taking into consideration the radial topology of passive distribution networks. Legacy distribution networks are evolving to become smart active networks. The following aspects are examples of the characteristics of this evolution:

- an increased number of small-scale DERs connected to distribution networks especially at the LV level;
- sensors, accompanied with the necessary ICT, deployed to provide better power quality to the end-users; and
- power electronic devices such as the soft-open-points (SOP) that can facilitate the integration of AC and DC grids [6], [175], [176].

The radial topology of distribution networks is unlikely to be retained once the concepts of smart grids are adopted, and the application of state estimation methods that were specially designed for radial networks would become infeasible. Therefore, a state estimator based on the IRWLS state estimation method, that was applied in transmission networks [148], is used in this study.

The Weighted Least Squares (WLS) algorithm, that minimises the sum of weighted squared residuals between the measured and the estimated values of the network states, is used to solve the state estimation problem. Equations (5.1) – (5.4) define the WLS algorithm.

$$J(x) = \min \sum_{i=1}^{m} \left(\mathbf{z}_i - \mathbf{h}_i(\mathrm{x})\right)^2 \mathrm{W}_{ii} \quad \textbf{(5.1)}$$

$$J(x) = (\mathbf{z} - \mathbf{h}(\mathbf{x}))^T \mathbf{W} (\mathbf{z} - \mathbf{h}(\mathbf{x})) \quad \textbf{(5.2)}$$

subject to

$$\mathrm{r}_i = \mathrm{z}_i - \mathrm{h}_i(\mathrm{x}) \quad \textbf{(5.3)}$$

where r_i is the residual of the i^{th} measurement and z_i is the i^{th} measurement. The vector $\mathbf{z}$ denotes a set of m measurements

$$\mathbf{z} = \begin{bmatrix} \mathrm{z}_1 \\ \mathrm{z}_2 \\ \vdots \\ \mathrm{z}_m \end{bmatrix} = \begin{bmatrix} \mathrm{h}_1(\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_n) \\ \mathrm{h}_2(\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_n) \\ \vdots \\ \mathrm{h}_m(\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_n) \end{bmatrix} + \begin{bmatrix} \mathrm{e}_1 \\ \mathrm{e}_2 \\ \vdots \\ \mathrm{e}_m \end{bmatrix} = \mathbf{h}(\mathbf{x}) + \mathbf{e} \quad \textbf{(5.4)}$$

given

$\mathbf{h}(\mathbf{x})^T = [\mathrm{h}_1(\mathbf{x}), \mathrm{h}_2(\mathbf{x}), \cdots, \mathrm{h}_m(\mathbf{x})]$

$\mathrm{h}_i(\mathbf{x})$ is the nonlinear function that relates the i^{th} measurement to the state vector $\mathbf{x}$;

$\mathbf{x}^T = [\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_n]$ is the system state vector denoting n system states;

$\mathbf{e}^T = [\mathrm{e}_1, \mathrm{e}_2, \cdots, \mathrm{e}_m]$ is the vector of measurement errors;

$\mathbf{r}^T = [\mathrm{r}_1, \mathrm{r}_2, \cdots, \mathrm{r}_m]$ is the vector of measurement residuals; and

$\mathbf{W}$ is the weight matrix. The weight matrix is a diagonal matrix where the weights of measurements are in the main diagonal. The assumption that measurement errors are independent and are not correlated, results in setting the off-diagonal elements of the weight matrix to zeros. The weight – that is equal to the reciprocal of the variance of a measurement $\left(1/\sigma_i^2\right)$ – reflects the accuracy of a measurement.

$$\mathbf{W} = \begin{bmatrix} \mathrm{W}_{11} & 0 & 0 & 0 \\ 0 & \mathrm{W}_{22} & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \mathrm{W}_{mm} \end{bmatrix} = \begin{bmatrix} \sigma_1^{-2} & 0 & 0 & 0 \\ 0 & \sigma_2^{-2} & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_m^{-2} \end{bmatrix} \quad \textbf{(5.5)}$$

In order to minimise the performance index, J, described in equation (5.1) a first-order condition must hold [177]:

$$\left.\frac{\partial J}{\partial \mathbf{x}}\right|_{\mathrm{x}^k} = 0 \qquad \textbf{(5.6)}$$

Evaluating equation (5.2) at the necessary conditions gives

$$\mathbf{H}(\mathbf{x}^{\boldsymbol{k}})^{\boldsymbol{T}}\mathbf{W}(\mathbf{z} - \mathbf{h}(\mathbf{x})) = 0 \qquad \textbf{(5.7)}$$

where $\mathbf{H}(\mathbf{x})$ is the measurement Jacobian matrix evaluated at iteration k

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} \frac{\partial \mathrm{h}_1}{\partial \mathrm{x}_1} & \frac{\partial \mathrm{h}_1}{\partial \mathrm{x}_2} & \cdots & \frac{\partial \mathrm{h}_1}{\partial \mathrm{x}_n} \\ \frac{\partial \mathrm{h}_2}{\partial \mathrm{x}_1} & \frac{\partial \mathrm{h}_2}{\partial \mathrm{x}_2} & \cdots & \frac{\partial \mathrm{h}_2}{\partial \mathrm{x}_n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \mathrm{h}_m}{\partial \mathrm{x}_1} & \frac{\partial \mathrm{h}_m}{\partial \mathrm{x}_2} & & \frac{\partial \mathrm{h}_m}{\partial \mathrm{x}_n} \end{bmatrix}_{\mathbf{x}^k} \qquad \textbf{(5.8)}$$

A linearized relationship between the measurements and the state variables is then found by expanding the function $\mathbf{h}(\mathbf{x})$ around a point $\mathbf{x}^k$ using the Taylor series expansion

$$\mathbf{h}(\mathbf{x}) = \mathbf{h}\left(\mathbf{x}^{\boldsymbol{k}}\right) + \Delta\mathbf{x}^{\boldsymbol{k}}\frac{\boldsymbol{\partial}\mathbf{h}(\mathbf{x})}{\boldsymbol{\partial}\mathbf{x}} + \text{higher order terms} \qquad \textbf{(5.9)}$$

The set of (5.9) equations can be solved using an iterative approach such as Newton-Raphson's method. At the $(k+1)^{th}$ iteration, the updated values of the state variables can be obtained from their values in the previous iteration using equation (5.10)

$$\mathbf{x}^{\boldsymbol{k+1}} = \mathbf{x}^{\boldsymbol{k}} + \left(\mathbf{H}\left(\mathbf{x}^{\boldsymbol{k}}\right)^{\boldsymbol{T}}\mathbf{W}\,\mathbf{H}\left(\mathbf{x}^{\boldsymbol{k}}\right)\right)^{-\mathbf{1}}\mathbf{H}\left(\mathbf{x}^{\boldsymbol{k}}\right)^{\boldsymbol{T}}\mathbf{W}\left(\mathbf{z} - \mathbf{h}\left(\mathbf{x}^{\boldsymbol{k}}\right)\right) \qquad \textbf{(5.10)}$$

At convergence, the solution $\mathbf{x}^{\boldsymbol{k+1}}$ corresponds to the weighted least squares estimates of the state variables. convergence can be determined either by satisfying

$$\max\left(\mathbf{x}^{k+1} - \mathbf{x}^{k}\right) \leq \epsilon \qquad \textbf{(5.11)}$$

or

$$\boldsymbol{J}^{k+1} - \boldsymbol{J}^{k} \leq \epsilon \qquad \textbf{(5.12)}$$

where ϵ is some predetermined convergence factor.

Measurements with high variances (i.e., low weights) have less impact upon the final solution of state estimation than those measurements that have low variances. The WLS estimators provide consistent and accurate estimates as compared to other state estimation methods. However, WLS estimators are affected by the accuracy of real-time and pseudo measurements, gross measurement errors and noisy measurements [9], [12], [131], [178].

IRWLS algorithms are more robust than the WLS algorithms [148], [149], [151], [172]. In IRWLS algorithms, $\mathbf{W}$, the weight matrix, includes functions that calculate the measurement residuals at each iteration of state estimation. The measurements are iteratively reweighted according to the values of their residuals. If the residual of any measurement is found to increase, the confidence in the measurement decreases and its weight is reduced. If the residual decreases, the confidence in the measurement is greater and its weight is increased [150]. Appendix A.2 demonstrates the principles of reweighting measurements in IRWLS estimators.

The set of measurements available for the present study includes busbar voltage magnitudes, busbar active and reactive power demands, and line active and reactive power flows. Each measurement can be calculated as a function of the state variables. The state variables were defined as the voltage magnitudes (V_k) and voltage phase angles (θ_k) at each busbar of the network [7]. The state vector was therefore defined as

$\mathbf{x}^T = [V_1, V_2, V_3, \cdots, V_N, \theta_2, \theta_3, \cdots, \theta_N,]$ and N is the number of busbars of the network.

Equations (5.13) and (5.14) illustrate relationship between the active and reactive power injections at each busbar of the network and the state variables.

$$P_k = V_k \sum_{j \in N_k} V_j \left(G_{kj} \cos\theta_{kj} + B_{kj} \sin\theta_{kj}\right) \quad \textbf{(5.13)}$$

$$Q_k = V_k \sum_{j \in N_k} V_j \left(G_{kj} \sin\theta_{kj} - B_{kj} \cos\theta_{kj}\right) \quad \textbf{(5.14)}$$

given $j = 1, 2, \cdots, N_k$ and $j \neq k$. N_k are the busbars connected to busbar k under consideration;

[7] Except the voltage phase angle (θ_1) at the grid connection point (Busbar 1)

G_{kj} is the sum of the conductance of all feeder sections that link busbars k and j; and B_{kj} is the sum of the susceptance of the corresponding feeder sections.

Equation (5.15) represents the flow of active power from busbar i to busbar j of the network. The reactive power flow from busbar i to busbar j is calculated according to equation (5.16)

$$P_{kj} = g_{kj}V_k^2 - V_kV_jg_{kj}\cos\theta_{kj} - V_kV_jb_{kj}\sin\theta_{kj} \quad \textbf{(5.15)}$$

$$Q_{kj} = -b_{kj}V_k^2 - V_kV_jg_{kj}\sin\theta_{kj} + V_kV_jb_{kj}\cos\theta_{kj} \quad \textbf{(5.16)}$$

given that g_{kj} is the conductance of the feeder section connecting busbars k and j and b_{kj} is the susceptance of this section. For the case of two busbars connected with a single section, then $G_{kj} = g_{kj}$ and $B_{kj} = b_{kj}$ provided that there are no tap-changing transformers connected across the feeder and that the shunt admittance of the feeder section is equal to zero.

5.3 Test Distribution Networks

Measurements collected (at a rate of one measurement every 10 seconds) from an actual LV microgrid were used to investigate the performance of the IRWLS estimator. Using different configurations of measurement type and location, the state estimator outputs were compared against the measurements to quantify the accuracy of the estimated state of the microgrid. The capability of the IRWLS estimator to detect and reduce the impacts of noise and gross errors in the real-time measurements was investigated.

Real-time and pseudo measurements derived from half-hourly smart metering data were used in conjunction with the IEEE 33 bus distribution network to assess the performance of the integrated load and state estimation algorithm to increase the observability of distribution network when a limited number of real-time measurements are available for the state estimator to utilise.

5.3.1 Low voltage microgrid

The Distributed Energy Resources Test Facility (DER-TF) is a practical test network established by Ricerca sul Sistema Energetico (RSE) in Italy. The network

represents a three-phase LV microgrid [179] that consists of several generators with different technologies (renewable and conventional), controllable loads and energy storage systems. Figure 5.1 shows the single line diagram of the DER-TF microgrid.

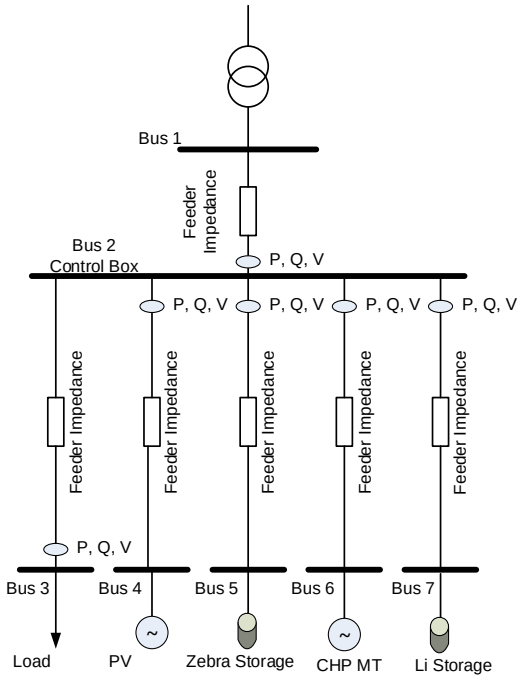


Figure 5.1 Single line diagram of DER-TF microgrid

The LV microgrid is connected to the distribution grid through a 23kV/0.4kV transformer. The microgrid consists of a photovoltaic (PV) array, a Zebra (molten salt) battery energy storage system (BESS), a Gas Combined Heat and Power (CHP) micro turbine, a Lithium BESS, and a programmable resistive and inductive load. A detailed description of the technical specifications of DER-TF microgrid in provided in Appendix (A.3).

In Figure 5.1, Busbar 1 is the grid connection point and Busbar 2 is the control box where all (except the load) metering devices were connected. Phase (Line-to-ground) and line-to-line voltage magnitudes and active and reactive line power flows of all feeder sections (except the feeder section that connects the load) were measured in real-time at Busbar 2. Active and reactive power demands of a busbar were calculated using the active and reactive line power flow measurements (along the feeder section that connects the busbar to Busbar 2).

5.3.2 Medium voltage distribution network

The IEEE 33 bus test system, shown in Figure 5.2, is a hypothetical MV distribution network rated at 12.66kV and comprises 32 busbars and 5 looping branches (tie lines).

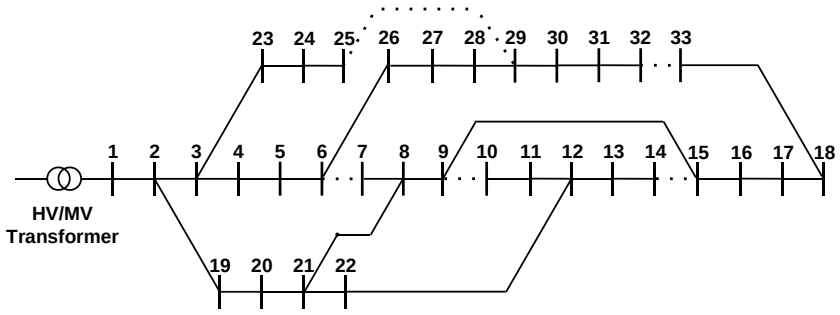


Figure 5.2 Single line diagram of the 33 bus distribution network

The peak load of the network is equal to 3841.12kW and 2396.06kVAr. Table 5.1 lists the active and reactive power injections at each busbar of the network. The positive sign in Table 5.1 indicates that the respective busbar is injecting power to the network and is therefore considered as a generator busbar, whereas the negative sign means that a load is connected to the busbar and is therefore absorbing power from the network

Table 5.1 Active and reactive power injections of the IEEE 33 bus network

Busbar number	***P injection (kW)***	***Q injection (kVAr)***	***Busbars number***	***P injection (kW)***	***Q injection (kVAr)***
1	+ 3802.19	+ 2694.60	**18**	– 90	– 40
2	– 100	– 60	**19**	– 90	– 40
3	– 90	– 40	**20**	– 90	– 40
4	– 120	– 80	**21**	– 90	– 40
5	– 60	– 30	**22**	– 90	– 40
6	– 60	– 20	**23**	– 90	– 50
7	– 200	– 100	**24**	– 420	– 200
8	– 200	– 100	**25**	– 420	– 200
9	– 60	– 20	**26**	– 60	– 25
10	– 60	– 20	**27**	– 60	– 25
11	– 45	– 30	**28**	– 60	– 20
12	– 60	– 35	**29**	– 120	– 70
13	– 60	– 35	**30**	– 200	– 600
14	– 120	– 80	**31**	– 150	– 70
15	– 60	– 10	**32**	– 210	– 100
16	– 60	– 20	**33**	– 60	– 40
17	– 60	– 20			

The network parameters are summarised in Appendix (A.4). The average R/X ratio of the feeder sections is equal to 1.46 and the active and reactive power losses of the network are 126.12kW and 96.01 kVAr. The active power loss was approximately 3.4% of the peak active power demand while the reactive power loss was about 4.2% of the peak reactive power demand of the network [180].

