Meaningful Information Extraction from IoT Measurements using Signal Information Processing

Georgia Elafoudi

Doctor of Philosophy

University of Strathclyde

2020

To my beloved parents Dimo and Angeliki,

Abstract

The advancements of modern technology have motivated researchers, business and governmental stakeholders to envision the idea of a unified network-based platform, known as the *Internet of Things*, that can interconnect devices and allow the bi-directional communication between relevant parties. The information exchanged through the Internet of Things can originate from a variety of fields, namely healthcare, environmental and infrastructure monitoring, transportation and logistics, smart grid and smart houses, which can produce a vast amount of diverse data in real time. One of the most important challenges, emerging from the implementation of the Internet of things, is the acquisition of meaningful information using the data derived from the various smart devices and sensors. Therefore, it is essential to identify suitable data processing and analysis approaches, able to adapt depending on the application-specific requirements, and consequently transform the available data into useful information that can be utilised by any various stakeholders.

The main focus of this research thesis was in the field of Non-Intrusive Load Monitoring (NILM) in order to provide solutions suitable for low resolution smart metering data, similar to the specifications of the smart meters selected for deployment from most utilities and governmental stakeholders. Through an extensive and up-to-date review of the NILM field presented in this thesis, it has been identified that only recently, researchers have focused on disaggregating using only active power aggregate data for feature extraction at low sampling rates. Therefore three unsupervised NILM methods were proposed as an outcome of this research, one using solely *Dynamic Time Warping* (DTW), a signal processing-based method, while the other two methods propose a combination of DTW and *k-means*, namely *DTW+kM* and *kDTW*, in order to address the computation complexity observed using the DTW-based method. The DTW+kM approach performs DTW for creating a library of appliance signatures and classification via clustering using k-means, and the *kDTW* is incorporating a DTW refinement post processing step in order to optimise the performance of the initial implementation. The proposed methods were evaluated and benchmarked against

various state-of-the-art NILM methods using the publicly available REDD [1] and REFIT [2, 3] datasets, and reported good performance.

Furthermore, the research presented in this thesis has investigated two other heterogeneous applications of the Internet of Things with respect to the emerging data challenge, and have proposed customised monitoring systems, and a variety of signal processing and machine learning approaches for analysing the corresponding data. More specifically, a prototype monitoring system was proposed for monitoring earthwork assets and preliminary findings reported in this thesis. In the context of visual content interaction using a video based eye tracking device, applicable in healthcare, computing, and even advertisement, the user's attention was evaluating using various statistical and signal processing methods, with the wavelet-based analysis being the best contestant for identifying features for extraction using pupil dilation and gaze fixation.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Lina Stankovic, for giving me the opportunity to investigate a very interesting field for my PhD research. Her guidance, support and patience during this long process have been invaluable and made the completion of this PhD feasible. Furthermore, I would like extend my thanks to my special advisor Dr. Vladimir Stankovic, for his helpful advice and input during my research.

With regards to the presented NILM work in Chapters 4 and 5, I would like to thank O. Parson for making his HMM code available, N. Batra, J. Kelly, and O. Parson for developing NILMTK and both B. Zhao and K. He from University of Strathclyde for providing their results, presented in their published works ([4–6]) for benchmarking purposes during during the implementation of the work presented in 4 and 5. Furthermore, I would like to thank Dr. J. Liao for her help and contribution as a co-author in the NILM oriented published work.

I would like to acknowledge the support of EPSRC (Engineering and Physical Sciences Research Council) for funding my PhD through their Doctoral Training Grant. Furthermore, I would like to thank S3C for funding the embankment monitoring project, presented in Chapter 6 and A. Stevenson, D. Lamont and R. Fletcher from Scottish Canals, for their contribution through their expert knowledge and geological survey and for allowing the use of Falkirk Wheel for the deployment. A special thanks would have to go to Prof. I. Andonovic for providing his advice and feedback during this project and J. Ralston from Freescale for providing hardware material and expert knowledge during the duration of the project

Most importantly, I would like to express my gratitude to Dr. C. Tachtatzis for his invaluable help and contribution throughout the embankment monitoring work, presented in Chapter 6. Without his firmware development, assistance throughout the hardware design and testing phase, the completion of this work would not have been feasible. Furthermore, I would like to thank both R. Fletcher from Scottish Canals and D. Murray from University of Strathclyde for assisting during the deployment of the

prototype system in Falkirk Wheel, and D. Murray for assisting with the collection of the initial data. A special thanks goes to the staff of the Mechanical and Electronics Workshop of the Department of Electronics and Electrical Engineering in University of Strathclyde for their assistance in producing the prototype casing and their help with the PCB printing process.

I would like to extend my thanks to the FP7 QoSTREAM project for funding my visual content interaction research, presented in Chapter 6, during my secondment at Florida Atlantic University (FAU) and FAU for accepting me as a visiting research scholar. I would like to express my sincere regards to Prof. H. Kalva for supervising me during this research project, for making available the equipment required for this research and for his important help and feedback. I would also like to thank both D. Pappusetty and Dr. V. Adzic, for their help during my research experiment using the video based eye-tracker, and all the people that participated in that experiment.

A special acknowledgement goes to my sister Christina, niece Georgia, and all my extended family and friends in Greece for the emotional support and their understanding for not being that social throughout the last years. I would like to thank my friends and colleagues at CIDCOM for the help and all the memorable moments, especially Hana, Jing, Jakub, Kamila, Xavier, and Krishna. Finally, my friends, Nena, Vangelis, Panos, Giannis, Danae, Nadia, Katerina, Tolis, and Alexandra, for always being there for me during the ups and downs of my PhD, and for making my everyday life enjoyable and full of happy memories.

Most importantly, I would like to thank my parents, Dimos and Angeliki, for their continuous support throughout this long process. Without them and their incredible love and understanding, I would not be able to to complete this PhD. This PhD thesis is dedicated to them.

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Georgia Elafoudi

List of Publications

A. G. Elafoudi, L. Stankovic, and V. Stankovic, "Power disaggregation of domestic smart meter readings using dynamic time warping," in *2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pp. 36–39, May 2014, Athens, Greece.

B. J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Power disaggregation for low-sampling rate data," in *2nd International Workshop on Non-Intrusive Load Monitoring (NILM)*, June 2014, Austin,Texas, USA.

C. J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Non-intrusive appliance load monitoring using low-resolution smart meter data," in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 535– 540, Nov 2014, Venice, Italy.

D. L. Stankovic, G. Elafoudi, C. Tachtatzis, I. Andonovic, G. Cassels, and A. Fickling, "Decision support system for proactive maintenance of earthworks assets," in *World Water Congress XV*, May 2015, Edinburgh, Scotland, UK.

E. G. Elafoudi, V. Stankovic, L. Stankovic, D. Pappusetti, and H. Kalva, "Evaluation of Signal Processing Methods for Attention Assessment in Visual Content Interaction", *in New Trends in Image Analysis and Processing – ICIAP 2015 Workshops: ICIAP 2015 International Workshops, BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, September 7-8, 2015, Proceedings,* (Cham), pp. 580–588, Springer International Publishing, 2015.

Author's Contribution to the Publications

This thesis's author describes her contribution to each of the papers listed in the List of Publications Section, as follows, and where these contributions are described in the thesis.

A. DTW Algorithmic research, development and testing of the proposed DTW- based method, paper writing, with supervisory input (algorithm design and paper writing) from other two authors. Contribution reported in Chapter 4.

B. DTW Algorithmic research, pre-processing, development and testing of the proposed DTW-based method, equally contribution with the first author for the producing HMM-based results and paper writing, with supervisory input from the other two authors. Contribution reported in Chapter 4.

C. DTW Algorithmic research, pre-processing, development and testing of the proposed DTW-based method, equally contribution with the first author for the producing HMM-based results and paper writing, with supervisory input from the other two authors. Contribution reported in Chapter 4.

D. Embankment monitoring background research, research for sensors and equipment used in the prototype, with supervisory input from the first, third and fourth authors, data pre-processing and analysis of the collected data (apart from Data communication section), paper writing, with supervisory input (data analytics and paper writing). Contribution reported in Chapter 6.

E. Experimental design and execution, background research, data pre-processing, data analysis and discussion, paper writing, with supervisory input (experimental design, analysis and paper writing) from second, third and fourth author (experimental design). Contribution reported in Chapter 6.