5.4 Low Voltage Microgrid State Estimation

Measurements were collected from the DER-TF LV microgrid and were used to validate the outputs of the IRWLS state estimator. Several case studies were carried out, taking into consideration the microgrid topology and the available DERs, to assess the impacts of the type of the measurement upon the estimated states of the network. Table 5.2 demonstrates the details of the implemented case studies.

Table 5.2 Case studies of state estimation in the DER-TF LV microgrid

	Base case (Reference case)	***Real-time substation measurements (RT_substation)***	***Real-time load power injection measurements (RT_load_PQ)***
Voltage magnitude measurements	All measurements were in real-time	- Only the voltage magnitude measurement at the substation was in real-time [8] - No other busbar V magnitude measurements	- Only the voltage magnitude measurements at the substation and the load were monitored in real-time [9] - No other busbar V magnitude measurements
Line PQ power flow measurements	All measurements were in real-time	- Only the line (P and Q) flows of Feeder 1 were in real-time - No other line (P and Q) flows were metered	- Only the line (P and Q) of Feeder 1 and Feeder 2 were in real-time - No other line (P and Q) flows metered
Busbar PQ demand measurements	All measurements were in real-time	- Only the substation (P and Q) demands were in real-time - All other (P and Q) demands were pseudo measurements	- Only the substation and the load (P and Q) demands were in real-time - All other (P and Q) demands were pseudo measurements

A redundancy of real-time measurements of the network was initially considered in the **Base case** study. This reference case study was carried out so that the actual values of the network state estimates were determined.

In the **RT_substation** case, the availability of real-time measurements was considered at the secondary of the MV/LV transformer. This set of measurements included the voltage magnitude measurement, active and reactive powers injected and active and reactive line power flows along the main feeder connecting the substation to

[8] All measurements were collected at the receiving end of Feeder section 1

[9] All measurements were collected at the receiving end of Feeder section 2

the rest of the LV network. Since the real-time measurements cannot guarantee full observability of the network in this case, the active and reactive power injections of all busbars, other than the substation busbars were assumed as pseudo measurements. State estimation was therefore run to extend the observability and determine the voltage magnitudes and angles at each busbar of the network.

The **RT_load_PQ** case study assumes that the substation measurements (voltage magnitude measurement, active and reactive powers injected and active and reactive line power flows) and the active and reactive power demands of the load (Busbar 3**)** were available in real-time. Smart meters were assumed to provide the real-time load measurements.

The outputs of the state estimator for the both case studies i.e. (**RT_substation**, and **RT_load_PQ**) were compared against the state estimator outputs of the **Base case** study. the root-mean-square error was applied to quantify the estimation errors – between the **RT_substation**, **RT_load_PQ** case studies and the **Base case.**

5.4.1 Impact of gross errors in real-time measurements

The real-time measurements collected from the DER-TF LV microgrid were partially impacted by gross errors. These errors in the measurements were caused by the short failure of the communication between the installed meters at Busbar 2 (the receiving end of Feeder section 1 in Figure 5.1), Busbar 3 (the receiving end of Feeder section 2 in Figure 5.1) and the PCs used to collect the measurements. The measurements collected at the sending end of the feeder section connecting the PV array to Busbar 2 (the control box) were also affected. During the temporary loss of communication, the measurements were recorded as zeros. The capability of the IRWLS state estimator to lessen the impacts of gross errors in real-time measurements was tested in this case.

Figure 5.3 illustrates a comparison of the measured voltage and the estimated voltage at the load busbar (Busbar 3).

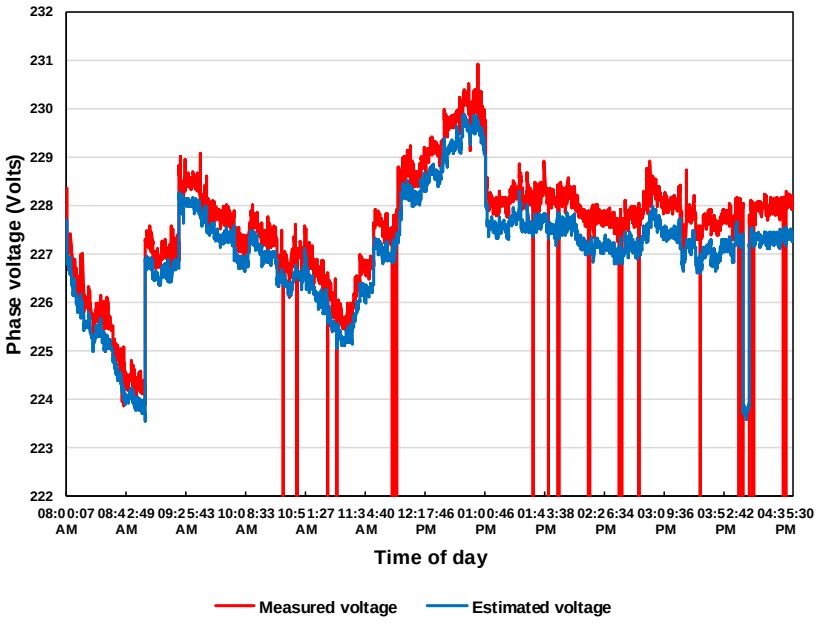


Figure 5.3 Comparison of the measured vs. estimated load voltage magnitude

Figure 5.3 shows that the voltage magnitude of the load busbar was reliably estimated even when the real-time measured voltage at that busbar was corrupted by gross errors. An offset between the estimated and the measured voltage magnitudes at Busbar 3 is observed in Figure 5.3. This indicates a possibility of bias within the meters installed at the load busbar. A test was carried out under no-load conditions to compare the measured voltages at Busbars 2 and 3. The no-load test confirmed that an offset of less than 1 volt was contained within the measurement devices under scrutiny. As a result, a recommendation was made to recalibrate or replace the measurement devices at Busbar 3. Based on the estimated voltage magnitudes and angles, the IRWLS state estimator is capable of calculating the active and reactive powers injected at each busbar of the network. Figure 5.4 shows a comparison of the estimated and measured active power injections at the load busbar.

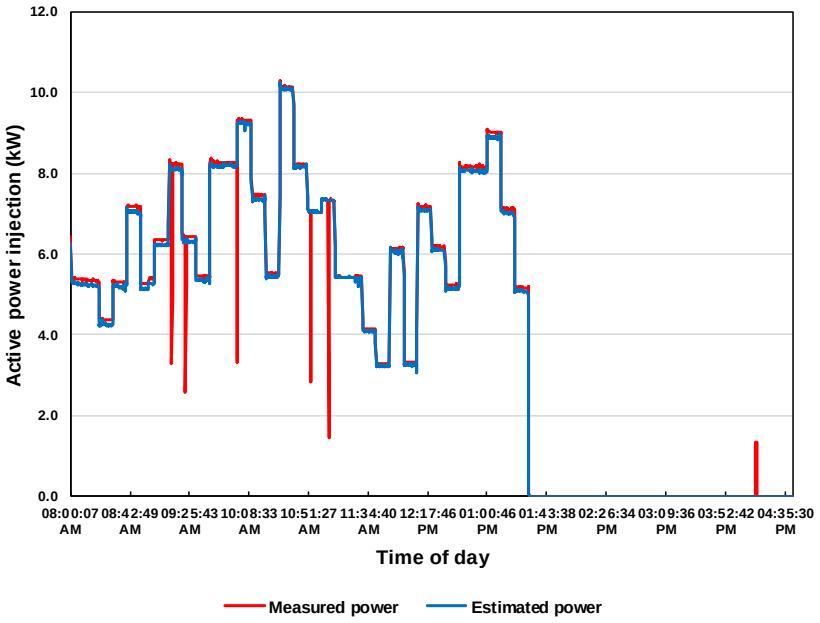


Figure 5.4 Comparison of estimated and measured active power injections at Busbar 3

Figure 5.4 shows that the active power demand, of the load busbar, was reliably estimated despite that real-time active power measurements contained some errors due to the temporary loss of communications. The estimated and measured reactive power demand of the load busbar shown in Figure 5.5 reveals that the IRWLS estimator was capable of removing the gross measurement errors provided that a set of redundant real-time measurements was available for the state estimator to use.

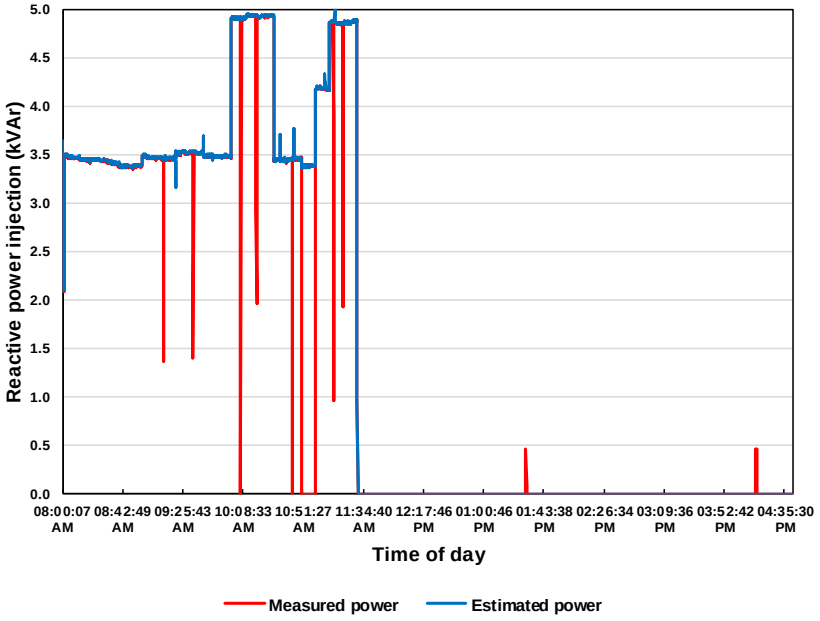


Figure 5.5 Comparison of estimated and measured reactive power injections at Busbar 3

5.4.2 Real-time substation measurements (RT_substation)

In a practical distribution network, the set of real-time measurements is generally available only at the primary substation. The set of measurements usually comprises of the voltage magnitude, the current and active and reactive line powers flowing through the main feeder section connecting the primary substation to the rest of the network, and the active and reactive powers injected by the primary substation to the network. The **RT_substation** case study was designed to simulate the availability of real-time measurements in a typical distribution network. This case study considers that the set of real-time measurements consists of the voltage measurement at Busbar 2, the active and reactive line power flows measured at the receiving end of Feeder section 1, and the active and reactive powers injected from the grid to the network. The active and reactive powers injections were determined using the power flows metered along Feeder section 1.

Some of the real-time measurements, collected at the receiving end of Feeder section 1, contained gross errors due to the temporary loss of communications (as stated earlier in **5.4.1**). Detailed analysis of the measurements' database indicate that the affected measurements can be divided into two sets. The first set includes a complete set of zero measurements of the phase voltages, line-to-line voltages, line currents, and active and reactive line power flows. On the other hand, the second set includes only zero (or corrupted) phase voltage measurements. The line-to-line voltages and the line power flows are not corrupted by gross errors.

Since the state estimator considers the real-time measurements at the receiving end of Feeder section 1 as leverage measurements in the **RT_substation** and **RT_load_PQ** case studies, the following approach was adopted to use zero error measurements as inputs to the IRWLS estimator. Initially, the set of real-time measurements was filtered by removing all the instances at which a complete failure of communications was observed. All other measurements, collected at other busbars of the network, with the same timestamps as the faulty substation measurements were also removed to guarantee the same number of measurements at all busbars of the network. At a further step, the faulty phase voltage measurements were calculated from the corresponding line-to-line voltage measurements according to equation (5.10)

$$V_{phase} = \frac{V_{line-to-line}}{\sqrt{3}} \quad \textbf{(5.7)}$$

An assumption was made that the LV microgrid is a balanced 3-phase network.

The filtered measurements were applied to the IRWLS estimator and the operating state of the microgrid was determined. The distribution of the RMSE of the estimated voltage magnitudes is shown in Figure 5.6. The errors between the **Base case** voltages and the estimated voltages according to the present case study are shown in this figure.

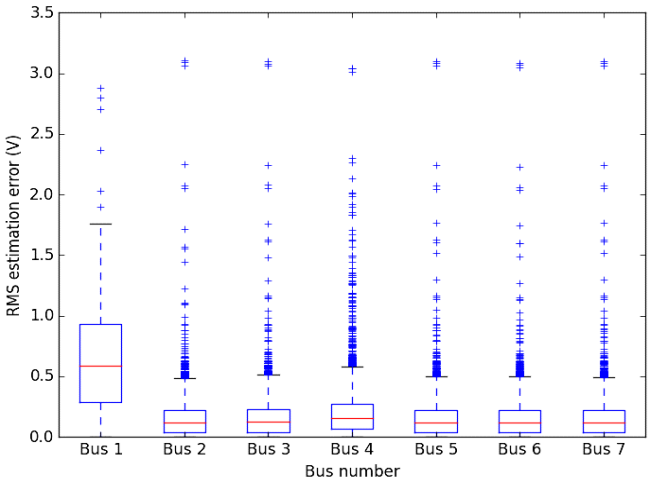


Figure 5.6 Error Distribution of the estimated busbar voltage magnitudes – RT_substation_case study

The largest distribution of the errors was observed at Busbar 1, which is the LV of the secondary substation, because no real-time measurements were installed at this busbar. The median of the voltage estimation error was equal to 0.6 volts at Busbar 1, whereas the median errors of the estimated voltages were approximately 0.15 volts at all other busbars. The maximum estimation errors were in the range of 0.5 volt (at Busbars 2 – 7) to 1.5 volts (at Busbar 1). Taking the error outliers, which were approximately 3 volts, into consideration, the maximum estimation errors were nearly 1.5% of the rated phase voltage, that is equal to 230 volts.

Figure 5.7 shows the distribution of RMSE between the estimated and **Base case** voltage angles at different busbars of the LV microgrid. Figure 5.7 reveals that the voltage angles exhibit the maximum RMS errors at Busbar 4 where the PV array was connected. The fluctuation of the generation of the installed PV panels, and not reaching the steady state in terms of the generated power both contributed to the observed distribution of the voltage angles errors at Busbar 4.

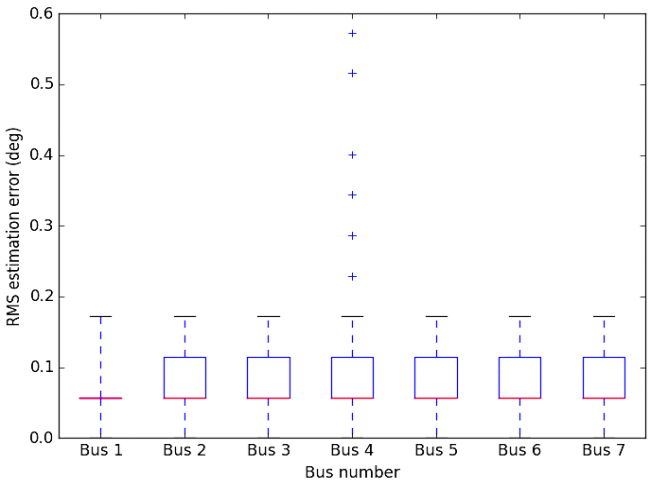


Figure 5.7 Error Distribution of the estimated busbar voltage phase angles RT_substation_case study

The distribution of the RMSE of the estimated active powers injections at each busbar of the network is shown in Figure 5.8.

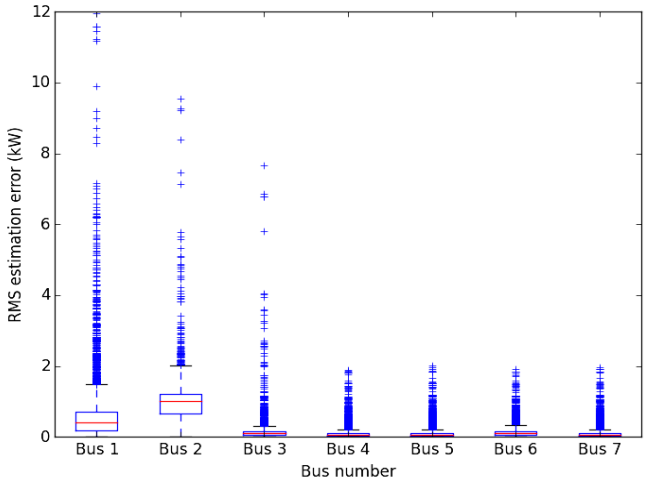


Figure 5.8 Error Distribution of the estimated active power injections RT_substation_case study

Figures 5.8 demonstrates that the errors of the estimated active power injected at each busbar of the network were less than 2 kW. Taking the error outliers into consideration, the maximum estimation errors was nearly 15 kW at Busbar 1, and 10 kW at Busbar 2. The maximum active power imported from the grid was approximately 50.31kW. Figure 5.9 illustrates the RMSE distribution of the estimated reactive power injections.

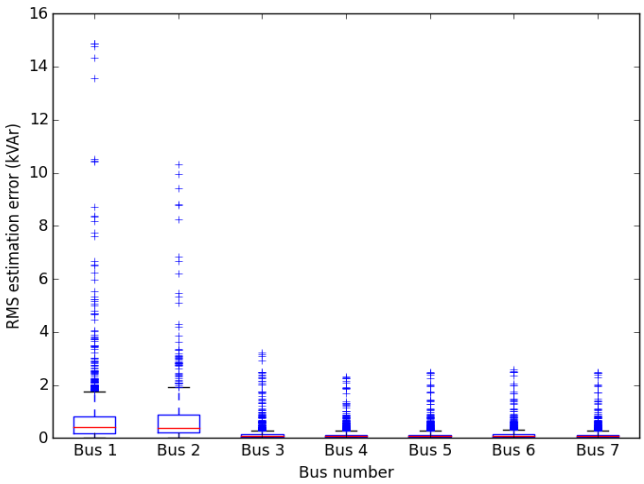


Figure 5.9 Error Distribution of the estimated reactive power injections RT_substation_case study

The median errors of the estimated reactive power injections were less than 1 kVAr, whereas the maximum values of the RMSE, with the error outliers included, were about 15 kVAr at Busbar 1 and 10 kVAr at Busbar 2. The maximum reactive power imported from the grid was approximately 51.84kVAr. Comprehensive analysis was carried out to investigate the reasons for the emergence of outlier errors in both estimated active and reactive power injections. These errors were a result of the fast switching of supplying the load from the grid to being supplied by the DGs. In this case, the measurements were collected from the network under transient conditions, with the steady state conditions not yet reached. The IRWLS estimator was designed only to provide a steady state solution for a balanced 3-phase electrical power network.

5.4.3 Real-time load power injection measurements (RT_load_PQ)

In this case study, real-time measurements are assumed to be available at the substation level, i.e. at the receiving end of Feeder section 1, and at Busbar 3, the load busbar, i.e. the receiving end of Feeder section 2. An assumption was made that smart meters delivered the average RMS voltages, and the active and reactive power injection measurements of the load [181]. State estimation was carried out with real-time measurements of the substation and the load, simultaneously with pseudo measurements representing the power injections of other busbars of the microgrid.

Figure 5.10 illustrates the distribution of the RMSE of the estimated voltage magnitudes at each busbar of the microgrid.

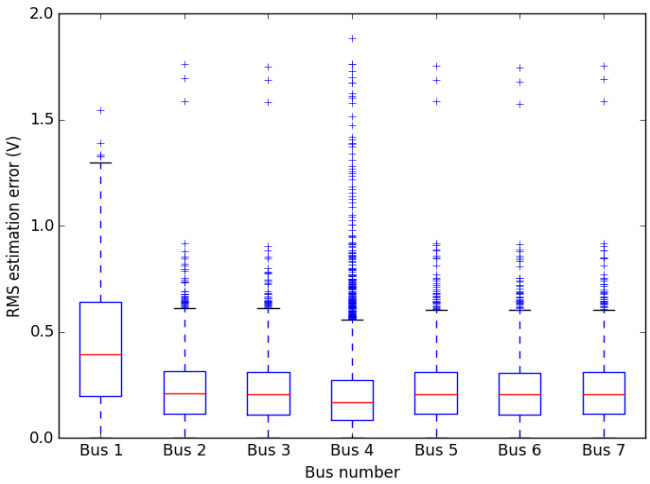


Figure 5.10 Error Distribution of the estimated busbar voltage magnitudes RT_load_PQ case study

Figure 5.10 shows that the distribution of the voltage estimation errors was less than 0.7 volts for all busbars of the network. When the maximum errors were taken into consideration, the RMSE of the estimated voltage magnitudes was within 1.5 volts of the **Base case** voltages. The distribution of the RMSE of the estimated voltage angles is shown in Figure 5.11.

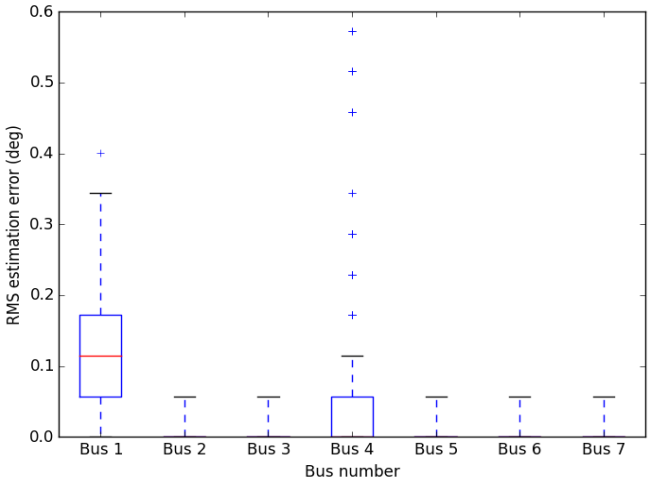


Figure 5.11 Error Distribution of the estimated busbar voltage angles RT_load_PQ case study

Figure 5.11 shows that the RMSE of the estimated voltage angles were approximately zero; and the maximum errors were less than 0.1 degree except at Busbar 1 where it was nearly 0.35 degrees. The error distribution of the estimated voltage angles at Busbar 4, the PV array, was observed as a result of the instability of the generated power at that busbar. This observation comes in accordance with the explanation provided in **(5.4.2)**.

Figure 5.12 shows the distribution of the RMSE of estimated (a) active power and (b) reactive power injections at each busbar of the network. In Figure 5.12, the outlier errors were approximately 12 kW for the estimated active power injection and 15 kVAr for the estimated reactive power injection. As explained in (**5.4.2**), these outliers were observed at instances when the steady state conditions of the network were not reached. In this case, the outliers can simply be ignored. The estimation errors of the active power injections remain below 2 kW. The estimated reactive power injections were within ±0.5 kVAr of their reference values.

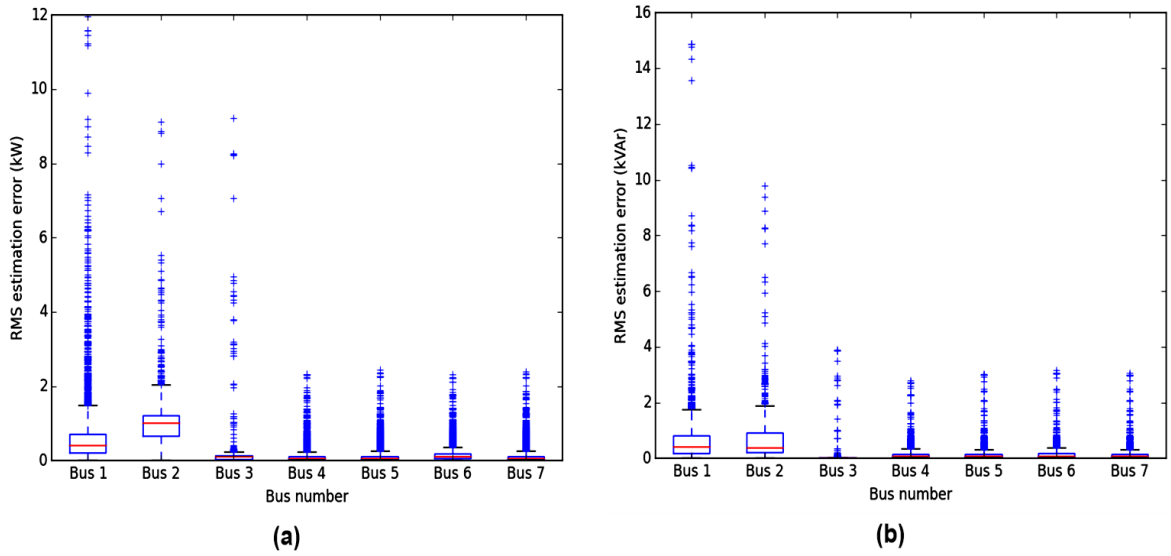


Figure 5.12 Error Distribution of the estimated busbar power injections - RT_load_PQ case study
(a) estimated active power errors and (b) estimated reactive power errors

5.4.4 Discussion of the estimation errors

The capability of the state estimator to filter gross errors in real-time measurements was demonstrated in **(5.4.1)**. The **RT_substation** case study in **(5.4.2)** assessed the accuracy of the network state estimates when only the substation measurements were monitored in real-time. Finally, the future scenario of real-time measurements of the substation and the load was simulated in the **RT_load_PQ** case of **(5.4.3)**. The latter study estimated the operating states of the LV microgrid; the estimation errors were quantified accordingly.

The estimation errors of the **RT_substation** case were compared against the estimation errors of the **RT_load_PQ** case study. The objective of the comparison was to quantify the impacts of the introduction of real-time load measurements upon the state estimator outputs.

A comparison of the errors of the estimated busbar voltage magnitudes for the **RT_substation** and **RT_load_PQ** case studies is shown in Figure 5.13. The distribution of the absolute estimation errors was depicted.

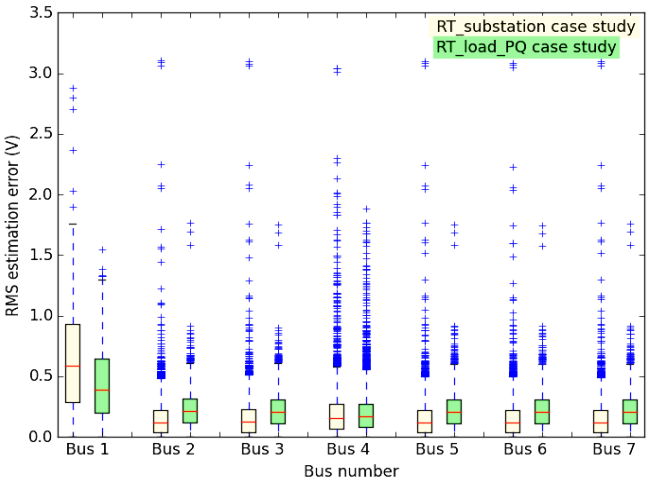


Figure 5.13 Comparison of the absolute errors of the estimated voltage magnitudes

Figure 5.13 clearly demonstrates that the estimation errors of the substation voltage magnitude were reduced considerably in the **RT_load_PQ** case study. However, in this case the estimated voltages at Busbar 2 through Busbar 7 picked up larger errors as compared to the voltage estimation errors of the **RT_substation** case. This observation is seen as a result of the detected offset in the voltage measurements at Busbar 3. It was found that utilising the correct voltage measurements of **RT_load_PQ** study by the state estimator renders less voltage estimation errors than the **RT_substation** case.

Additionally, it was found that the application of real-time power injection measurements of the load increases the accuracy of the estimated voltage angles. Figure 5.14 illustrates a comparison of the estimation errors of the voltage angles for the aforementioned case studies.

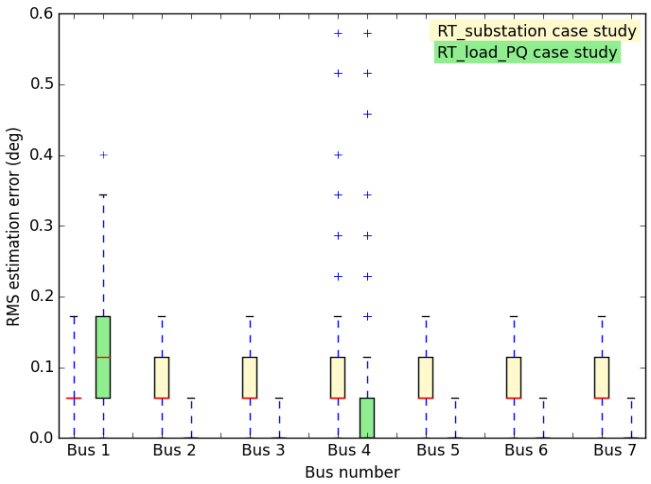


Figure 5.14 Comparison of the absolute errors of the estimated voltage angles

Figure 5.14 shows that the errors of the estimated voltage angles were approximately equal to zero in the **RT_load_PQ** case study. The increase in the angle estimation errors at Busbar 1 was because of sudden dips in the real-time measured active and reactive power injections at Busbar 3, which are shown in Figure 5.15. These dips were not removed in the filtering process described in **(5.4.2**) because the exclusion criterion was based on the phase voltage magnitudes and not active and reactive power measurements.

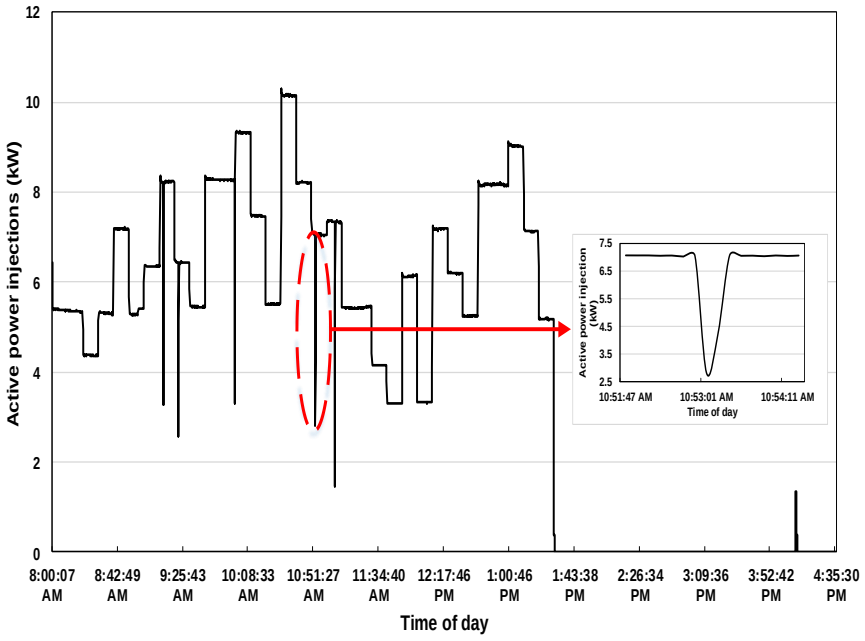


Figure 5.15 Dips in the power injections of the load

Figure 5.16 illustrates the distributions of the estimation errors of the active and reactive power injections.