Georgia Elafoudi

Contents

Figures and Tables

Figures

Tables

Acronyms and abbreviations

AC	Air Conditioning
ACS-F1/ACS-F2	Appliance Consumption Signatures
AdaBoost	Adaptive Boost
ADS	Automated Driving System
AEFI	Actual Energy Fraction Index
AFAMAP	Additive Factorial Approximate Maximum A-Posteriori Probability
AIA	Artificial Immune Algorithm
AIC	Akaike information Criterion
AMI	Advanced Metering Infrastructure
AMPds	Almanac of Minutely Power Datasets
ANIALM	Adaptive Non-Intrusive Appliance Load Monitoring
ANN	Artificial Neural Network
Apps	Appliances
ARMA	Autoregressive Moving Average Model
AV	Autonomous Vehicle
B	Bathroom GFI
BIC	Bayesian Information Criterion
BLUED	Building Level fUlly Labeled Electricity Disaggregation
BMS	Building Management System
BP	Back-Propagation
BP-ANN	Back-Propagation Artificial Neural Network
BR	Base-load Removal
BSS	Blind Source Separation
BST	British Summer Time
CAN	Controlled Area Network
CCWT	Complex Continuous Wavelet Transform
CDM	Committee Decision Mechanisms
CMW	Complex Morlet Wavelet

CNN	Convolutional Neural Networks
CO	Combinational Optimisation
COOLL	Controlled On/Off Loads Library
COMBED	COMmercial Building Energy Dataset
CP	Constraint Programming
CURE	Clustering Using Representatives
CUSUM	Cumulative Sum
CWT	Continuous Wavelet Transform
dAE	De-noising Auto-encoders
DB	Daubechies
DBN	Dynamic Bayesian Network
DBSCAN	Density Based Spatial Clustering of Applications with Noise
DD	Duration-Dependent
DDoS	Distributed Denial of Service
DDSC	Discriminative Disaggregation Sparse Coding
dFCM	Dynamic Fuzzy c-Means
DFT	Discrete Fourier Transform
DNN	Deep Neural Networks
DoS	Denial of Service
DP	Dynamic Programming
DRED	Dutch Residential Energy Dataset
DT	Decision Tree
DTW	Dynamic Time Warping
DTW+kM	Combined DTW and k-means Method
DW	Dishwasher
DWT	Discrete Wavelet Tranform
E	Electronics
EA	Evolutionary Algorithm
ECG	Electrocardiogram
ECO	Electricity Consumption & Occupancy
EEFI	Estimated Energy Fraction Index
EIG	Eigenvalues
EMF	Electromagnetic Field

EMI	Electromagnetic Interference
EPRI	Electric Power Research Institute
ERP	Enterprise Resource Planning
F	Fridge
FHMM	Factorial Hidden Markov Model
FP-growth	Frequent Pattern Growth
FP-tree	Frequent Pattern Tree
FSC	Food Supply Chain
FFT	Fast Fourier Transform
FSM	Finite State Machines
FZ	Freezer
FZZ	Fridge-Freezer
GA-SVM-MKL	Genetic Algorithm Support Vector Machine Multiple Kernel Learning
GLR	Generalised Likelihood Ratio
GMM	Gaussian Mixture Model
GNB	Gaussian Naive Bayes
GP	Genetic Programming
GPRS	General Packet Radio Service
GPS	Global Positioning System
GRU	Gated Recurrent Unit
GSP	Graphical Signal Processing
GUI	Graphical User Interface
H	Heater
HA	Historical Average
HDFS	Hadoop Distributed File System
HDP-HSMM	Hierarchical Dirichlet Process Hidden Semi-Markov Model
HES	Household Electricity Use Study
HCI	Human Computer Interaction
HMM	Hidden Markov Model
HSID	Hybrid Signature-based Iterative Disaggregation
HSVM	Hierarchical SVM
HSVM	Hierarchical SVM
HVAC	Heating, Ventilation, and Air Conditioning
IAMs	Individual Appliance Monitors

IAW	Instantaneous Admittance Waveform
iAWE	Indian Dataset for Ambient Water and Energy
ICT	Information and Communications Technology
ID3	Iterative Dichotomiser 3
IhepcDS	Individual household electric power consumption Data Set
IHH	In-Home Healthcare
IID	Independent and Identically Distributed
IoT	Internet of Things
ISDTW	Iterative Subsequence Dynamic Time Warping
ISODATA	Iterative Self-Organising Data Analysis Technique
K	Kettle
kDTW	Combined DTW and k-means Method using DTW Refinement
k-NN	k-Nearest Neighbour
k-NNR	k-Nearest Neighbour Rule
KO	Kitchen Outlet
LI	Linear Interpolation
LLD-Max	Log Likelihood Ratio Detector with Maxima
LPF	Low-Pass Filter
LSTM	Long Short Term Memory
LUS	Least Unified Residue
LVQ	Learning Vector Quantisation
MAP	Maximum A Posteriori Probability
MAR	Multilevel Association Rules
MCMC	Markov-Chain Monte Carlo
MCO	Most Common Occurrence
MFCCs	Mel-frequency Central Coefficients
MLE	Maximum Likelihood Estimation
MLP	Multilayer Perceptron
MMS	Multimedia Messaging Service
MP	Matching Pursuit
MSE	Mean Square Error
M-SVM	Multiple-class SVM

MultiBoost	Multi-class AdaBoost
MW	Microwave
NaN	Not-a-Number
NDE	Normalised Disaggregation Error
NFC	Near Field Communication
NHTSA	National Highway Traffic Safety Administration
NILM	Non-Intrusive Appliance Monitoring
NMF	Non-negative Matrix Factorisation
OCTES	Opportunities for Communities Through Energy Storage
OGSA	Open Grid Service Architecture
PALDi	Particle Filter-Based Load Disaggregation
PARAFAC	PARAllel FACtor
PCA	Principal Component Analysis
PCB	Printed Circuit Board
PDT	Particle-Based Distribution Truncation
PF	Particle Filtering
PKNN	Probabilistic K-Nearest Neighbour
PLAID	Plug Load Appliance Identification Dataset
PNN	Probabilistic Neural Network
PSD	Power Spectral Density
PSO	Particle Swarm Optimisation
PVC	Polyvinyl Chloride
PZT	Piezoelectric Transducers
RAN	Radio Access Network
RBF	Radial Basis Function
RBSA	Residential Building Stock Assessment
RDD	Resilient Distributed Datasets
RECAP	Real Time Recognition and Profiling Appliances
RMSE	Root Mean Square Error
MFNN	Multi-layer Feed-forward Neural Network
REDD	Residential Energy Disaggregation Dataset
ReLU	Rectifier Linear Unit
RF	Radio Frequency

RFID	Radio Frequency Identification
RMS	Root Mean Square
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
S	Stove
SAE	Society of Automotive Engineers
SCADA	Supervisory Control And Data Acquisition
sdAE	Stacked De-noising Auto-encoders
SMF	Structured Mean Field
SIQP	Segmented Integer Quadratic Programming
SMPS	Switch Mode Power Supply
SOA	Service Oriented Architecture
STFT	Short-time Fourier Transform
STMF	Source Separation Via Tensor and Matrix Factorisations
StS	Steady-States
SVD	Singular Value Decomposition
SVM	Support Vector Machine
T	Toaster
TC	Temporal Correlation
TECA	Total Energy Correctly Assigned
TinyEARS	Tiny Energy Accounting and Reporting System
TPW	Transient Power Waveform
UK-DALE	UK Domestic Appliance-Level Electricity
UO	Unknown Outlets
VQ	Vector Quantisation
WD	Washer-Dryer
WM	Washing Machine
WMN	Wireless Mesh Networks
WMRA	Wavelet Multi-Resolutional Analysis
WN	WeekNum
WSN	Wireless Sensor Network
WSRF	Web Services Resource Framework

Nomenclature

Symbol	Meaning	Units
Acc	Total Estimation Accuracy	-
Acc_i	Estimation Accuracy for appliance i	-
f	Frequency	Hz
FP	False Positive	-
F_M	F- Measure or F-score	-
FN	False Negative	-
I	Current	A(Ampere)
I_{avg}	Average Current	A(Ampere)
I_{peak}	Peak Current	A(Ampere)
I_{rms}	Root Mean Square Current	A(Ampere)
P	Active Power	W (Watt)
pf	Power Factor	-
PR	Precision	-
p_t^i	Ground Truth Power Consumption for appliance i	
	at time t	W (Watt)
$\hat{p}_t^i$	Estimated Power Consumption for appliance i	
	at time t	W (Watt)
R_w	Wenner Resistance	Ω(Ohm)
Q	Reactive Power	VAR
RE	Recall	-
V	Voltage	V (Volt)
V_{peak}	Peak Voltage	V (Volt)
V_{rms}	Root Mean Square Voltage	V (Volt)
S	Apparent Power	VA
TP	True Positive	-
ρ_w	Resistivity	Ωm

U_T	Turn-on Transient Energy	W (Watt)
Φ	Phase Angle	$^\circ$ or $rads$ (radians)

Chapter 1
Introduction

1.1 Introduction

The concept of *Internet of Things* (IoT) and its potentials have been the focus of both researchers, governments and companies in recent years. IoT can be described as a network, that can interconnect and create a unified platform for physical devices, vehicles, household appliances and items that are equipped with electronics, sensors, etc [7–10]. This interconnection will allow the exchange and eventually smart utilisation of data from all involved parties such as consumers, industry, utilities etc.