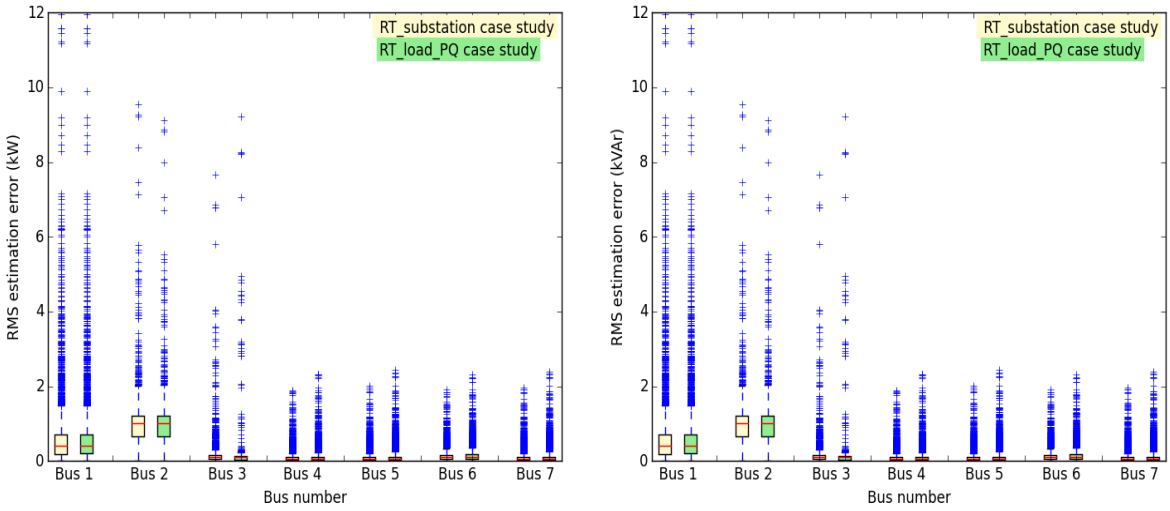


Figure 5.16 Comparison of the absolute errors of the estimated power injections (a) estimated active power errors and (b) estimated reactive power errors

Because of the biased voltage measurements at the load busbar, the estimation errors of the injected active powers, Figure 5.16(a), at Busbars 1 through 7 had approximately the same values in both **RT_substation** and **RT_load_PQ** case studies.

The error distribution of the estimated reactive power injections in both **RT_substation** and **RT_load_PQ** case studies attained similar trends. The introduction of more real-time measurements to the state estimator implies better quality state estimates. However, the offset in the real-time measurements of the load adversely affected the accuracy of the state estimates of the network.

5.5 Integrated Load and State Estimation

The performance of the IRWLS state estimator was investigated in Section **(5.4)** using actual measurements collected from an LV microgrid. In this study, the IRWLS estimator was applied to evaluate the operating states of the IEEE 33 bus test network using both real-time and pseudo measurements derived from smart metering data. The smart meter measurements were collected from the Irish smart metering CBT [45]. Pseudo measurements were provided by the k-means based load estimation algorithm **(Chapter 4)**.

For the present case study, the following assumptions were made;

a. The IEEE 33 bus test system represents an urban residential MV network,

b. A total number of 384 residential customers [166] are connected to the LV side of each MV/LV transformer; and

c. The active and reactive power consumption of customers was monitored using smart meters deployed to their premises.

d. A set of real-time measurements was available at the 12.66kV side of the substation transformer. The voltage magnitude, active and reactive powers injections at Busbar 1 and active and reactive power flows at the sending end of Feeder section 1, connecting Busbars 1 and 2, were assumed to be measured.

The following approach was adopted to construct the daily load profiles of each MV/LV transformer

1. Initially, 384 residential customers were randomly allocated to the LV side of a secondary substation.

2. The daily load profiles of all the customers were aggregated at each half hour time step to create the aggregated daily load profile.

3. A random percentage in the range of 6-10% [182], [183] of the daily consumption was added, to each aggregated daily load profile, to reflect the aggregated power losses of service cables and MV/LV transformer.

4. Aggregated daily load profiles allocated to different secondary substations were normalised, with regard to their maximum value, and then scaled up to match the P injection entries illustrated in Table 5.1.

5. The same approach illustrated in **Step 4** was repeated but with regard to the Q injection entries. The daily load profile of the reactive power at the MV side of each MV/LV transformer was constructed accordingly.

6. The aggregated daily load profiles of the active and reactive power injections were constructed over a period of 8 days extending from 20th until the end of 27th July 2009.

7. The aggregated daily profiles of the first seven days, 20th July until 26th July 2009, were segmented in the [16, 24] even range of hours. This was in accordance with the findings of **Chapter 4**. The segmented profiles were applied to train the k-means based load estimation algorithm. Two sets of cluster centres were

obtained. The first set included the cluster centres of the active power injections, whereas The second set included the cluster centres of the reactive power injections.

8. The eighth day (27th July) was used to test the performance of the integrated load and state estimation algorithm. The aggregated profiles, constructed in **Step 1** through **Step 8**, of the test period were utilised by the IRWLS estimator.

9. Under normal operating conditions, it was assumed that the half-hourly measurements, of the aggregated profiles, were delivered in real-time to the state estimator. Under contingent conditions of communication failure, pseudo measurements were used to replace the missing measurements of smart meters. Pseudo measurements were obtained from the k-means based load estimation algorithm (Cluster centres from **Step 8** following the approach of **Chapter 4**).

The estimation errors between the base case and estimated operating states of the network were quantified in terms of the Mean Absolute Percentage Error (MAPE). Equation (5.12) defines the MAPE [184].

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\mathrm{x}_{base\ case}(t) - \mathrm{x}_{estimated}(t)}{\mathrm{x}_{base\ case}(t)}\right| \times 100 \qquad \textbf{Equation 5.8}$$

given that x is the state variable defined in terms of the voltage magnitude and voltage phase angles. Pypower [185], a MATPOWER [186] based power system simulation package was used to obtain the base case operating state of the 33 bus network through the application of Newton-Raphson load flow analysis.

The accuracy of the estimated states of the network was investigated for different durations of pseudo measurements. At the same time, different segmentation time windows of the cluster centres were applied to estimate the required pseudo measurements. Figure 5.17 illustrates the MAPE profile of the estimated voltage magnitude at each busbar of the network. The black line denotes the error profile of the estimated busbar voltages when the duration of pseudo measurements was one hour, whereas the light grey line shows the MAPE profile when the duration of pseudo measurements increases to 24 hours. In both cases, the nearest cluster centre, with a segmentation time window that was equal to 16 hours, was applied to estimate the pseudo measurements.

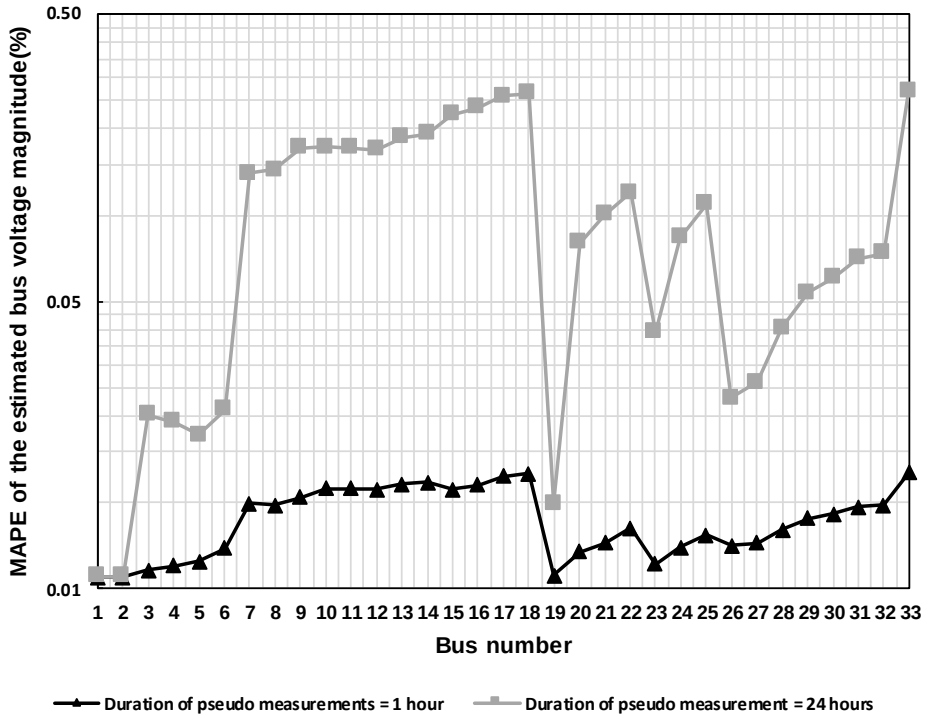


Figure 5.17 MAPE of the estimated busbar voltage magnitudes

Figure 5.17 shows that one hour of pseudo measurements results in an error, of the estimated voltage magnitudes, that varies between 0.01-0.03%, and that the error increases to be nearly 0.5% when the duration of pseudo measurements increases to 24 hours. The maximum errors of the estimated voltage magnitudes were observed at Busbars 18 and 33. These busbars are located at the end of the feeder. The availability of real-time voltage measurements only at the substation (Busbar 1) results in that the estimated voltages establish high errors at the busbars furthest to the substation.

The distribution of the MAPE, of the estimated voltage magnitudes at Busbar 33, for different segmentation time windows of cluster centres is shown in Figure 5.18. The increase in the segmentation the time windows used to estimate the pseudo measurements results in larger voltages estimation errors. A conclusion was therefore made that the segmentation time window of 16 hours is the best time window to estimate the lost measurement and provide the required pseudo measurements to the state estimator.

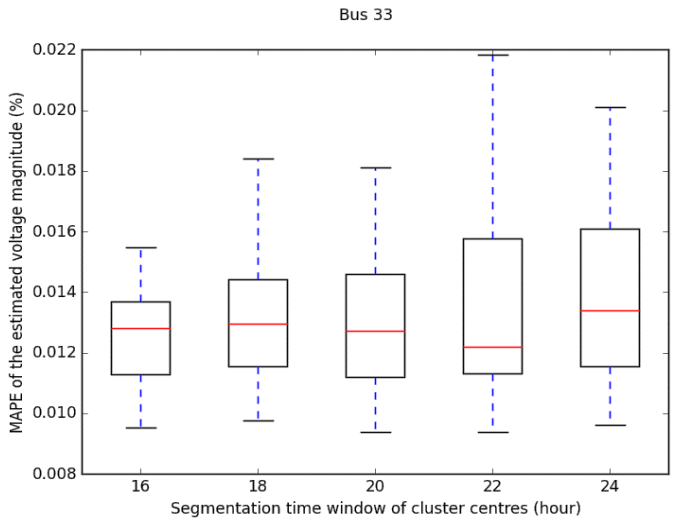


Figure 5.18 MAPE distribution of estimated voltage for different segmentation time windows – Busbar 33

Figure 5.19 shows the error distribution of the estimated voltages at Busbar 33 for different durations of pseudo measurements; a segmentation time window of 16 hours was applied to estimate the pseudo measurements.

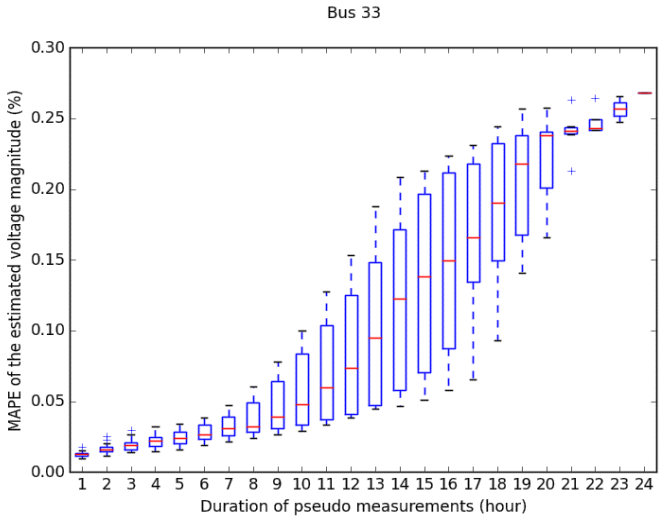


Figure 5.19 MAPE distribution of estimated voltage for different durations of pseudo measurements – Busbar 33

Figure 5.19 shows that up to 7 hours of pseudo measurements result in an estimation error that was approximately equal to 0.05%. The increase in the durations of pseudo measurements increases the voltage estimation errors.

Figure 5.20 shows the MAPE distribution of the estimated active power injections at Busbar 33. A time window of 16 hours of past measurements was employed to calculate the required pseudo measurements. Up to 10 hours of pseudo measurements yield approximately 10% errors in the estimated active power injections at Busbar 33. In other words, the active power injections at Busbar 33 were estimated with an

approximate accuracy of 90% when 10 hours of pseudo measurements were applied to the state estimator. The MAPE distribution of the reactive power injections at Busbar 33 is illustrated in Figure 5.21.

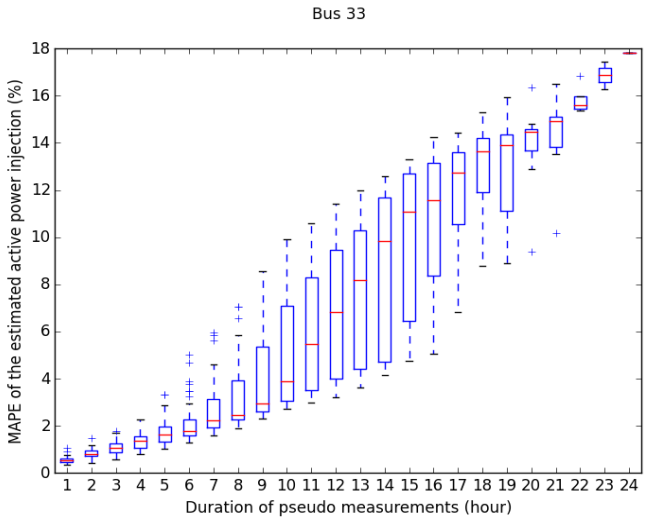


Figure 5.20 MAPE distribution of estimated active power injections for different durations of missing measurements – Busbar 33

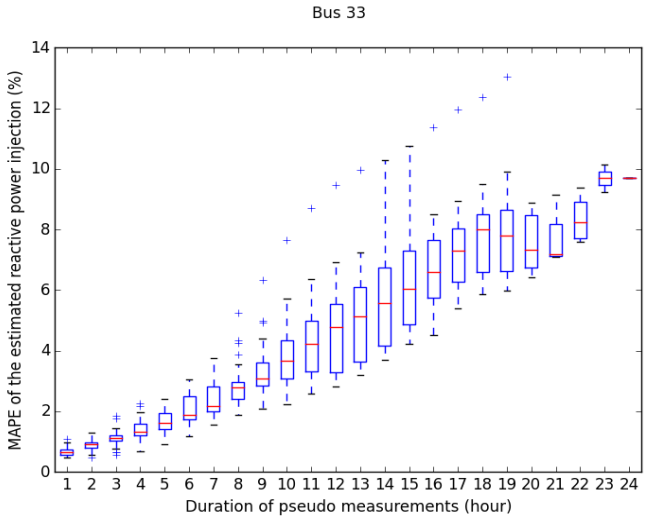


Figure 5.21 MAPE distribution of estimated reactive power injections for different durations of missing measurements – Busbar 33

For the same range of 1-10 hours of pseudo measurements, the estimated reactive powers injected at Busbar 33 attained an error of approximately 6%. When the duration of pseudo measurements increases to 24 hours, the estimation errors of the power injections – at Busbar 33 – were approximately 18% for the active power and 10% for the reactive power.

5.6 Conclusions

In this chapter, an integrated load and state estimation algorithm was developed to extend the observability of medium and low voltage distribution networks.

The k-means cluster analysis method was used to establish the basis of the load estimation algorithm. Canberra distance and segmented cluster centres in the range of 16-24 hours were applied in **Chapter 5** to provide pseudo measurements to the state estimator.

The state estimator was capable of accurately estimating the operating state(s) of the networks when different sets of measurements were applied. The estimation results, of the RSE LV microgrid, reveal that gross measurement errors were removed; and that the bias in measurement devices was detected when redundant real-time measurements were available. Accurate state estimates were produced when minimum sets of real-time measurements were applied. The availability of real-time measurements of the load was found to improve the accuracy of the estimated state of the network.