Smart meters in households can provide information about power consumption, that can potentially allow better billing plans for customers, information about the condition of the appliances, and even maintenance notifications for utilities. *Smart appliances* equipped with sensors could interconnect to the IoT through smart meters, and establish a fully automated household appliance network. Smart meters and in extension *Smart Grid* represent one of the most promising applications of IoT [11].

Furthermore, monitoring of large infrastructures, such as bridges, canals, embankments, can provide information about the condition of the asset and early notifications for potential failures, that can be costly and dangerous [12]. Similarly, data from monitoring different environmental parameters, such as temperature, humidity, and even livestock health monitoring can assist to the advancement of all sectors of agricultural industry. This information can lead to better management and automation of the farming processes, minimise risk and therefore ensure increased production [13–15].

Eye-tracking devices and sensors have numerous applications in healthcare, automobile, aviation, psychology, computer science, advertisement and many more (see in [16–19]). The use of eye-tracking devices in vehicles, together with car cameras, could provide information about the driver's attention during driving and even recognise

fatigue (see in [18, 19]) and potential use of alcohol [20]. Furthermore, if used in laptops, computers and even smart phones, software and advertising companies could evaluate customers attention through visual content interaction [16]. Although these applications and their existence in the IoT could raise more questions about the ethical use of these data and would require lawmakers to propose laws that would ensure the safety and privacy of all members involved, they could still provide solutions to existing problems.

All the above applications of IoT represent only a small amount of sectors that will be part of this cloud-like network. Each entity in the IoT has its own features and measurements that can be monitored, thus the variability of the data leads to different approaches for data analysis in order to acquire useful information. For this reason, it is highly important to investigate both traditional signal processing methods, machine learning algorithms, and in general data mining processes that could be suitable for processing these smart data, finding patterns and eventually provide solutions for problems and concerns linked to each discipline.

1.2 Data Science in IoT

Data Science, according to [21], can be explained as a "concept to unify statistics, data analysis and their related methods, but also comprises its results" and includes three phases: "design for data", data collection and data analysis.

"Design for data" means the understanding of the problem and investigation behind which experiments or surveys are required. Data collection includes the databases, devices and sensors required for the measurements defined through the "design for data" phase, and the understanding of the properties of the collected data. Data analysis incorporates machine learning, statistical and signal processing, that can assist in the classification and identification of patterns using various data. This phase allows the acquisition of expert knowledge through the efficient interpretation of data [21].

All phases can be clearly implemented in the IoT concept, and especially the derivation of appropriate data analytics and signal processing tools for the extraction of useful information, which is mainly the focus of the research work presented in this thesis.

1.3 Non-Intrusive Load Monitoring

One interesting application of IoT is the field of Smart Grid and smart metering. The roll-out of smart meters in households has been on the agenda of many governments worldwide. The deployment of smart meters will provide a variety of energy data that can potentially be used in order to benefit both users and utilities, for itemised billing, competitive pricing plans and maintenance information. It is really important for the successful deployment of the smart metering technology to minimise both the cost and the disturbance of the household [22].

Disaggregating individual appliance usage from the total, aggregated energy consumption captured at the smart meter, is referred to as *Non-Intrusive Load Monitoring* (NILM) and was first proposed by Hart in [23]. Load disaggregation [24] is beneficial to customers to determine which appliances are the most energy consuming ones, which are faulty, and when it is time to replace or service an old appliance, characterised by a changing energy signature. Furthermore, it will provide a better forecasting of demand for suppliers, more efficient monitoring of smart grid from the network operator's side, and useful information for both appliance manufacturers and policy makers that can be applied to the manufacturing of smarter and more efficient appliances and the proposal and implementation of the appropriate policies. Since [23], many NILM algorithms have been proposed, using different measurements and sampling rates depending of the available sensor technology, but more frequently using sampling rates in the order of kHz [22, 25–29].

Both utilities and providers require to maintain a low cost for the smart meter deployment and in addition low computational and storage cost [22]. For this purpose, low sampling smart meters are both more affordable and the volume of data produced and requiring storage are less compared to higher sampling rates. Even with the advancements of NILM algorithms and methods, there are still algorithms and methods that could be investigated in order to increase the efficiency of the disaggregation for low sampling rates [30, 31].

1.4 Research Motivation and Aim

The motivation behind this research was driven by the fascination of the concept of IoT and in general the challenges that would arise from the volume and variability of data produced. The attempt to answer questions about how diverse data can be monitored, can be analysed and what useful information can be derived from these data, triggered the involvement in terms of research in three different and challenging fields.

While technology is ready for smart meter deployment, in order to maximise the gain from smart metering, NILM methods can provide disaggregation of energy usage for the purposes of itemised billing and improved or competitive pricing plans. Due to the cost of sensing technology, computational and storage cost, low sampling smart meters are the most important candidates for deployment [22], which motivated this research to investigate efficient NILM methods, suitable for sampling rates under $1Hz$.

Furthermore, an embankment and similar types of earthworks require extensive monitoring and maintenance, which is normally done using expensive equipment and complex geological surveys performed by experts. A failure of such infrastructure can be catastrophic, as many of such assets are located near roads, railways and other public infrastructures. For this purpose, affordable and accurate monitoring solutions are of great importance, in addition to a better understanding of which features should be monitored, and how this information can be further processed.

Eye movements and changes in pupil dilation are known to provide information about viewer's attention and interaction with visual content. Through IoT, this information could be readily available and utilised in various applications and therefore provide further advancements in multiple disciplines, such as healthcare, psychology, automotive, software and in general computer interaction. These exciting possibilities were the foundation of the involvement in the research field of human interaction with visual content.

Therefore, the main aims of this research can be presented in the form of the following research questions:

R.Q.1 What are the advancements and limitations of the available NILM methods, and how they can address successfully the IoT smart metering problem?

R.Q.2 How can DTW be applied in the context of NILM and load disaggregation and how it can address the problem of low sampling rates available in commercial smart meters; what are the advantages and limitations, compared to other state-of-the-art NILM methods?

R.Q.3 Which approaches would be suitable foroptimising for complexity and performance accuracy with a view to address practical deployments in the field?

R.Q.4 What are the challenges of earthworks monitoring and how they can be addressed by the use of low-cost IoT-enabled devices; which are the most important features worth monitoring and what useful information can be inferred for the condition of an earthworks asset?

R.Q.5 Which features should be monitored for assessing user's attention when interacting with visual contents, and which signal processing techniques would be suitable for identifying patterns and obtaining meaningful information?

Chapter 2 elaborates in answering question **R.Q.1**, as it introduces the concept of NILM, an extensive review of its background, a plethora of proposed NILM methods for performing load disaggregation/appliance classification, their limitations and comparison with the proposed methods of this thesis. **R.Q.2** is answered in Chapter 3, where a NILM approach performing load disaggregation solely using DTW has been presented. This method is unsupervised, as it does not require any previous knowledge of the appliances and their specifications, in order to disaggregate the load. A library of appliance signatures is created using unlabelled historical aggregate load data in order to classify the obtained disaggregation results. The performance of the proposed DTW-based approach is evaluated in Chapter 4, using a variety of state-of-the-art NILM methods for benchmarking purposes, that will be defined further on this thesis.

R.Q.3 is answered in Chapter 5, where two methods are proposed, that use a combination of k-means and the DTW-based approach, proposed in Chapter 4, in order to reduce the cost limitations of the original proposed method. The first method, defined as *DTW+kM* is using DTW for obtaining the library of appliance signatures, by optimising the original event detection, and in order to initialise the clusters, that will be used by k-means, which is used for classification. The second method, *kDTW*,

is performing event detection in the same way as the *DTW+kM* method, performs initial classification using k-means, and finally uses DTW as a refinement method in order to increase the performance of the method in terms of accuracy. These methods are also unsupervised and they only require an initialisation k-means clusters, using information regarding the number of appliances, and average power consumption, which are obtained using the library of appliance signatures. Similarly to Chapter 3, the library of appliance signatures for the two proposed NILM approaches is created using unlabelled historical aggregate load data.