The obtained results show that the developed load and state estimation algorithm can be integrated with distribution management systems to extend the observability of distribution networks. The voltage phasors and active and reactive power injections were estimated accurately at each busbars of the IEEE 33 busbars MV test network. The efficacy of the developed load and state estimation algorithm was investigated using aggregated smart meter measurements and pseudo measurements – that were obtained through the developed load estimator. Taking into consideration the availability of real-time smart meter measurements, the segmentation time window – that was used to estimate the lost measurements and produce the required pseudo measurements – and the duration of pseudo measurements, several case studies were carried out.

Chapter 6 Conclusions and Future Work

6.1 Conclusions

6.1.1 Cluster analysis of smart meter measurements

In Chapter 3, an algorithm based on the k-means cluster analysis method was developed and used to extract consumption patterns from smart meter measurements. Cluster centres represent the patterns that were extracted. The cluster analysis algorithm was applied to examine the grouping of load profiles of individual (at the 0.433kV low voltage level) and aggregated (at the 11kV medium voltage level) smart meters (see Figure 3.1).

The algorithm was applied in order to cluster the daily and segmented load profiles of the smart meters. Whereas a daily load profile consists of 48 half-hourly measurements, a segmented load profile extends over a time window that is less than or equal to 24 hours (or 48 half-hourly steps). The 2, 4, 6, 8, …, 24 hours' time windows were used, on a rolling basis, to create the segmented load profiles (see Figure 3.5).

A set of training load profiles was applied to train the algorithm and obtain the cluster centres. The mean value of the Average Euclidean Distance (AvED) was calculated between the training load profiles and their cluster centres. The mean AvED and the AvED threshold, that was defined as a 1-10% of the mean consumption of the training load profiles, were used to determine the required number of the clusters.

A set of test load profiles was applied to test and validate the algorithm. The AvED, between the test load profiles and cluster centres that were obtained as a result of clustering the training load profiles, was applied to allocate each test load profile to its nearest cluster. The overall maximum absolute difference and mean AvED were used to quantify the differences between the test load profiles and their nearest training cluster centres.

Three weeks of load profiles, that were collected from 96 randomly selected domestic smart meters, were used in this study. One week of load profiles was used as a training set whereas two weeks of load profiles were used as a test set.

The results show that:

- The daily load profiles of aggregated smart meters were grouped, by the cluster analysis algorithm, into weekdays and weekend clusters when the

AvED threshold was varied between 8-10% of the mean consumption of the training load profiles.

- The cluster analysis algorithm grouped the daily load profiles of aggregated smart meters into weekdays, Saturday, and Sunday clusters when the AvED threshold was equal to 7% of the mean consumption of the training load profiles.

- The daily load profiles of aggregated smart meters were not grouped into separate weekdays and weekend clusters when the AvED threshold was relaxed from 1-6% of the mean consumption of the training load profiles.

- The daily load profiles of a limited number of individual smart meters were clustered, by the algorithm, into weekdays, Saturday, and Sunday clusters when AvED threshold was relaxed from 1-10% of the mean consumption of the training profiles.

- For the majority of individual smart meters, each daily load profile was the centre of its own cluster despite relaxing the AvED threshold from 1-10% of the mean consumption of the training profiles.

- The 2-hours was the best time window of segmenting the load profiles of individual and aggregated smart meters. Clustering the segmented profiles showed that the smallest number of clusters, per number of input profiles of the training period, was obtained for the 2-hour segmented load profiles.

- The smallest classification differences between the segmented test load profiles, of individual and aggregated smart meters, and their nearest training cluster centres were obtained for the 2-hour segmented load profiles.

Based on the results that were obtained, it was concluded that:

- Clustering the daily load profiles of aggregated smart meters into weekdays and weekends is feasible through the application of the cluster analysis algorithm that was developed. This is achieved when the AvED threshold is relaxed to be in the range of 8-10% of the mean consumption.

- For the same 8-10% values of AvED threshold, clustering the daily load profiles of individual smart meters into weekdays and weekend is infeasible. The randomness and variability of the power consumption, of an LV

customer, explain this conclusion. In order to cluster the daily profiles of individual smart meters into weekdays and weekends, the range of AvED threshold would be in the range of 150-200% of the mean consumption.

- The segmented load profiles of individual and aggregated smart meters show greater degrees of similarity at the shortest, 2-hour, segmentation time window. This means that the clustering of the 2-hour segmented load profiles produces fewer and denser clusters than the clustering of other segmented load profiles.

6.1.2 Load estimation of smart meter measurements

In Chapter 4, a load estimation algorithm based on the k-means cluster analysis method was developed and used to estimate missing and future measurements of smart meters. The load estimation algorithm applied the cluster centres, of previously clustered smart meter measurements, in order to estimate the measurements of individual and aggregated smart meters.

The load estimation algorithm investigated two approaches in order to produce the required load estimates. In the first approach, load estimation using daily cluster centres, 48 half-hourly measurements (the measurement to be estimated plus 47 half-hourly measurements that precede it) were paired with the nearest cluster centre that was obtained as a result of clustering the training daily load profiles (see Figure 4.2).

In the second approach, rather than the application of daily cluster centres, segmented cluster centres (with 2, 4, 6, …, 24 segmentation time windows) were used to estimate the missing measurements. In this manner, given that r is the segmentation time window (in hours), $2r$ half-hourly measurements (one measurement to be estimated plus the $2r - 1$ measurements that precede it) were paired with the nearest segmented cluster whose length is $2r$ half-hourly measurements (see Figure 4.3).

Different distance functions were applied in order to pair the load profiles (that had missing measurements) with the nearest cluster centres. Canberra, Manhattan, Pearson correlation and Euclidean distance functions (see Table 4.1) were compared in terms of producing load estimates with the smallest estimation errors. The estimation errors, between the actual and estimated smart meter measurements, were quantified using the overall maximum absolute error and the mean value of root-mean square error.

One week of training load profiles, that were collected from 96 randomly selected domestic smart meters, was applied to train the algorithm and obtain the cluster centres.

An AvED threshold that was equal to 1% of the mean consumption of the training load profiles, was used to determine the required number of the clusters. One week of test load profiles was used in order to test the load estimation algorithm. For each test daily profile, the durations of lost measurements from 1 to 24 consecutive hours were simulated. The measurements were estimated iteratively, i.e. only one half-hourly measurement was estimated at a time.

The results that were obtained show that:

- The load estimation error, of estimating the missing measurements of individual smart meters, was within 10-50% of the mean daily consumption of the training load profiles. This load estimation error was obtained when Pearson correlation distance function was used (see Figure 4.11).

- The load estimation error, of estimating the missing measurements of aggregated smart meters, was within 10% of the mean daily consumption of the training load profiles. This load estimation error was obtained when Canberra distance function was used (see Figure 4.19).

- The smallest errors, of estimating the measurements of individual and aggregated smart meters, were obtained when the load estimation algorithm used the segmented cluster centres (see Figures 4.7 and 4.15). As compared to the application of daily cluster centres to estimate the load, the application of segmented cluster centres resulted in smaller load estimation errors.

- The smallest load estimation errors were obtained when segmented cluster centres in the range of 16-24 hours were applied to estimate the missing measurements of individual and aggregated smart meters (see Figure 4.17). The largest load estimation errors were observed when the 2-hours' segmented cluster centres were used to estimate the missing measurements. Load estimation errors decreased as longer segmented cluster centres were applied.

The following conclusions were drawn from the results that were obtained:

- Pearson correlation is the best distance function to use, in the load estimation algorithm, when the load of individual smart meters is to be estimated.

- Canberra is the best distance function to be used in the load estimation algorithm when the load of aggregated smart meters is to be estimated.

- The application of segmented cluster centres in the range between 16-24 hours results in significantly smaller load estimation errors than the application of daily cluster centres. This also means that the utilisation of the past 16 hours of measurements, to estimate any missing half-hourly measurement, produces higher accuracy load estimates than the application of a day-before measurements.

6.1.3 Integrated load and state estimation

In Chapter 5, an integrated load and state estimation algorithm was developed and used to estimate the operating state of a distribution network. The estimation algorithm was based on the k-means based cluster analysis load estimator and the Iteratively Re-Weighted Least Squares (IRWLS) state estimator.

The IRWLS state estimator was tested using measurements that were collected from a low voltage (LV) microgrid (see Figure 5.1). The set of measurements included real-time measurements (where metering devices were installed) and pseudo-measurements (these were assumed to be obtained from low accuracy line power flow measurements).

Real-time measurements comprised busbar voltage magnitudes, active and reactive line power flows, and busbar active and reactive power injections. Real-time power injections of a busbar were assumed to be equal to the line power flows (that were measured along the feeder connecting that specific busbar to the microgrid).

Pseudo-measurements were the active and reactive power injections at each busbar of the microgrid. The values of the pseudo-measurements were equal to the line power flows, but were assumed to have half the initial weight that was assigned to real-time power injection measurements.

Three case studies were implemented to assess the accuracy of the outputs of the state estimator (voltage magnitude, voltage phase angle and active and reactive power injections at each busbar of the microgrid). The first case study simulated the availability of a complete set of real-time measurements at all busbars and feeders of the microgrid. The second case study simulated the availability of real-time measurements only at the LV side of the substation. The third case study simulated the availability of real-time measurements at the LV side of the substation and at the load busbar.

The outputs of the state estimator for the first case study were assumed to be the exact operating state of the microgrid. The estimation errors between the outputs of the

state estimator (for each of the second and third case studies) and the exact operating state (first case study) were quantified using the RMSE.

The performance of the integrated load and state estimation was investigated using the IEEE 33 bus medium voltage test network (see Figure 5.2). An assumption was made that the network is an urban domestic distribution network. An overall number of 384 domestic smart meters [166] were connected to each MV/LV secondary substation. Feeder, service lines and transformer power losses were taken into account by adding 6-10% [182], [183] of the consumption to each half hour in a daily load profile. The aggregated daily load profiles, of the 384 smart meters, were scaled to the specified active and reactive power injection (see Table 5.1).

The aggregated load profiles were segmented in the range 16-24 hours. The segmented load profiles of the active and reactive powers were separately clustered. The k-means based load estimation algorithm applied segmented cluster centres in conjunction with Canberra distance function in order to estimate the missing measurements of aggregated smart meters. The durations between 1-24 hours of missing measurements were simulated.

Real-time (aggregated smart meter) measurements and pseudo measurements, busbar active and reactive power injections that were obtained through load estimation algorithm, were input to the IRWLS state estimator in order to estimate the operating state of the 33 bus network. The exact operating state was obtained by running load flow. The state estimation errors, between the state estimator outputs and the exact operating state of the 33 bus network, were quantified in terms of the Mean Absolute Percentage Error (MAPE).

The results show that:

- The IRWLS state estimator was capable of removing gross errors in the real-time measurements. When a set of redundant real-time measurements was available for the state estimator to use, gross measurement errors and offsets in the real-time measurements were detected and removed.

- The availability of real-time measurements at the load increased the accuracy of the estimated state of the LV microgrid. The installation of real-time active and reactive power injections of the load produced nearly zero errors in the estimated voltage phase angles at each busbar of the LV microgrid. The estimation errors of the voltage magnitude were not reduced as a result of the biased voltage measurements at the load.

- Accurate operating states of the IEEE 33 bus network were estimated when the integrated load and state estimation algorithm was used. A maximum error (of the MAPE of the estimated voltage magnitude) that was equal to 0.013% was observed at the end of feeder busbar, Busbar 33, with one hour of pseudo measurements. When the duration of pseudo measurement was increased to 24 hours, the MAPE of the estimated voltage magnitude was approximately 0.3% at Busbar 33.

The following conclusions were drawn:

- The deployment of real-time measurements at the load (busbar and feeder section), reduces the errors of the estimated voltage magnitude and voltage phase angle at other busbars of a distribution network.

- Smart meter measurements, both real-time and pseudo-measurements, that are used in conjunction with state estimation have the capability to extend the observability of a distribution network.

- The integrated load and state estimation algorithm estimates the busbar voltage magnitude and voltage phase angle with small estimation errors. Therefore, the algorithm has the potential to be applied in distribution network monitoring and operation.

6.2 Recommendations for future research

6.2.1 Cluster analysis of smart meter measurements

The proposed clustering algorithm was based on the k-means cluster analysis method. Although the k-means is one of the most widely used clustering methods, it does not provide the global optimal solution; only a sub-optimal solution is produced. Therefore, it is suggested to investigate other clustering methods that have the potential to find a global optimal solution.

At the same time, the k-means method is a non-intelligent method (i.e., does not have the capability of incremental learning from the sets of inputs and outputs). Nowadays, the integration of more DERs into LV distribution networks renders the power consumption of individual households more dynamic and volatile. Therefore, it is necessary to develop the present algorithm using artificial intelligence (AI) methods. In this way, the new clustering algorithm will have the capability to learn during the training

phase. The clustering efficiency is expected to increase accordingly and more accurate load estimates are anticipated.

In the present research, an adaptive approach was adopted to define the required number of clusters. The researcher recommends the development of a smart approach to define the essential number of clusters. The "smart k-means" method will initially scan the input profiles. Based on the proximity – i.e., distances between the profiles - the number of clusters can be determined.

The training period included the load profiles of individual smart meters that were collected over one week. This period of measurements was selected because of computational limitations. The researcher strongly recommends exploring the use the multi-core multi-processor supercomputers. The clustering will be facilitated and larger sets of profiles can be clustered. Additionally, dynamic sets of measurements are also suggested. In this sense, dynamic refers to an up to date set of smart meter measurements.

6.2.2 Load estimation of smart meter measurements

The load estimation algorithm was developed using the principles of pattern recognition. Distance functions and cluster centres were used to estimate the half-hourly measurements of smart meters. The selected distance functions are the most widely reported in literature. The researcher recommends investigating the performance of alternative distance functions. Artificial intelligence techniques encompass a variety of distance functions which can be applied to improve the developed load estimation algorithm.

On the other hand, all the measurements preceding the one estimated were assigned equal weights in the proposed algorithm. The researcher strongly recommends applying a variable weight scheme to estimate the measurements. Thus, the most recent measurements preceding the missing one will be assigned high weights. The weights will decrease as the previous measurements move far from the missing one. The missing measurement is the same as the estimated one.

6.2.3 Integrated load and state estimation

An integrated load and state estimation algorithm was designed by establishing the link between the developed k-means based load estimation and the IRWLS state estimator.

Large estimation errors were encountered when the steady state conditions of the network were not yet met. The researcher suggests improving the IRWLS estimator

to take into consideration the steady state and transient operating conditions of the network.

Alternatively, the IRWLS estimator utilises an exact network model. In this case, the network parameters are assumed to be 100% accurate. The present work can be improved by considering the uncertainty in network parameters. Investigating the impact of the feeder resistance and reactance upon state estimates is recommended.

Finally, a closed-loop smart load and state estimation algorithm is suggested. An AI based cluster analysis method can be employed in the new smart algorithm. The outputs of the state estimator [10] can be fed back to the clustering algorithm. As a result, the cluster centres will be updated periodically. High accuracy load estimates (pseudo measurements) are anticipated.

Based on the results of the integrated load and state estimation algorithm, the researcher strongly recommends developing a meter placement algorithm. The algorithm will consider the placement of additional meters to increase the accuracy of the state estimates and compensate for unexpected loss of smart meter measurements.

6.3 Research contributions

In summary, the contributions of this research can be highlighted as

- The development of a clustering algorithm based on the k-means cluster analysis method to group similar consumption and extract characteristic load profiles, defined in terms of the cluster centres, from these patterns.

- The development of a load estimation algorithm, based on the k-means cluster analysis method and principles of pattern recognition, to estimate missing and future measurements of smart meters and provide the required pseudo measurements to state estimation algorithms.

- The development of an integrated load and state estimation algorithm, capable of using limited MV real-time and LV smart meter measurements, to extend the observability of MV distribution networks.