Chapter 6 will provide the answers for both questions **R.Q.4** and **R.Q.5** by presenting two case studies from different IoT application domains. The first case study investigates earthworks monitoring, using a prototype monitoring system and performing initial assessment of the obtained results, and the second one focuses on evaluating the users' attention, when interacting with visual content, using an eye tracker device.

1.5 Novel Contributions

This research is based in concept of IoT and provides an understanding of the variety of information that would be present in IoT and methods of processing and utilising this information.

In summary, the main contributions of this thesis are the following:

- A novel unsupervised NILM method using *Dynamic Time Warping* for low sampling rate disaggregation, described in Chapter 4.
- Proposal of two unsupervised methods that combine *k-Means* and DTW, for reducing the computation complexity and for improving the performance accuracy of the original method, described in 5. The DTW+kM method uses DTW for creating the library of appliance signatures, in order to initialise the k-means clustering used for classification, and the kDTW method incorporates a DTW post-processing refinement for improving the disaggregation process.

Furthermore, this thesis presents supplementary contributions and preliminary findings regarding two other IoT applications. These contributions can be summarised, as follows:

- Prototype a low-cost platform with a variety of sensors located on one sensor node for gathering real-time data for resistivity, ground movement and pressure, that can enable prediction of earthworks failure, in the specific application a canal embankment. This contribution is described in Chapter 6.

- Evaluation of data science techniques for analysing pupil dilation and gaze fixation data for assessing human attention on visual content interaction. This contribution is described in Chapter 6.

- An implementation of a wavelet-based signal processing method and an evaluation of its suitability for extracting useful information on visual content interaction, described in Chapter 6.

1.6 Structure of the Thesis

This thesis consists of seven chapters and three appendices and is organised as follows:

Chapter 2 provides a detailed and up-to-date background for *Non-Intrusive Load Monitoring*. Chapter 4 introduces a novel *Dynamic Time Warping* NILM approach for power disaggregation using low sampling rates. This method is implemented for disaggregation purposes on three US houses from REDD dataset [1] and two UK houses from REFIT dataset [2, 3] and is evaluated using a variety of state-of-the-art NILM algorithms, namely DT (*Decision Tree*) [32], supervised GSP (*Graphical Signal Processing*), [33], unsupervised GSP [4], FHMM (*Factorial Hidden Markov Model*) available in NILMTK toolbox [34] and HMM (*Hidden Markov Model*)[35].

Chapter 5 proposes two unsupervised NILM methods, that combine *k-means* and the DTW NILM method, in order to reduce the complexity of the proposed method in Chapter 4. The performance evaluation of the proposed methods in 5 is performed using the same houses, as in the case of Chapter 4, with benchmarks the DTW-based proposed in the same chapter.

Chapter 6 presents two case studies based on different IoT applications, and data processing techniques that can be used for inferring useful information. The first case study aims to propose a prototype monitoring solution for embankments and similar earthworks and initial analysis of the monitored features. The second case

study investigates signal processing methods for attention assessment in virtual content interaction. For both case studies the potential use of DTW algorithm is discussed.

Chapter 7 consolidates the main findings of this thesis, discusses any limitations of the presented work and consequently proposes research approaches for future work.

Appendix A provides the categorisation of the different appliance types used in residential households. Appendix B includes detailed results with benchmarks all the NILM methods used in Chapter 4 for the methods presented in Chapter 5 in the form of tables and graphs. Finally, Appendix C provides a detailed background for IoT, as a supplementary material for easing the understanding of IoT for any interested reader.

Chapter 2

Non Intrusive Load Monitoring Background

2.1 Introduction

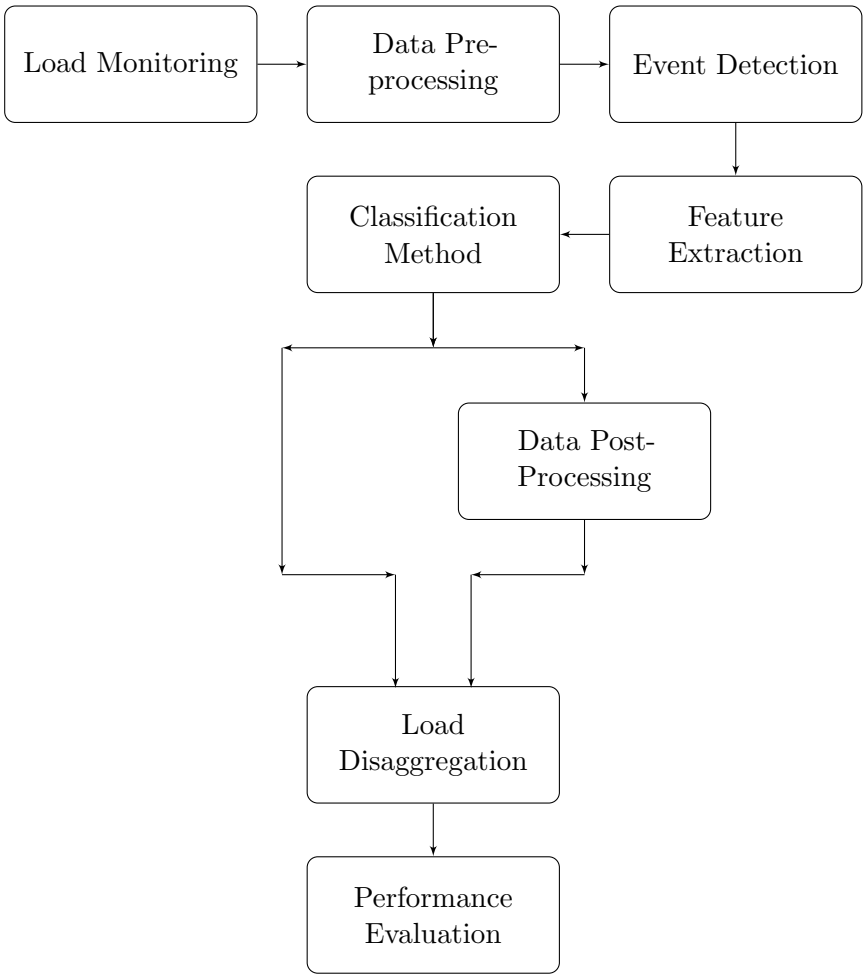
Disaggregation of individual appliance usage from the total, aggregated energy consumption captured at the energy monitor, is referred to as *Non-Intrusive Load Monitoring* (NILM) and was first proposed by Hart in [23]. Load disaggregation [24] is beneficial to customers to determine which appliances are the most energy consuming ones, and which are faulty. It could also be beneficial to system operators to monitor the effect of smart grid fluctuations on the residential communities, and to suppliers to better forecast demand, although and thus it is an important aspect of smart homes, smart buildings and smart grids. Although the use of personal energy data from suppliers could raise privacy concerns, policy makers could develop appropriate policies in order to ensure privacy and anonymity of any residential data used. Since [23], many NILM algorithms have been proposed to adapt to advances in smart metering technology capturing energy measurements at a range of sampling rates.

The block diagram in Figure 2.1 shows a graphical representation of the different steps that researchers need to take into account when developing a NILM method. This chapter has been structured as per Figure 2.1 to explain each of the building blocks involved in NILM. Note that not all the building blocks presented in Figure 2.1 are necessarily present in all NILM implementations, due to differing quality of raw input data, algorithm used, appliances being disaggregated, accuracy and complexity constraints.

The NILM process typically consists of data acquisition and data pre-processing, event detection and feature extraction followed by a classification method, that allows

the disaggregation. After data acquisition (Section 2.2), data pre-processing (Section 2.3) can be done in the form of interpolation and data filling for replacing missing measurements and can be extended to power normalisation of the signal, filtering (for signal smoothing and getting rid of sudden peaks), and thresholding to remove small power loads that would appear as noise as well as so-called base-load, from appliances that are always running.

Figure 2.1: Explanation of the various steps of a NILM Process.



Next, event detection (Section 2.4) is done to identify events of appliances switching on and off or appliance state changes. Event detection is followed by feature extraction (Section 2.5), where various features are extracted from the identified event windows. Classification (Section 2.6) is then used to group sets of extracted windows which have similar characteristics, such as power levels, time profile, reactive components etc. Some approaches, have proposed an additional step that performs data post-processing

(Section 2.7), commonly know as *active learning* or *model tuning*, which allows the refinement and optimisation of the classification methods, as seen in [5]. The final output of each NILM method is the disaggregated load, as seen in Figure 2.1. This output can be evaluated in terms of performance, using various metrics that will be discussed in Section 2.8, in order to identify the effectiveness of a NILM method. This step is essential for evaluation of a NILM algorithm at research stage or validation of a NILM algorithm during real world field trials.

2.2 Load Monitoring and Data Acquisition

According to Figure 2.1, the first step of a NILM method is load monitoring, which essentially means monitoring of the power consumption using a smart meter or relevant sensors. Makonin in [36] has listed the different aspects that NILM researchers need to take into account during this step, and these can be summarised in the measurement types, sensing types and sampling rate.

The measurement types refer to the load parameters that can be measured and can affect directly which features can be extracted from the data. The appliance features can be found in Section 2.5, where a range of different measurement types and features, that have been used in previous NILM works, are briefly presented, such as active power, reactive power, voltage, current, Electromagnetic Interference (EMI) spectrum analysis.

By sensing types, Makonin [36] refers to the *monitoring techniques* with regards to the number of sensors and points required during load monitoring, which can be categorised in in *single-point sensing* and *multi-point sensing* techniques [36, 37]. *Single-point sensing* uses only one point for monitoring power consumption, while *multi-point sensing* use multiple locations by using sensors either at each appliance, known as *Individual Appliance Monitors* (IAMs), or by using circuit breakers. More details regarding the above monitoring techniques, and any advantages and disadvantages can be found in the works of [36, 37].

Sampling rate is the rate in which the load data are measured. It is highly important in NILM, as it can affect the type of features that can be extracted and the cost of the equipment used for monitoring and data storage purposes. Sampling rates in the field of

NILM can be found in Section 2.2.1. Finally, the result of every load monitoring NILM process is the data acquisition of a variety of measurement types. A plethora of NILM datasets, such as AMPds [38], REDD [1] and REFIT [2, 3] with different measurement types and sampling rates are currently publicly available for NILM researchers and can be used for testing NILM algorithms and also as a ground truth comparison (see [3, 39, 40] for more details).

2.2.1 Sampling Rates

The data acquired through different smart meters, sensors, etc. vary in terms of sampling rate, and thus the NILM algorithms analysing them can be classified into two categories: *high sampling rate* and *low sampling rate*-based.

- *High sampling rate*-based NILM methods are algorithmic methods that have used sensor technology with sampling rates $> 60Hz$. Generally, higher sampling rates can provide higher accuracy, but both the cost of the sensing technology and the computational and storage cost is higher [22]. For example, Patel *et al.* [25] have used $100kHz$, Laughman *et al.* [26] $8kHz$, Berges *et al.* [27] $10kHz$ and Chang *et al.* [28, 29] $15kHz$.

- *Low sampling rate*-based NILM methods are the methods that use sampling rates $\leq 60Hz$, which are currently the focus of NILM research, as these rates can be achieved using cost-effective equipment, that is readily available [30]. For example Hart in [23] has used both $1Hz$ and a higher frequency at $7680Hz$ for capturing current and voltage waveforms and Ruzzelli *et al.* [41] have used 1 sample per min.

For example, low-rate NILM approaches can use mainly *steady-state* parameters, such as active or real power [35], reactive power using $1Hz$ [23], power factor [41], and additionally voltage or current waveforms [26, 42], but without receiving higher harmonics. At high sampling rates, it is possible to detect *transient voltage noise* [25] and *transient* parameters [30, 37] including higher harmonics current and voltage signatures [26, 42, 43]. *Feature extraction* parameters will be further described in Section 2.5.

2.3 Data Pre-processing

NILM datasets and generally smart metering data, similarly with other data-driven research fields, can suffer from measurement noise, corruption and missing data. Therefore, many researchers have focused on developing effective *data pre-processing* techniques, such as interpolation, forward filling, and filtering, that could resolve this issue, as seen in [5]. Generally, these techniques originate from the field of statistics and signal processing, and can be adapted for the case of load data.

Data pre-processing is an important step for developing a NILM method during research, but it can lead to higher accuracies due to the cleaning process, thus it is important to clearly define any pre-processing process, as Makonin and Popowich suggested in [44]. Furthermore, a NILM algorithm should be robust and able to perform online disaggregation using the real-time measurements, which are susceptible to measurement noise and corruption.

2.3.1 Data Imputation Methods

Imputation refers to the process of filling missing data, and could be applied for the case of corrupted data, when identified and removed from a dataset [45]. According to Peppanen *et al.* [45], multiple variations of *interpolation* can present a simple and suitable solution for performing imputation in smart metering.

The simplest version of interpolation is the *nearest neighbour approach*, where missing data are replaced with the value of the closest available data, either using forward or backward filling, or an average of the closest values [45]. This method is suitable for short intervals of missing data and has been used for pre-processing REFIT dataset in [3], in NILMTK [34] and UK-DALE [46].

Furthermore, *Linear Interpolation* (LI), is suitable for slightly longer intervals of missing data, as it performs data imputation using two samples. LI can offer both fast and low complexity processing [45], but the estimation accuracy is reduced as the intervals are increased. Other variations of interpolation include polynomial and splined interpolation, but they are generally computationally expensive.

Peppanen *et al.* [45] suggested, for extended periods of missing data, the use of samples of historical data, from previous hour, day or month. A *historical average* (HA)

imputation approach was proposed, which can perform better than LI imputation, but depends on the characteristics of the data and frequent patterns in the historic data [45]. This idea has been extended to a combined LI and HA imputation approach, named *optimal weighted average* (OWA) imputation, which estimates missing data samples, as the weighted average of the LI and the HA imputed values [45].

2.3.2 De-noising using Signal Processing Techniques

Smart metering data and data from IAMs can be expressed as time-series signals, as they obtain continuous measurements of electrical quantities (e.g. power consumption, current, voltage), depending on the sampling rate available in each device. They can be susceptible to noise due to the monitoring devices, in addition to unknown load noise, transient spikes and fluctuations [5]. In order to address this problem, signal processing techniques, such as *total variation reguralisation*, filtering and smoothing can be applied in the context of NILM.

Kolter and Jaakkola *et al.* [47] have proposed the use of *1-D Total Variation Regularisation*, a technique commonly used for image processing, in order to remove any outliers and reduce the effect of rarely used appliances. Furthermore, *median filtering* is commonly used in NILM research for noise and spike removal and signal smoothing [48–55]. While it is efficient at removing noise and outliers by usually preserving the edges, it has shown difficulty in addressing periodic curves (e.g. sine, triangle, etc.) [55]. In [51], a two stage median filtering process was applied for preparing the data, which first reduces the noise and then smoothens the signal. Pattem *et al.* [52] have proposed a pre-processing method, which employs quantisation, calculation of the quantisation error waveform, median filtering and down-sample/window-wise constant smoothing.

According to Weiss *et al.* [55], a mean filter has low computational cost, but it has a difficulty in removing large outliers and could potentially erase edges that correspond to an on/off event of an appliance. A kernel-weighted average filter (Nadaraya-Watson filter with Gaussian kernel) was proposed in [55], which allows attenuation of the signal fluctuation, while preserving the edges [55], but it can increase the computational complexity. For this purpose, the authors selected the use of a combined median and mean filter, which performs close to the optimal solution compared to the kernel-weighted average filter [55].

An adaptive *Cumulative Sum* (CUSUM) filter was proposed in [56] for noise removal and data smoothing. Furthermore, in [57], the authors have introduced the use of *denoising auto-encoders* (dAE) for dimensionality reduction.

In [58], a pre-processing step was proposed for heuristically estimating and removing the base load before disaggregation using a *Graphical Signal Processing* (GSP)-based disaggregation. Zhao *et al.* [5] have designed two types of GSP filters for, based on total variation regularisation and *bilateral filtering* (BF), which follows an initial median filter. BF can provide a smoother signal, but can split the sharp edges into multiple segments, thus filtering out edges that could potentially correspond to actual appliance events [5].

2.4 Event Detection

After data pre-processing, NILM methods require to develop methods for detecting the operation of an appliance (event), commonly known as *event detection*. Event detection methods can be categorised in: *edge detection* and *probabilistic event detection*. *Edge detection* uses the rising and falling edges of power consumption. *Probabilistic event detection* treats edges as a probability distribution.

Event detection can be affected by the different types of consumer appliances present in a household. These appliances can be classified in four categories: ON/OFF appliances (eg. toaster), *Finite State Machines* (FSM) (eg. washing machine), continuously variable consumer devices (eg. dimmable lights) and permanent consumer devices (eg. smoke alarms). For more details the reader can refer to Appendix Section A.

2.4.1 Edge Detection

A household's power consumption constantly changes, depending on the operation of different appliances. Each appliance operation can create rising edges, when the appliance is turned on, and falling edges when the appliance is turned off, or during the state change of an appliance, which correspond to the transient part of the signal. When these edges are high enough the operation of the specific appliance can be identified as an event. This method of event detection, known as *edge detection* or *expert heuristics* [59, 60], is commonly used in NILM algorithms.