[10] Estimated active and reactive powers

Appendix A.1 Illustration of a box-whisker plot

A box-whisker plot [187] presents and compares a set of data (load profiles for example) without making any assumptions about the statistical distribution of the data. For a given dataset, the following approach is adopted to construct the box-whisker plot.

1. Sort the data ascendingly and find the median;
2. Determine the median of the lower-half and upper-half of data. the median of the lower half is the first quartile (Q1), whereas the median of the upper half is the third quartile (Q3).
3. Calculate the inter-quartile range

   $$Inter\ quartile\ range = IQR = Q3 - Q1$$

4. Determine the upper and lower whiskers

   $$Upper\ whisker = Q3 + (1.5 \times IQR)$$

   $$Lower\ whisker = Q1 - (1.5 \times IQR)$$

5. Mark any values greater than the upper whisker or less than the lower whisker as outliers.

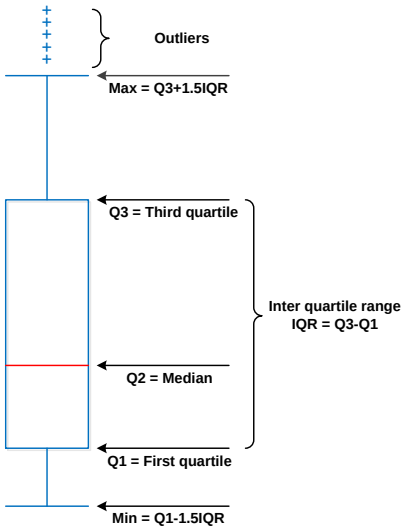


Figure A.1 Illustration of box-whisker plot

Appendix A.2 IRWLS reweighting principles

IRWLS estimators [150] minimise the sum of weighted squared residuals between the measured and the estimated values of the network states, as shown in equation (1), subject to the constraints given by the measurement equations shown in equation (2):

$$\min \sum_{i=1}^{m} \left(\mathrm{z}_i - \mathrm{f}_i(\mathrm{x})\right)^2 \mathrm{W}_i \quad \textbf{(1)}$$

$$\text{Subject to } \mathbf{z} = \mathbf{f}(\mathbf{x}) + \mathbf{e} \quad \textbf{(2)}$$

where $\mathbf{z}$ is the measurements vector, $\mathbf{f}$ is the vector of nonlinear measurement functions, $\mathbf{x}$ is the vector of system state variables, $\mathbf{e}$ is the vector of measurement errors, W_i is the weight of the i^{th} measurement, i is the measurement index, and m is the number of the measurements.

Equation (1) is formulated in the same way as the WLS estimator resulting in equation (3)

$$\min \left(\mathbf{z} - \mathbf{f}(\mathbf{x})\right)^T \mathbf{W} \left(\mathbf{z} - \mathbf{f}(\mathbf{x})\right) \quad \textbf{(3)}$$

given that $\mathbf{W}$ is the weight matrix which is a diagonal matrix such that the weights of the measurements are in the main diagonal. The weight, that is equal to the reciprocal of the variance of a measurement $\left(1/\sigma_i^2\right)$, reflects the accuracy of the measurement.

Measurements are normalised with respect to their standard deviations

$$\mathbf{s} = \frac{\mathbf{z}}{\boldsymbol{\sigma}} = \left[\frac{\mathrm{z}_1}{\sigma_1}, \frac{\mathrm{z}_2}{\sigma_2}, \cdots, \frac{\mathrm{z}_m}{\sigma_m}\right] \quad \textbf{(4)}$$

In a general form, $\mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x})/\boldsymbol{\sigma}$, i.e. for each i, $\mathrm{h}_{\boldsymbol{i}}(\mathbf{x}) = \mathrm{f}_i(\mathbf{x})/\sigma_{\boldsymbol{i}}$. Consequently, the normalised residual between the i^{th} measurement and its calculated value is defined as:

$$\mathrm{R}_{\boldsymbol{i}} = |\mathrm{s}_i - \mathrm{h}_i(\mathbf{x})| \quad \textbf{(5)}$$

The diagonal elements of the weight matrix, W_i, are modified iteratively (with iteration count α) according to the relationship

$$W_i^{\alpha} \propto \frac{1}{R_i^{\alpha}} \quad \textbf{(6)}$$

To avoid convergence problems that might result from the division by very small residuals, the range of weights was limited between minimum and maximum weight thresholds. The minimum weight threshold was set to 0.001 while the maximum weight threshold was unity. Equations (7) – (9) define the criteria of measurement re-weighting.

$$R_{\boldsymbol{max}}^{\boldsymbol{k}} = \max_i\left(R_i^k\right) \quad \textbf{(7)}$$

If $R_i^k > 0.001\ R_{\boldsymbol{max}}^{\boldsymbol{k}}$ then:

$$W_i^k = 0.001\ R_{\boldsymbol{max}}^{\boldsymbol{k}}/R_i^k \quad \textbf{(8)}$$

else:

$$W_i^k = 1 \quad \textbf{(9)}$$

Appendix A.3 DER – TF LV microgrid

The single line diagram of the DER-TF LV microgrid is shown in Figure A.2.

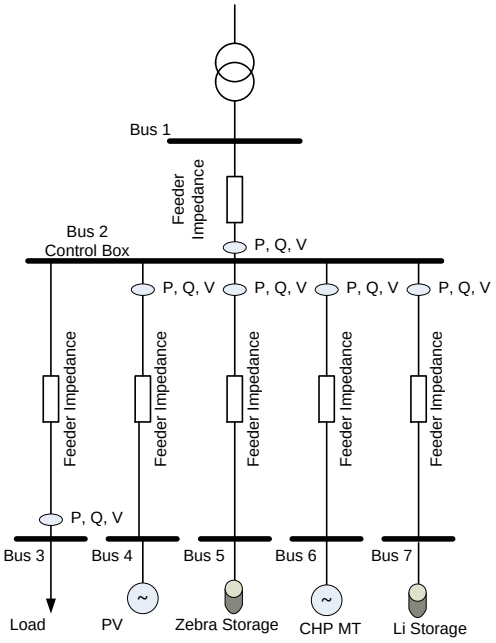


Figure A.2 Single line diagram of the RSE DER-TF

The line parameters are listed in Table A.1

Table A.1 Network parameters of RSE DER-TF LV microgrid

Feeder number	*From Busbar*	*To Busbar*	*Feeder length (km)*	*Resistance (Ω/km)*	*Reactance (Ω/km)*
1	1	2	0.104	0.081	0.079
2	2	3	0.065	0.160	0.084
3	2	4	0.375	0.500	0.073
4	2	5	0.065	0.200	0.084
5	2	6	0.050	0.200	0.084
6	2	7	0.065	0.160	0.084

The DERs that were available for this project include:

- A programmable resistive and inductive load (93 kW, 69 kVAr at 400 V); load can be set independently on different phases (active power 3x31 kW, D=1kW; reactive power 3x23 kVAr, D=0.75 kVAr); load was connected to Feeder 2;

- A PV array with a peak active power of 18 kW and power factor close to 1; connected to Feeder 3;
- A Zebra Storage System (150kVA Inverter and two Zebra batteries), connected to Feeder 4, with the following characteristics:
    - Maximum active power during battery discharge: 30 kW;
    - Maximum active power during battery charge: 20 kW;
    - Reactive power: 0 – 60 kVAr (inductive or capacitive);
- A Gas CHP Micro Turbine with a controllable active power from 50 to 100 kW and power factor from 0.8 to 1 (inductive and capacitive); connected to Feeder 5;
- A Lithium Storage System (100kVA Inverter and 18 module Lithium batteries), connected to Feeder 6, with the following characteristics:
    - Maximum active power during battery discharge: 40 kW;
    - Maximum active power during battery charge: 20 kW;
    - Reactive power: 0 – 36 kVAr (inductive or capacitive).

Appendix A.4 IEEE 33 Bus MV distribution network

The single line diagram of the IEEE 33 bus MV distribution network is shown in Figure A.3.

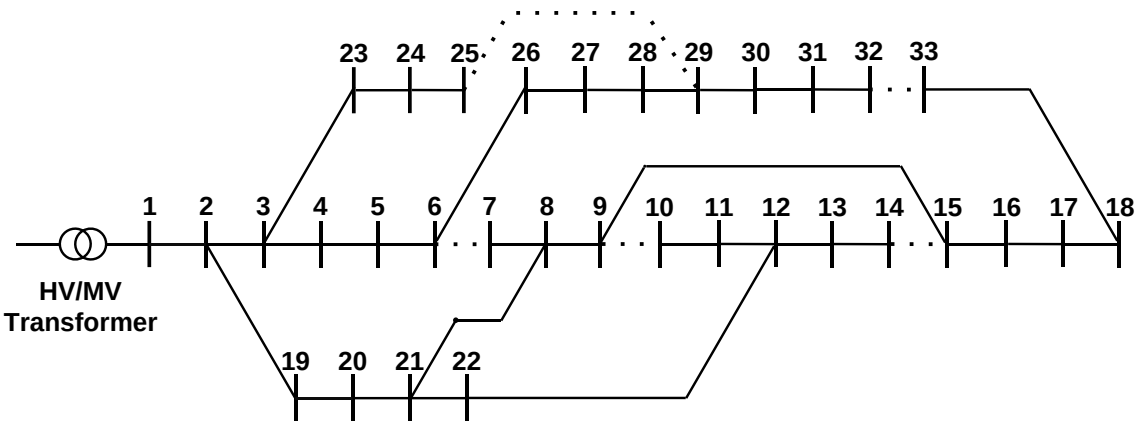


Figure A.3 Single-line diagram of the IEEE 33 bus MV network

Table A.2 lists the network parameters of the network whose rated voltage is 12.66kV

Table A.2 Network parameters of IEEE 33 Bus network

Feeder number	*From Busbar*	*To Busbar*	*Resistance (Ω)*	*Reactance (Ω)*
1	1	2	0.0922	0.0470
2	2	3	0.4930	0.2512
3	3	4	0.3661	0.1864
4	4	5	0.3811	0.1941
5	5	6	0.8190	0.7070
6	8	7	0.7114	0.2351
7	8	9	1.0300	0.7400
8	11	10	0.1966	0.0650
9	12	11	0.3744	0.1238
10	12	13	1.4680	1.1550
11	13	14	0.5416	0.7129
12	15	16	0.7463	0.5450
13	16	17	1.2890	1.7210
14	17	18	0.7320	0.5740
15	2	19	0.1640	0.1565
16	19	20	1.5042	1.3554
17	20	21	0.4095	0.4784
18	21	22	0.7089	0.9373
19	3	23	0.4512	0.3083
20	23	24	0.8980	0.7091
21	24	25	0.8960	0.7011
22	6	26	0.2030	0.1034
23	26	27	0.2842	0.1447
24	27	28	1.0590	0.9337
25	28	29	0.8042	0.7006
26	29	30	0.5075	0.2585
27	30	31	0.9744	0.9630
28	31	32	0.3105	0.3619
29	21	8	2.0000	2.0000
30	9	15	2.0000	2.0000
31	22	12	2.0000	2.0000
32	18	33	0.5000	0.5000
34	6	7	0.1872	0.6188
35	9	10	1.0440	0.7400

References

[1] United Nations Framework Convention of Climate Change, “Kyoto Protocol.” [Online]. Available: http://unfccc.int/kyoto_protocol/items/2830.php. [Accessed: 04-Feb-2016].

[2] DECC, “UK Renewable Energy Roadmap,” 2011. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/48128/2167-uk-renewable-energy-roadmap.pdf. [Accessed: 30-Mar-2016].

[3] European Commission, “2020 Climate & Energy Package.” [Online]. Available: http://ec.europa.eu/clima/policies/strategies/2020/index_en.htm. [Accessed: 04-Feb-2016].

[4] European Commission, “Overview of Europe 2020 Targets,” 2014. [Online]. Available: http://ec.europa.eu/europe2020/pdf/targets_en.pdf. [Accessed: 04-Apr-2016].

[5] N. Jenkins, C. Long, and J. Wu, “An Overview of the Smart Grid in Great Britain,” *Engineering*, vol. 1, no. 4, pp. 413–421, 2015.

[6] G. T. Heydt, “The Next Generation of Power Distribution Systems,” *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 225–235, 2010.

[7] J. A. Momoh, *Smart Grid: Fundamentals of Design and Analysis*. John Wiley & Sons, 2012.

[8] J. Ekanayake, N. Jenkins, K. Liyanage, J. Wu, and A. Yokoyama, *Smart Grid: Technology and Applications*. John Wiley & Sons, 2012.

[9] R. Singh, B. C. Pal, and R. B. Vinter, “Measurement Placement in Distribution System State Estimation,” *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 668–675, 2009.

[10] R. Singh, B. C. Pal, and R. A. Jabr, “Statistical Representation of Distribution System Loads Using Gaussian Mixture Model,” *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 29–37, 2010.

[11] F. Schweppe and J. Wildes, “Power System Static-State Estimation, Part I: Exact Model,” *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no.

1, pp. 120–125, 1970.

[12] J. Wu, Y. He, and N. Jenkins, “A Robust State Estimator for Medium Voltage Distribution Networks,” *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1008–1016, 2013.

[13] European Smart Metering Industry Group, “The 20-20-20 Goals,” 2012. [Online]. Available: http://esmig.eu/page/20-20-20-goals. [Accessed: 30-Mar-2016].

[14] DECC, “Delivering UK Energy Investment,” 2014. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/331071/DECC_Energy_Investment_Report.pdf. [Accessed: 17-Feb-2016].

[15] DCC, “Building a smart metering network for Great Britain,” 2015. [Online]. Available: https://www.smartdcc.co.uk/media/338770/15574_building_a_smart_metering_network_v3.pdf. [Accessed: 28-Feb-2016].

[16] DECC, “Smart Meters , Smart Data , Smart Growth,” 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/397291/2903086_DECC_cad_leaflet.pdf. [Accessed: 27-Aug-2015].

[17] S. Firth and J. Palmer, “The Potential for Smart Meters in a National Household Energy Survey,” 2013. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/275488/smart_meters_and_a_national_energy_survey.pdf. [Accessed: 30-Mar-2016].

[18] Bristol Smart Energy City, “Bristol Smart Energy City Collaboration Appendix B: Technical - Data and IT,” 2015. [Online]. Available: https://bristol-smart-energy.cse.org.uk/wiki/B:_Technical_-_Data_and_IT#T1:_Application_scales_for_datasets. [Accessed: 21-Feb-2016].

[19] DECC, “Smart Metering Implementation Programme,” 2013. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/267393/Smart_Metering_Summary_Plan.pdf. [Accessed: 04-Feb-2016].

[20] DECC, “Appendix E - DCC USER INTERFACE SERVICES SCHEDULE.” [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/41

6318/Annex_F_-_User_Interface_Services_Schedule__SEC_Appendix_E__-_March_2015_-__Amended__Conclusions_.pdf. [Accessed: 18-Feb-2016].

[21] World Energy Council, “2015 Energy Trilemma Index Benchmarking the sustainability of national energy systems,” 2015. [Online]. Available: https://www.worldenergy.org/wp-content/uploads/2015/11/20151030-Index-report-PDF.pdf. [Accessed: 04-Feb-2016].

[22] DECC, “Smart Metering Implementation Programme Smart Metering Equipment Technical Specifications Version 1.59,” 2014. [Online]. Available: http://www.eua.org.uk/sites/default/files/SMIP E2E SMETS2 V1 57 CLEAN version for notification.pdf. [Accessed: 30-Mar-2016].

[23] Beama, “Consumer Access Devices Applications for data in the Consumer Home Area Network (C HAN) and wider market considerations,” 2014. [Online]. Available: http://www.beama.org.uk/resourceLibrary/consumer-access-devices-a-beama-guide.html. [Accessed: 04-Feb-2016].

[24] DECC, “Consultation on amending Smart Meter In-home Display Licence Conditions,” 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/448900/IHD_Licence_Condition_Consultation_Final_for_publication_2707.pdf. [Accessed: 30-Mar-2016].