Many researchers have used predefined thresholds in order to detect the relevant events, such as Norford and Leeb in [61] ($3kW$ and $5kW$ for commercial buildings), Baranski and Voss [62] ($80W$), and Tsai and Lin [63] ($\Delta I_{itensity} \geq 0.03A$).

In [32], which is a published work based partly on the research work of this thesis, the rising and falling edges of active power have been used, using an adaptive threshold W, which is based to the minimum state transition that needs to be detected and the maximum variation of the active power within one appliance. This edge detection method will be explained in Chapter 3.

Three-dimensional clustering, based on real power, reactive power and harmonics, has been used in [26] for identifying the edges corresponding to appliance operations, while Leeb *et al.* in [64] have used a complex multi-scalar edge detector. Some researchers have further categorised the last techniques as *matched filters event detection*, as they use a known or template signal and correlate it with an unknown signal, in order to identify the existence of the template signal in the unknown signal [59, 60].

2.4.2 Probabilistic Event Detection

Probabilistic event detection, as in [36, 59, 60], is an event detection method that a probability distribution is used for the edges, instead of normal edge detection, as discussed in Section 2.4.1. *Generalised likelihood ratio* (GLR) has been used for event detection in [27, 59, 65, 66], while Jin *et al.* [67] have used a robust adaptive *goodness-of-fit* (GOF) χ^2 test event detector. Furthermore, Nguyen *et al.* [68] have proposed an online event detection using the CUSUM adaptive filter, while Pereira in [60, 69] have proposed the use of a *Log Likelihood Ratio Detector with Maxima* (LLD-Max) event detection algorithm. Details regarding the specific probabilistic event detection methods can be found in the relevant bibliography, as this research was focused on the use of a simple edge detection method, as described in section 2.4.1.

2.5 Feature Extraction

Following event detection, NILM methodology requires a *feature extraction* step before a classification method is performed. During this step, different appliance features are build using the measured data, commonly referred in NILM bibliography as *appliance signatures*. The extracted features can be classified as: *steady-state features*, *transient features* and *non-traditional features*, as seen in [31, 70]. This classification can be seen graphically in Figure 2.2.

2.5.1 Steady-State Features

Steady-State features refer to the features extracted from signals that either do not change or change insignificantly over time. In terms of signal processing, steady-state signals can be expressed as a finite number of sinusoids or a fixed sum of sinusoids [71].

- **Power Change**: Real or active power (P) and reactive power (Q), are commonly used features in NILM research [35, 61, 72–77]. Real power is power consumed during the ON state of an appliance, while reactive power is the power that is stored in components and then released again back to the source through each cycle, caused by the inductive and capacitive components of the appliances. Clustering in $P-Q$ space has been used in [24, 78], while *power factor* (pf) has been incorporated in [79]. pf is the ratio between real and *apparent power* (S) ($|S| = \sqrt{P^2 + Q^2}$) and it usually varies between 0 and 1, depending if the load is more reactive or more resistive [79].

- **V-I Features**: Raw waveforms have been used for appliance identification in [80], while a variety of V-I features, such as *peak current* (I_{peak}), *root mean square* (RMS) current (I_{rms}), *peak voltage* (V_{peak}) and RMS voltage (V_{rms}), have been included in [41, 42, 81, 82]. Fourier series analysis has been used in order to determine input current harmonics in [26, 42, 83–86].

- **V-I Trajectory**: A 2-dimensional (2-D) *Voltage-Current (V-I) trajectory* can be used for appliance classification, and wave shape features can be extracted, such as *asymmetry*, *looping direction*, *area*, *curvature of mean line*, etc. [43, 87, 88].

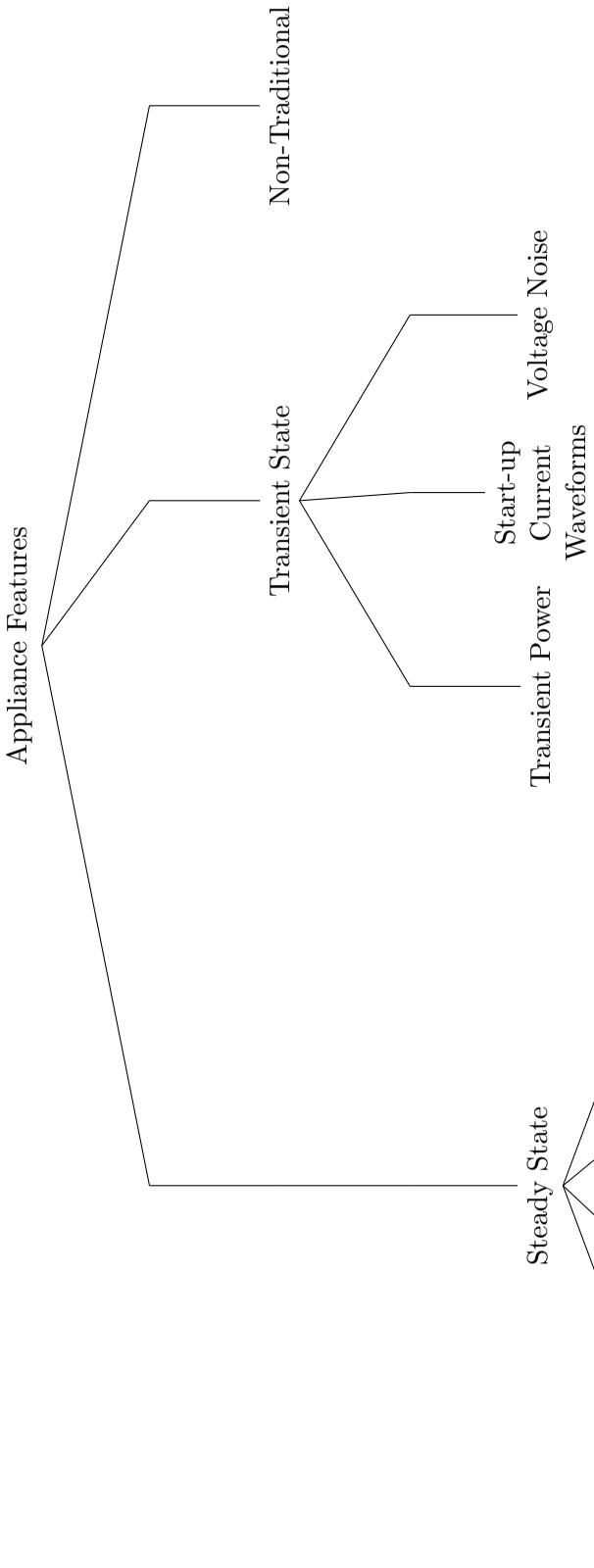


Figure 2.2: Taxonomy of Appliance Features as in [31] and [70].

- **Voltage Noise**: Steady-state electrical noise (also known as *Electromagnetic Interference* (EMI)) present on the power lines, can be used for identification of the switching of appliances and also for devices that do not generate transients, such as appliances equipped with *Switch Mode Power Supply* (SMPS) [25, 89].

2.5.2 Transient Features

A signal, according to [71], has a *transient state* when its Fourier expansion requires an infinite number of sinusoids. Transient features are difficult to detect and require higher sampling frequency, which increases the implementation cost. Appliances with identical or similar power consumption can potentially be distinguished by using transient features, such as start-up current [24], which can increase the performance of NILM algorithms.

- **Transient Power**: Spectral envelopes have been used for transient event classification in [64, 90]. Furthermore, Laughman *et al.* [26] have investigated the use of both active and reactive power in steady and transient state together with higher harmonics. Harmonic content features have been included for appliance classification in [91], harmonic energy of the current harmonics in [92], and the transient energy and response time in [29, 84].

- **Start-up Current Transients**: Start-up transients have been used in [61], while power spikes that occur during transition state of an appliance have been considered in [93]. Furthermore, energising and de-energising transient features for each appliance, derived by the current waveform, have been investigated in [63].

- **Transient Voltage Noise**: Transient voltage noise is the noise that occurs during the transient events, such switching on/off, changing state in multi-state appliances and has been used as a feature in [25].

2.5.3 Non-Traditional Features

In addition to the traditional steady-state and transient features, many researchers have considered some features that cannot be included in any of the above categories, which can be referred as *non-traditional features* [31, 70].

Wang and Zheng in [94] have suggested the representation of fast switching appliance events as Triangles and steady working events, as Rectangles (for more details see [94]). Liang *et al.* [42], have included additional features such as *instantaneous admittance waveform (IAW)*, and *eigenvalues* (EIG).