[25] DECC, “In-Home Display Licence Conditions: Consultation response,” 2016. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/497078/IHD_Policy_Framework_Licence_Conditions_Post_Consultation_Decisions_Final_for_Publication.pdf. [Accessed: 30-Mar-2016].

[26] C. Sastry, R. G. Pratt, V. Srivastava, and S. Li, “Use of Residential Smart Appliances for Peak-Load Shifting and Spinning Reserves Cost / Benefit Analysis,” 2010. [Online]. Available: http://www.pnnl.gov/main/publications/external/technical_reports/PNNL-20110.pdf. [Accessed: 30-Mar-2016].

[27] S. Nistor, “Residential Demand Response in The Power System,” Cardiff University, 2015.

[28] DECC, “Smart Metering Implementation Programme Communications Hub

Technical Specifications Version 1.46," 2014. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/381536/SMIP_E2E_CHTS.pdf. [Accessed: 17-Feb-2015].

[29] Zigbee Alliance, "ZigBee Smart Energy," 2016. [Online]. Available: http://www.zigbee.org/zigbee-for-developers/applicationstandards/zigbeesmartenergy/#features. [Accessed: 16-Feb-2016].

[30] DECC, "Award of Smart Meters DCC Licence." [Online]. Available: https://www.gov.uk/government/news/award-of-smart-meters-dcc-licence. [Accessed: 26-Aug-2015].

[31] R. Tafazolli, "Smart Metering System for the UK Technologies review," 2013. [Online]. Available: https://m2m.telefonica.com/system/files_force/SM_Report_07-06-2013 2_9.pdf?download=1. [Accessed: 27-Aug-2015].

[32] Telefónica, "Telefónica UK Signs £1.5bn Smart Meter Deal." [Online]. Available: http://news.o2.co.uk/?press-release=telefonica-uk-signs-1-5bn-smart-meter-deal. [Accessed: 26-Aug-2015].

[33] Ofcom, "Measuring Mobile Broadband Performance in the UK 4G and 3G network performance," 2014. [Online]. Available: http://stakeholders.ofcom.org.uk/binaries/research/broadband-research/mbb-nov14.pdf. [Accessed: 30-Aug-2015].

[34] Ofcom, "Infrastructure Report The first Communications Infrastructure Report," 2012. [Online]. Available: http://stakeholders.ofcom.org.uk/binaries/research/telecoms-research/infrastructure-report/Infrastructure-report2012.pdf. [Accessed: 30-Aug-2015].

[35] Arqiva, "Arqiva signs £625 million contract to provide smart metering communications service." [Online]. Available: http://www.arqiva.com/news/press-releases/arqiva-signs-625-million-contract-to-provide-smart-metering-communications-service/. [Accessed: 26-Aug-2015].

[36] Sensus, "Sensus T866 MicroRTU™ Techical Specifications," 2015. [Online]. Available: http://sensus.com/documents/10157/34791/DDS-10002. [Accessed:

27-Aug-2015].

[37] Sensus, “FlexNet,” 2015. [Online]. Available: http://sensus.com/web/uk/communications/product-line/uk-flexnet/learn-more. [Accessed: 27-Aug-2015].

[38] DCC, “Smart Meter Key Infrastructure , Infrastructure Key Infrastructure and DCC Key Infrastructure,” 2015. [Online]. Available: Smart Meter Key Infrastructure , Infrastructure Key Infrastructure and DCC Key Infrastructure. [Accessed: 18-Feb-2016].

[39] CGI Group, “UK DECC Selects CGI as DSP for GB SMIP,” 2013. [Online]. Available: http://www.cgi-group.co.uk/uk-department-of-energy-and-climate-change. [Accessed: 17-Feb-2016].

[40] DECC, “SEC4A and Transitional Arrangements: government conclusions and further consultation text.” [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/380831/SEC4A_transitional_arrangements_gov_conclusions_and_further_consultation_text_comment.pdf. [Accessed: 27-Aug-2015].

[41] Ofgem, “The GB gas distribution network.” [Online]. Available: https://www.ofgem.gov.uk/gas/distribution-networks/gb-gas-distribution-network. [Accessed: 18-Feb-2016].

[42] National Grid, “Distribution Network Operator (DNO) Companies.” [Online]. Available: http://www2.nationalgrid.com/UK/Our-company/Electricity/Distribution-Network-Operator-Companies/. [Accessed: 18-Feb-2016].

[43] Gemserv, “Smart Energy Code,” 2016. [Online]. Available: https://www.smartenergycodecompany.co.uk/docs/default-source/sec-documents/smart-energy-code-4.8/sec-4-8---10th-february-2016.pdf?sfvrsn=5. [Accessed: 22-Feb-2016].

[44] DECC, “Smart Metering Implementaion Programme Government Response to the Consultation on New Smart Energy Code Content,” 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/416318/Annex_F_-_User_Interface_Services_Schedule__SEC_Appendix_E__-_March_2015_-__Amended__Conclusions_.pdf. [Accessed: 15-Feb-2016].

[45] Commission for Energy Regulation, “Electricity Smart Metering Technology Customer Behaviour Trials (CBT) Findings Report,” *Appendices to Information Paper CER11080a*, 2011. [Online]. Available: http://www.cer.ie/docs/000340/cer11080(a)(i).pdf. [Accessed: 16-Jul-2015].

[46] Commission for Energy Regulation, “Electricity Smart Metering Technology Customer Behaviour Trials (CBT) Findings Report,” *Appendices to Information Paper CER11080ai*, 2011. [Online]. Available: http://www.cer.ie/docs/000340/cer11080(a)(ii).pdf. [Accessed: 01-Sep-2015].

[47] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, “Smart meters for power grid: Challenges, issues, advantages and status,” *Renewable and Sustainable Energy Reviews*, vol. 15, no. 6, pp. 2736–2742, 2011.

[48] Commission for Energy Regulation, “Electricity Smart Metering Technology Trials Findings Report,” *Information Paper CER11080b*, 2011. [Online]. Available: https://www.ucd.ie/t4cms/Electricity Smart Metering Technology Trials Findings Report.pdf. [Accessed: 01-Sep-2015].

[49] Ofcom, “Infrastructure Report 2014,” *Ofcom’s second full analysis of the UK’s communications infrastructure*, 2014. [Online]. Available: http://stakeholders.ofcom.org.uk/binaries/research/infrastructure/2014/infrastructure-14.pdf.

[50] DECC, “Smart meter rollout for the domestic sector (GB) Impact Assessment,” 2011. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/42740/1485-impact-assessment-smart-metering-implementation-p.pdf. [Accessed: 24-Feb-2016].

[51] DECC, “Smart meter roll-out for the non-domestic sector (GB) Impact Assessment,” 2012. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/48805/4907-smart-meter-rollout-non-domestic-ia-resp.pdf. [Accessed: 24-Feb-2016].

[52] A. Carvallo and J. Cooper, *The Advanced Smart Grid Edge Power Driving Sustainability*. Artech House, 2015.

[53] Jack Danahy, “That Smart Grid Data Surge We Mentioned Earlier? You Can’t

Ignore It," 2009. [Online]. Available: http://www.smartgridnews.com/story/smart-grid-data-surge-we-mentioned-earlier-you-can-t-ignore-it/2009-11-03. [Accessed: 21-Feb-2016].

[54] Landis+Gyr, "How to handle big data from energy to information," *Pathway 02*, 2012. [Online]. Available: http://www.landisgyr.eu/newsletters-archive/pathway-02-english/. [Accessed: 22-Feb-2016].

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

[56] F. McLoughlin, A. Duffy, and M. Conlon, "Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study," *Energy and Buildings*, vol. 48, pp. 240–248, 2012.

[57] M. J. Zaki and W. Meira Jr., *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press, 2014.

[58] D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*. Massachusets Institute of Technology, 2001.

[59] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts And Techniques*. Elsevier, 2011.

[60] M. Brown, "Data mining techniques," 2012. [Online]. Available: http://www.ibm.com/developerworks/library/ba-data-mining-techniques/. [Accessed: 24-Feb-2016].

[61] S. Ramos, J. M. Duarte, F. J. Duarte, and Z. Vale, "A data-mining-based methodology to support MV electricity customers' characterization," *Energy and Buildings*, vol. 91, pp. 16–25, 2015.

[62] G. J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, "Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1120–1128, 2007.

[63] I. P. Panapakidis, T. A. Papadopoulos, G. C. Christoforidis, and G. K. Papagiannis, "Pattern recognition algorithms for electricity load curve analysis of buildings," *Energy and Buildings*, vol. 73, pp. 137–145, 2014.

[64] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, “Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data,” *Applied Energy*, vol. 87, no. 11, pp. 3538–3545, 2010.

[65] M. Espinoza, C. Joye, R. Belmans, and B. De Moor, “Short-Term Load Forecasting , Profile Identification , and Customer Segmentation : A Methodology Based on Periodic Time Series,” *IEEE Transactions on Power Systems*, vol. 20, no. 3, pp. 1622–1630, 2005.

[66] R. Granell, C. J. Axon, and D. C. H. Wallom, “Clustering disaggregated load profiles using a Dirichlet process mixture model,” *Energy Conversion and Management*, vol. 92, pp. 507–516, 2015.

[67] I. Dent, U. Aickelin, and T. Rodden, “Application of a clustering framework to UK domestic electricity data,” in *11th Annual Workshop on Computational Intelligence (UKCI)*, 2011, pp. 1–6.

[68] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, “Development of Low Voltage Network Templates—Part I: Substation Clustering and Classification,” *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3036–3044, 2015.

[69] J. P. Gouveia and J. Seixas, “Unraveling electricity consumption profiles in households through clusters: combining smart meters and door-to-door surveys,” *Energy and Buildings*, vol. 116, pp. 666–676, 2016.

[70] K. A. Heller and Z. Ghahramani, “Bayesian hierarchical clustering,” in *Proceedings of the 22nd International Conference on Machine Learning*, 2005, pp. 297–304.

[71] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*. Springer-Verlag, 2005.

[72] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, 2001.

[73] S. V. Verdu, M. O. Garcia, F. J. G. Franco, N. Encinas, A. G. Marin, A. Molina, and E. G. Lazaro, “Characterization and identification of electrical customers through the use of self-organizing maps and daily load parameters,” in *IEEE PES Power Systems Conference and Exposition*, 2004, pp. 1240–1247.

[74] S. Kaski, “Data exploration using self-organizing maps,” Helsinki University of

Technology, 1997.

[75] D. Roussinov and H. Chen, “A scalable self-organizing map algorithm for textual classification: A neural network approach to thesaurus generation,” *Communication Cognition and Artificial Intelligence (CC-AI)*, vol. 15, no. 1–2, pp. 81–111, 1998.

[76] N. Allinson, H. Yin, L. Allinson, and J. Slack, *Advances in Self-Organising Maps*. Springer-Verlag, 2001.

[77] K. Le Zhou, S. L. Yang, and C. Shen, “A review of electric load classification in smart grid environment,” *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 103–110, 2013.

[78] S. Bandyopadhyay and S. Saha, *Unsupervised Classification: Similarity Measures, Classical and Metaheuristic Approaches, and Applications*. Springer-Verlag, 2013.

[79] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, “An Electric Energy Consumer Characterization Framework Based on Data Mining Techniques,” *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 596–602, 2005.

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

[81] Y.-I. Kim, J.-H. Shin, J.-J. Song, and I.-K. Yang, “Customer clustering and TDLP (typical daily load profile) generation using the clustering algorithm,” in *IEEE Transmission & Distribution Conference & Exposition: Asia and Pacific*, 2009, pp. 1–4.

[82] I. P. Panapakidis, M. C. Alexiadis, and G. K. Papagiannis, “Electricity customer characterization based on different representative load curves,” in *9th International Conference on the European Energy Market (EEM)*, 2012, pp. 1–8.

[83] S. Ramos, J. M. Duarte, F. J. Duarte, Z. Vale, and P. Faria, “A data mining framework for electric load profiling,” in *IEEE PES Conference on Innovative Smart Grid Technologies Latin America (ISGT LA)*, 2013, pp. 1–6.

[84] I. Panapakidis, M. Alexiadis, and G. Papagiannis, “Evaluation of the performance of clustering algorithms for a high voltage industrial consumer,” *Engineering*

Applications of Artificial Intelligence, vol. 38, pp. 1–13, 2015.

[85] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, “Load Pattern-Based Classification of Electricity Customers,” *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1232–1239, 2004.

[86] H.-A. Cao, C. Beckel, and T. Staake, “Are domestic load profiles stable over time? An attempt to identify target households for demand side management campaigns,” in *39th Annual Conference of the IEEE Industrial Electronics Society (IECON)*, 2013, pp. 4733–4738.

[87] J. Kwac, J. Flora, and R. Rajagopal, “Household Energy Consumption Segmentation Using Hourly Data,” *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420–430, 2014.

[88] F. Rodrigues, J. Duarte, V. Figueiredo, Z. Vale, and M. Cordeiro, *A Comparative Analysis of Clustering Algorithms Applied to Load Profiling*. Springer-Verlag, 2003.

[89] J. J. López, J. A. Aguado, F. Martín, F. Muñoz, A. Rodríguez, and J. E. Ruiz, “Hopfield–K-Means clustering algorithm: A proposal for the segmentation of electricity customers,” *Electric Power Systems Research*, vol. 81, no. 2, pp. 716–724, 2011.

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

[91] G. Chicco, R. Napoli, and F. Piglione, “Comparisons Among Clustering Techniques for Electricity Customer Classification,” *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 933–940, 2006.

[92] J. D. Rhodes, W. J. Cole, C. R. Upshaw, T. F. Edgar, and M. E. Webber, “Clustering analysis of residential electricity demand profiles,” *Applied Energy*, vol. 135, pp. 461–471, 2014.

[93] M. Koivisto, M. Degefa, M. Ali, J. Ekström, J. Millar, and M. Lehtonen, “Statistical modeling of aggregated electricity consumption and distributed wind generation in distribution systems using AMR data,” *Electric Power Systems Research*, vol. 129, pp. 217–226, 2015.

[94] A. Mutanen, M. Ruska, S. Repo, and P. Jarventausta, “Customer Classification and Load Profiling Method for Distribution Systems,” *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1755–1763, 2011.

[95] M. Koivisto, P. Heine, I. Mellin, and M. Lehtonen, “Clustering of Connection Points and Load Modeling in Distribution Systems,” *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1255–1265, 2013.

[96] T. Räsänen, J. Ruuskanen, and M. Kolehmainen, “Reducing energy consumption by using self-organizing maps to create more personalized electricity use information,” *Applied Energy*, vol. 85, no. 9, pp. 830–840, 2008.

[97] G. Chicco, R. Napoli, and F. Piglione, “Application of clustering algorithms and self organising maps to classify electricity customers,” in *IEEE Power Tech Conference Proceedings*, 2003, vol. 1, pp. 373–379.

[98] I. Benítez, A. Quijano, J. L. Díez, and I. Delgado, “Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers,” *International Journal of Electrical Power and Energy Systems*, vol. 55, pp. 437–448, 2014.

[99] F. Martinez Alvarez, A. Troncoso, J. C. Riquelme, and J. S. Aguilar Ruiz, “Energy Time Series Forecasting Based on Pattern Sequence Similarity,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 8, pp. 1230–1243, 2011.

[100] D. Gerbec, S. Gasperic, and F. Gubina, “Determination and allocation of typical load profiles to the eligible consumers,” in *IEEE Power Tech Conference Proceedings*, 2003, vol. 1, pp. 368–372.

[101] Y.-I. Kim, J.-M. Ko, and S.-H. Choi, “Methods for generating TLPs (typical load profiles) for smart grid-based energy programs,” in *IEEE Symposium on Computational Intelligence Applications In Smart Grid (CIASG)*, 2011, pp. 1–6.

[102] P. R. S. Jota, V. R. B. Silva, and F. G. Jota, “Building load management using cluster and statistical analyses,” *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 8, pp. 1498–1505, 2011.

[103] I. P. Panapakidis, M. C. Alexiadis, and G. K. Papagiannis, “Deriving the optimal number of clusters in the electricity consumer segmentation procedure,” in

Proceedings of the 10th International Conference on the European Energy Market (EEM), 2013, pp. 1–8.

[104] S. M. Bidoki, N. Mahmoudi-Kohan, M. H. Sadreddini, M. Zolghadri Jahromi, and M. P. Moghaddam, “Evaluating different clustering techniques for electricity customer classification,” in *IEEE PES Transmission and Distribution Conference and Exposition*, 2010, pp. 1–5.

[105] H. Xue, Q. Jia, N. Wang, and Z. Bo, “A dynamic state estimation method with pmu and scada measurement for power systems,” *Power Engineering*, vol. 5, pp. 848–853, 2007.

[106] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, and F. J. G. Franco, “Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps,” *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1672–1682, 2006.

[107] R. Lamedica, L. Santolamazza, G. Fracassi, G. Martinelli, and A. Prudenzi, “A novel methodology based on clustering techniques for automatic processing of MV feeder daily load patterns,” in *Power Engineering Society Summer Meeting*, 2000, vol. 1, pp. 96–101.

[108] C. H. Jin, G. Pok, Y. Lee, H.-W. Park, K. D. Kim, U. Yun, and K. H. Ryu, “A SOM clustering pattern sequence-based next symbol prediction method for day-ahead direct electricity load and price forecasting,” *Energy Conversion and Management*, vol. 90, pp. 84–92, 2015.

[109] M. E. Baran, L. A. A. Freeman, F. Hanson, and V. Ayers, “Load Estimation for Load Monitoring at Distribution Substations,” *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 164–170, 2005.

[110] H. Liao and D. Niebur, “Load profile estimation in electric transmission networks using independent component analysis,” *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 707–715, 2003.

[111] N. I. A. Tawalbeh, “Daily load profile and monthly power peaks evaluation of the urban substation of the capital of Jordan Amman,” *International Journal of Electrical Power & Energy Systems*, vol. 37, no. 1, pp. 95–102, 2012.

[112] L. J. Soares and M. C. Medeiros, “Modeling and forecasting short-term electricity

load: A comparison of methods with an application to Brazilian data," *International Journal of Forecasting*, vol. 24, no. 4, pp. 630–644, 2008.

[113] R. Ramanathan, R. Engle, C. W. J. Granger, F. Vahid-Araghi, and C. Brace, "Short-run forecasts of electricity loads and peaks," *International Journal of Forecasting*, vol. 13, no. 2, pp. 161–174, 1997.

[114] K. Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: From big data to big insights," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 215–225, 2016.

[115] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 1352–1372, 2015.

[116] S. A. Soliman and A. M. Al-Kandari, *Electrical Load Forecasting*. Butterworth-Heinemann, 2010.

[117] F. McLoughlin, A. Duffy, and M. Conlon, "Evaluation of time series techniques to characterise domestic electricity demand," *Energy*, vol. 50, pp. 120–130, Feb. 2013.

[118] H. Willis, A. Schauer, J. D. Northcote-green, and T. Vismor, "Forecasting Distribution system Loads Using Curve Shape Clustering," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 4, pp. 893–901, 1983.

[119] J. Wan and K. Miu, "Weighted least squares methods for load estimation in distribution networks," *Power Systems, IEEE Transactions on*, vol. 18, no. 4, pp. 1338–1345, 2003.

[120] H. Wang and N. N. Schulz, "A load modeling algorithm for distribution system state estimation," in *IEEE PES Transmission and Distribution Conference and Exposition*, 2001, vol. 1, pp. 102–106.

[121] J. Wu and Y. Yu, "CBR-based Load Estimation for Distribution Networks," in *IEEE Mediterranean Electrotechnical Conference MELECON*, 2006, pp. 952–955.

[122] E. Manitsas, R. Singh, B. Pal, and G. Strbac, "Distribution System State Estimation Using an Artificial Neural Network Approach for Pseudo Measurement Modeling," *IEEE Transaction on Power Systems*, vol. 27, no. 4, pp. 1–9, 2012.

[123] A. P. Sakis Meliopoulos and F. Zhang, “Multiphase power flow and state estimation for power distribution systems,” *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 939–946, 1996.

[124] C. Gómez-Quiles, A. Gómez-Exposito, and A. De La Villa Jaén, “State estimation for smart distribution substations,” *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 986–995, 2012.

[125] Y. Deng, Y. He, and B. Zhang, “A branch-estimation-based state estimation method for radial distribution systems,” *IEEE Transactions on Power Delivery*, vol. 17, no. 4, pp. 1057–1062, 2002.

[126] I. Roytelman and S. M. Shahidehpour, “State estimation for electric power distribution systems in quasi real-time conditions,” *IEEE Transactions on Power Delivery*, vol. 8, no. 4, pp. 2009–2015, 1993.

[127] C. N. Lu, J. H. Teng, and W. E. Liu, “Distribution system state estimation,” *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 229–240, 1995.

[128] M. E. Baran and A. W. Kelley, “State estimation for real-time monitoring of distribution systems,” *IEEE Transactions on Power Systems*, vol. 9, no. 3, pp. 1601–1609, 1994.

[129] E. Manitsas, R. Singh, B. Pal, and G. Strbac, “Modeling of pseudo-measurements for distribution system state estimation,” in *IET-CIRED SmartGrids for Distribution Seminar*, 2008, pp. 1–4.

[130] R. Singh, B. C. Pal, R. a. Jabr, and R. B. Vinter, “Meter placement for distribution system state estimation: An ordinal optimization approach,” *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2328–2335, 2011.

[131] R. Singh, B. C. Pal, and R. a. Jabr, “Choice of estimator for distribution system state estimation,” *IET Generation, Transmission & Distribution*, vol. 3, no. 7, pp. 666–678, 2009.

[132] K. Samarakoon, J. Wu, J. Ekanayake, and N. Jenkins, “Use of delayed smart meter measurements for distribution state estimation,” *IEEE Power and Energy Society General Meeting*, vol. 15, 2011.

[133] R. Khorshidi, F. Shabaninia, and T. Niknam, “A new smart approach for state

estimation of distribution grids considering renewable energy sources," *Energy*, vol. 94, pp. 29–37, 2016.

[134] A. Angioni, T. Schlosser, F. Ponci, and A. Monti, "Impact of Pseudo-Measurements From New Power Profiles on State Estimation in Low-Voltage Grids," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 1, pp. 70–77, 2016.

[135] I. Dzafic, M. Gilles, R. A. Jabr, B. C. Pal, and S. Henselmeyer, "Real Time Estimation of Loads in Radial and Unsymmetrical Three-Phase Distribution Networks," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4839–4848, 2013.

[136] A. Abur and A. G. Exposito, *Power System State Estimation Theory and Implementation*. New York: CRC Press, 2004.

[137] Y.-F. Huang, S. Werner, J. Huang, N. Kashyap, and V. Gupta, "State Estimation in Electric Power Grids: Meeting New Challenges Presented by the Requirements of the Future Grid," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 33–43, 2012.

[138] G. N. Korres, N. D. Hatziargyriou, and P. J. Katsikas, "State estimation in Multi-Microgrids," *European Transactions on Electrical Power*, vol. 21, no. 2, pp. 1178–1199, 2011.

[139] M. S. Kurzyn, "Real-Time State Estimation for Large-Scale Power Systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 7, pp. 2055–2063, 1983.

[140] K. Li, "State estimation for power distribution system and measurement impacts," *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 911–916, 1996.

[141] I. Cobelo, A. Shafiu, N. Jenkins, and G. Strbac, "State estimation of networks with distributed generation," *European Transactions on Electrical Power*, vol. 17, no. 1, pp. 21–36, 2007.

[142] M. E. Baran and a. W. Kelley, "A branch-current-based state estimation method for distribution systems," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 483–491, 1995.

[143] W. M. Lin, J. H. Teng, and S. J. Chen, “A highly efficient algorithm in treating current measurements for the branch-current-based distribution state estimation,” *IEEE Transactions on Power Delivery*, vol. 16, no. 3, pp. 433–439, 2001.

[144] H. Wang and N. N. Schulz, “A Revised Branch Current-Based Distribution System State Estimation Algorithm and Meter Placement Impact,” *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 207–213, 2004.

[145] B. S. Everitt and A. Skrondal, *The Cambridge Dictionary of Statistics*. Cambridge University Press, 2010.

[146] P. J. Green, “Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives,” *Journal of the Royal Statistical Society*, vol. 46, no. 2, pp. 149–192, 1984.

[147] X. Zhou, R. Molina, F. Zhou, and A. K. Katsaggelos, “Fast iteratively reweighted least squares for LP regularized image deconvolution and reconstruction,” in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 1783–1787.

[148] M. R. Irving and M. J. H. Sterling, “Robust State Estimation for Large Scale Power Systems,” in *Proceedings of IEE Conference on Power System Monitoring and Control*, 1986, pp. 261–264.

[149] B. C. Clewer, M. R. Irving, and M. J. H. Sterling, “Robust state estimation in power systems using sparse linear programming,” *IEE Proceedings C - Generation, Transmission and Distribution*, vol. 132, no. 3, pp. 123–131, 1985.

[150] M. M. Smith, R. S. Powell, M. R. Irving, and M. J. H. Sterling, “Robust algorithm for state estimation in electrical networks,” *IEE Proceedings C - Generation, Transmission and Distribution*, vol. 138, no. 4, pp. 283–288, 1991.

[151] M. R. Irving and C. N. Macqueen, “Robust algorithm for load estimation in distribution networks,” *IEE Proceedings - Generation, Transmission and Distribution*, vol. 145, no. 5, pp. 499–504, 1998.

[152] J. Abonyi and B. Feil, *Cluster Analysis for Data Mining and System Identification*. Springer-Verlag, 2007.

[153] J. Kogan, *Introduction to Clustering Large and High-Dimensionl Data*. Cambridge University Press, 2007.

[154] P. D. Diamantoulakis, V. M. Kapinas, and G. K. Karagiannidis, “Big data analytics for dynamic energy management in smart grids,” *Big Data Research*, vol. 2, no. 3, pp. 94–101, 2015.

[155] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*. SIAM, 2007.

[156] J. A. Hartigan, *Clustering algorithms*. John Wiley & Sons Ltd., 1975.

[157] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*. John Wiley & Sons Ltd., 2011.

[158] C. C. Aggarwal and C. K. Reddy, *Data Clustering Algorithms and Applications*. Chapman and Hall/CRC, 2013.

[159] K. Y. Yeung, C. Fraley, A. Murua, A. E. Raftery, and W. L. Ruzzo, “Model-based clustering and data transformations for gene expression data,” *Bioinformatics*, vol. 17, no. 10, pp. 977–987, 2001.

[160] J. Valente de Oliveira and W. Pedrycz, *Advances in Fuzzy Clustering and its Applications*. John Wiley & Sons Ltd., 2007.

[161] E. Müller, S. Günnemann, I. Assent, and T. Seidl, “Evaluating Clustering in Subspace Projections of High Dimensional Data,” *Proceedings of the VLDB Endowment*, vol. 2, no. 1, pp. 1270–1281, 2009.

[162] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, “Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications,” in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1998, pp. 94–105.

[163] L. Parsons, E. Haque, and H. Liu, “Subspace clustering for high dimensional data: a review,” *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 90–105, 2004.

[164] K. Kailing, H. Kriegel, and P. Kroger, “Density-Connected Subspace Clustering for High-Dimensional Data,” in *Proceedings of the 4th SIAM International Conference on Data Mining (SDM)*, 2004, pp. 246–256.

[165] Irish Social Science Data Archive, “Irish Smart Metering Measurements.” [Online]. Available: http://www.ucd.ie/issda/data/commissionforenergyregulationcer/. [Accessed: 10-Mar-2016].

[166] S. Ingram, S. Probert, and K. Jackson, “The impact of small scale embedded generation on the operating parameters of distribution networks,” *PB Power, Department of Trade and Industry (DTI)*, 2003. [Online]. Available: http://webarchive.nationalarchives.gov.uk/20100919181607/http://www.ensg.gov.uk/assets/22_01_2004_phase1b_report_v10b_web_site_final.pdf. [Accessed: 30-Mar-2016].

[167] M. J. L. de Hoon, S. Imoto, J. Nolan, and S. Miyano, “Open source clustering software,” *Bioinformatics*, vol. 20, no. 9, pp. 1453–1454, 2004.

[168] E. Zivot and J. Wang, *Modeling Financial Time Series with S-PLUS*. Springer-Verlag, 2006.

[169] H.-S. Park, S. R. Thomas, and R. L. Tucker, “Benchmarking of Construction Productivity,” *Journal of Construction Engineering and Management*, vol. 131, no. 7, pp. 772–778, 2005.

[170] B. S. Duran and P. L. Odell, *Cluster Analysis: A Survey (Volume 100 of Lecture Notes in Economics and Mathematical Systems)*. Springer-Verlag, 1974.

[171] K. Jain, Anil and C. Dubes, Richard, *Algorithms for Clustering Data*. Prentice-Hall, Inc., 1988.

[172] J. Wu and N. Jenkins, “Self-adaptive and robust method for distribution network load and state estimation,” in *16th Power System Computation Conference (PSCC)*, 2008, pp. 1–6.

[173] C. Muscas, M. Pau, P. A. Pegoraro, and S. Sulis, “Effects of Measurements and Pseudomeasurements Correlation in Distribution System State Estimation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 12, pp. 2813–2823, 2014.

[174] K. B. Samarakoon, “Use of Smart Meters for Frequency and Voltage Control,” Cardiff University, 2012.

[175] A. Gómez-Exposito, A. Abur, A. de la Villa Jaén, and C. Gómez-Quiles, “A Multilevel State Estimation Paradigm for Smart Grids,” *Proceedings of the IEEE*, vol. 99, no. 6, pp. 952–976, 2011.

[176] G. Strbac, I. Konstantelos, and R. Moreno, “IET Report : Emerging Modelling

Capabilities for System Operations," 2015. [Online]. Available: http://www.theiet.org/sectors/energy/documents/modelling-6.cfm?type=pdf. [Accessed: 30-Mar-2016].

[177] L. L. Grigsby, *Power System Stability and Control*, Third. CRC Press, 2012.

[178] G. A. Taylor, M. R. Irving, N. Nusrat, R. Liao, and S. Panchadcharam, "Smart distribution network operation: Emerging techniques and standards," *IEEE Power and Energy Society General Meeting*, vol. 1, pp. 1–6, 2011.

[179] "RSE website." [Online]. Available: http://www.rse-web.it/laboratori/laboratorio/32. [Accessed: 01-Jan-2016].

[180] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Transactions on Power Delivery*, vol. 4, no. 2, pp. 1401–1407, 1989.

[181] DECC, "Smart Metering Implementation Programme: Smart Metering Equipment Technical Specifications Version 1.58," 2014. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/381535/SMIP_E2E_SMETS2.pdf. [Accessed: 30-Mar-2016].

[182] OfGem, "Electricity distribution losses: A consultation document," no. January, p. 55, 2003.

[183] Sohn Associates Limited, "Electricity Distribution System Losses Non-Technical Overview," 2009. [Online]. Available: https://www.ofgem.gov.uk/sites/default/files/docs/2009/05/sohn-overview-of-losses-final-internet-version.pdf. [Accessed: 20-Feb-2016].

[184] C. W. Chase Jr., *Demand-Driven Forecasting: A Structured Approach to Forecasting*. John Wiley & Sons, 2013.

[185] R. Lincoln and P. S. E. R. C. PSERC, "Pypower," 2015. [Online]. Available: https://pypi.python.org/pypi/PYPOWER. [Accessed: 22-Jan-2016].

[186] R. D. Zimmerman, C. E. Murillo Sánchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[187] R. McGill, J. W. Tukey, and W. A. Larsen, “Variations of Box Plots,” *The American Statistician*, vol. 32, no. 1, pp. 12–16, 1978.