Berges *et al.* [95] have investigated the use of data from environmental sensors (such as light intensity, sound level, etc.), in order to increase disaggregation accuracy, while *electromagnetic field*-sensors (EMF) have been used in [96] for appliance usage detection by monitoring changes in both magnetic and electric field. The control/status signal provided by the *Building Management System* (BMS) have been included in [97], while audio signals features, such as *Mel-frequency cepstral coefficients* (MFCCs), have been used for appliance detection in *Tiny Energy Accounting and Reporting System* (TinyEARS) [98].

Behavioural features, such as on/off duration, date and time, dependency between appliances (e.g microwave, kettle and toaster use during lunch time), and daily schedule of the occupants, have been considered in [47, 75, 77]. In [22], weather, location and demographics have been considered as additional features.

2.6 NILM Learning and Classification Methods

After feature extraction, NILM learning and classification methods are required in order to group sets of extracted features which similar characteristics.

Figure 2.3 provides a graphical representation of the categorisation of the NILM classification methods, and can allow a better understanding of he structure of this section. Based on the employed classification method, all NILM algorithms can be classified as supervised and unsupervised.

Supervised NILM methods use labeled appliance events to train classifiers, which means that they are highly dependent on user intervention and *"a-priori"* knowledge of all system parameters, such as individual appliance data [31, 70, 99].

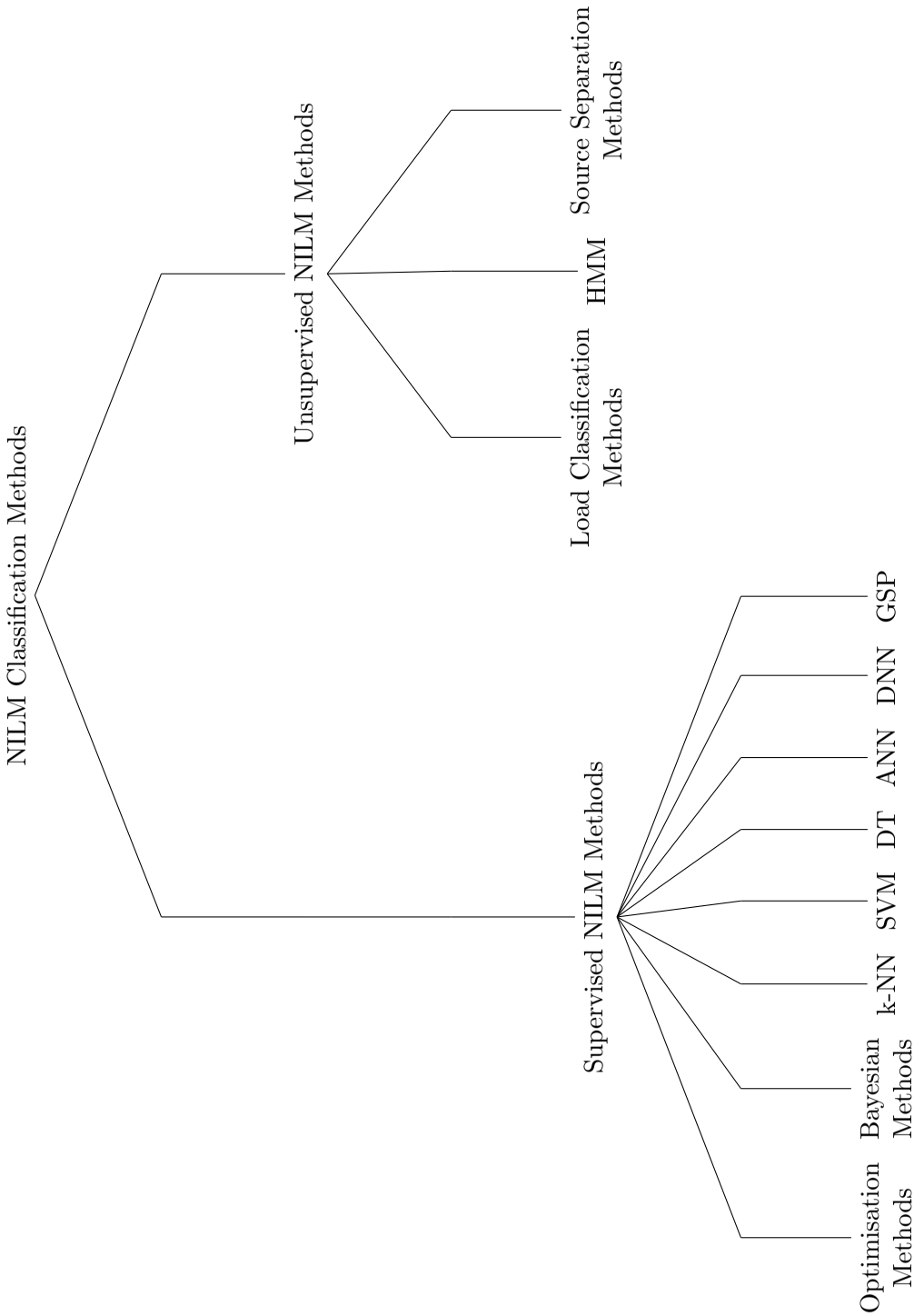


Figure 2.3: Taxonomy of NILM Classification Methods.

Unsupervised methods, on the other hand, do not require labeled sets or individual appliance data, minimising user intervention [31, 70, 99]. According to [99] and [70], unsupervised NILM methods should be able to train and disaggregate *online*, and continuously adapt to any changes that occur on the power system, such as addition, removal or substitution of an appliance.

2.6.1 Supervised Methods

Supervised NILM methods can be distinguished in *optimisation methods* and *pattern recognition methods* [31, 70, 99]. As the pattern recognition methods represent the majority of the NILM methods available in NILM literature, it has been decided to categorise them according to the algorithms used for classification, as found in Figure 2.3.

Hart [24] has performed simple clustering in the two-dimensional active-reactive power space, that is, for each window, active and reactive power readings are collected and input to a cluster analysis algorithm; clustering is followed by building an appliance model as a finite-state machine, tracking behaviour using the Generalised Viterbi algorithm and finally identifying appliances.

Following Hart's work in [24], researchers have experimented with a range of pattern recognition methods for load disaggregation, which can be summarised in the following: *Bayesian methods*, *k-Nearest Neighbour* (k-NN), *Support Vector Machines* (SVM), *Decision Tree* (DT), *Artificial Neural Networks* (ANN), *Deep Neural Networks* (DNN) and *Graphical Signal Processing* (GSP) methods.

2.6.1.1 Optimisation Methods

Optimisation methods are the NILM methods that approach load disaggregation as an optimisation problem. According to [31], an optimisation method compares the extracted features of an unknown load with known loads available in an appliance signature library, and tries to identify the closest match by minimising the error between the known and unknown extracted features. Mathematically, this can be represented by the following equation:

$$class = \underset{i}{\operatorname{argmin}} \|\hat{y}_i - y_i\| \tag{2.1}$$

, where $\hat{y}_i$ is the appliance feature available in the signature library and the y_i is the new feature extracted during the new unknown event [31].

Liang *et al.* [42] have incorporated both optimisation and pattern recognition algorithms for load disaggregation (namely *least residue (LR)* algorithm and ANN), using a variety of features, such as current waveform, active/reactive power, harmonics, IAW, EIG, etc. The authors have proposed the use *Committee Decision Mechanisms* (CDM) in order to identify the best possible answer in the disaggregation problem utilising both single features and single algorithm disaggregation. The CDMs used in their work were *most common occurrence* (MCO), *least unified residue* (LUS) and *maximum-likelihood estimation* (MLE). The authors have used both high and low sampling rate data (12kHz and 1Hz respectively), while the proposed methods of this thesis focused on disaggregation using only low sampling rate. Furthermore, the methods in this thesis have only used active power as a feature, while the methods in [42] provide both single and multiple features disaggregation and have included a combined PQ feature, rather than only active power.

Suzuki *et al.* [80] have proposed a NILM method that uses *integer programming* on current waveforms, while this thesis offers NILM solutions using only active power. As some appliances have similar waveforms, the authors have used a unification method to overcome this problem, which depends on *"heuristic maneuver"*, but did not investigate appliances, whose current waveform change continuously. According to [80], the proposed method does not require *relearning* when a new appliance is installed in a household, as the signature of each appliance can be obtained by using only one period of current waveform per appliance. Similarly, the proposed methods in this thesis do not require relearning of the whole household when a new appliance is added. The signature of the new appliance can be obtained through the aggregate data and then compared to the existing signatures in the library and if the signature does not present any similarity to the library signatures, the library will be updated.

Furthermore, Baranski and Voss in [100] have used an *Iterative Self-Organizing Data Analysis Technique* (ISODATA) algorithm, in order to cluster the different appliances

that have been monitored using optical sensors on Ferrari meters. The method of [62] does not require training and instead builds a histogram of historical data and uses a sub-optimal genetic algorithm to match a large set of on-off events to appliance presences in time and refined further with the Viterbi algorithm [24]. The algorithm shows good efficiency for large loads, but suffers from high computational complexity and requires the use of fuzzy sets in order to overcome overlapping events. Event detection was performed using low pass filtering and rule-based pre-conditioning of the data, whereas the proposed approaches of this thesis use an simple edge detection approach, as described in 2.4.1.

According to [31], the optimisation problem is more complex in the case of aggregated data, as the algorithm needs to perform both individual matching and appliance combination matching, in order to identify appliances that work simultaneously. In general, through the work of the researches in [42, 62, 80, 100], it has been apparent that these methods have high complexity and, especially when unknown loads occur in the aggregated load data. Furthermore, as already mentioned above, it is difficult to distinguish individual loads with similar features when the occurrence of different appliances overlap.

More recently, researchers, such as in [101], [48] and [102], have tried to combine *Particle Filtering* (PF) with *Hidden Markov-Models* (HMM). Though these techniques are unsupervised, it has been decided to include them here as an example of the progress of techniques that use optimisation methods. In [101], an appliance estimation approach of on/off appliances modelled by HMM, was proposed using PF. Their experimental results showed that a higher number of particles is required.

In [48], the authors have proposed *Particle Filter-Based Load Disaggregation* (PALDi) based on the work published in [101]. The authors noticed that higher number of particles improved the algorithm's performance, and on the other hand higher number of appliances with constant number of particles resulted to lower accuracy. For this purpose, the number of particles was selected according to the number of appliances. The authors have generated the aggregate data from known appliances using the data available for the specific appliances for House 1 in REDD dataset [1], while the proposed methods of this thesis have used the actual aggregate data available for the same test house, where all appliance operations are present (known or unknown). Furthermore,

the authors in [48] have performed smoothing and de-noising of the data, while the only pre-processing step used in this thesis was forward-filling for any missing values and downsampling of the REDD dataset from $1Hz$ to $1min$ (see Chapter 4 for more details).

Wong *et al.* [102] have proposed a *particle-based distribution truncation* (PDT) method with a *duration-dependent* (DD) state transition model. According to the authors, their results have indicated an efficient disaggregation performance with less complexity compared to Viterbi algorithm and conventional PF methods. The authors have used for testing purposes the last third of the data available from House 2 in REDD dataset [1], presumably using the original sampling rate available, namely $1Hz$, as there is no mention about any downsampling. The PDT method was able to disaggregate lighting, while the proposed methods in this thesis were not successful in disaggregating it. The method proposed in [102] was not able to successfully disaggregate washing machine, which was also the case for the methods used in this thesis. In terms of performance evaluation, even thought the authors in [102] have used similar metrics, namely precision and recall, there cannot be a direct comparison between the PDT method and the DTW-based methods of this thesis, due to the different testing period and sampling rate.

Kong *et al.* [103] have proposed a hybrid NILM method based on *segmented integer quadratic programming* (SIQP) and *constraint programming* (CP), in order to increase computation efficiency and reduce the search space. The authors were able to significantly increase computation efficiency compared to PF, and still obtained comparable accuracy compared to PF-based techniques.

The authors in [103] have tested their method using House 1 from REDD dataset [103], and were able to disaggregate bathroom GFI, electric heat, lighting and oven, while the proposed methods of this thesis were unable to do so, which will be discussed further in Chapters 4 and 5. It is important here to note that the method in [103] performed the testing using three scenarios, with very limited parts of the dataset, namely $6hrs$, $24hrs$ and $48hrs$, whereas the proposed methods of this thesis have used 16 days for testing, that do not match the time periods used in [103]. Therefore, there can not be a direct comparison of the obtained accuracy for the commonly disaggregated appliances among the methods in [103] and this thesis.

2.6.1.2 Bayesian-based Methods

Marchiori *et al.* [76] have proposed a *Bayesian approach* for disaggregation using only active power and state change information and a *heuristic approach* using $P-Q$ space as features. A *Naive Bayes classifier* has been used for recognising appliance states given the aggregate active power and the detected state change. Their *Bayesian approach* was able to perform better compared to the proposed *heuristic approach*, if the appliances have stable power, but the appliances included in their experiment were limited. Additionally, it has been assumed that each appliance's state is independent from the states of other appliances. This assumption, though accurate for most appliances, would not be accurate for appliances that often have correlated operation, such as TVs and DVD players. In terms of comparison, the Bayesian approach in [76] offer a probabilistic disaggregation approach, whereas the methods proposed in this thesis approach the disaggregation using edge detection.

Sanquer *et al.* [104] have introduced a *Hierarchical Bayesian*-based method, where both feature extraction and class learning can be performed together using a database of electrical transients. The proposed hierarchical Bayesian learning method is based on the smooth transition regression and a *Markov-chain Monte Carlo* (MCMC) algorithm is used for sampling the *posterior distributions* of the features and to *"learn the class-specific parameters"*. The authors have assessed the performance of their method using a very limited database of real-world electrical transients for only two classes of appliances (vacuum cleaner and refrigerator), there for the efficiency of this method has not been tested using multiple classes of appliances. The proposed methods of this thesis were able to obtain good disaggregation performance using a variety of larger datasets (REDD [1] and REFIT [2, 3]) and a bigger variety of appliances.

In [105], a NILM method was proposed that uses a *Dynamic Bayesian Network* (DBN) to incorporate user behaviour and a *Bayes filter* to perform online inference. The input data in this method are both observation sequences and label sequences. The observation sequences are constructed using a range of features from the energy data measured at circuit-level, while the label sequences are the operating states of all appliances at any *"time slice"*, manually labelled whenever there is a state change. The authors have constructed the observation model using two different methods, one

using *Gaussian Mixture Model* (GMM) and the other using *Discretisation*, which was able to acquire the best performance. The proposed method can be classified as a *temporal method* and was compared with some well known *non-temporal* methods such as k-NN, Naive Bayes and SVM, and the obtained results shown that Bayes filter with GMM or discretisation were able to outperform these methods. Additionally, when compared to Viterbi algorithm, the proposed method obtained better or comparable results. According to the authors, taking into account user behaviour can essentially improve performance and can provide more accurate states. It is important here to state, that the proposed methods of this thesis were focused on obtaining disaggregation using only the power consumption available in the aggregate data measured at a smart meter (single-point) and did not take into account any user behaviour information, that would potentially increase the performance, according to the authors claim in [105].

2.6.1.3 k-NN Methods

Some researchers have investigated the use of *k-Nearest Neighbour* (k-NN) as classification method, as seen in [66, 79, 89, 106–108]. More specifically, in [79, 106], the authors have proposed two *Steady-States* (StS) Recognition approaches in order to obtain the appliance signatures, which use the step-changes in active and reactive powers and power factor. The approach presented in [79] uses the ratios between the rectangular areas between successive state values, while the approach in [106] uses a MinMaxSteady-State algorithm. In both works, the authors have used as classification methods both k-NN ($k = 5$) and variations of SVM, and they obtained on average higher accuracy when using the 5-NN method. Furthermore, it is apparent that the works presented in [79, 106] performed disaggregation using multiple features, while this thesis investigates disaggregation using only one feature, the active power obtained through the aggregate power consumption measured at the smart meter.

Gupta *et al.* [89] have used k-NN classifier with $k = 1$ and an *Euclidean distance* metric with inverse weighting, and claimed 100% accuracy, but have not provided more details about the obtained performance results. In terms of features, the authors in [89] have used frequency domain features using *Fast Fourier Transform* (FFT), while the features used in this thesis are obtained from the time series aggregate active power.

Berges *et al.* [66], as already mentioned in 2.4.2, have implemented a probabilistic

event detection based using GLR and have investigated the use of both active and reactive features, and Fourier regression coefficients. In order to perform classification, the authors in [66], have employed various classifiers, such as 1-NN, *Gaussian Naive Bayes* (GNB), DT and *Multi-class AdaBoost* (MultiBoost), but favoured the use of 1-NN, as this methods was able to obtain the highest accuracy when using Fourier regression coefficients. Furthermore, Berges *et al.* [27] have improved their 1-NN based method using *Euclidean distance* measurements and were able to obtain accuracy $\sim$ 85%. The authors in [27, 66] have used both active and reactive power, while this thesis obtains only active power as a feature. Furthermore, the features used in [27, 66] were obtained using higher sampling rate ($20Hz$) compared to the sampling rates used in this thesis, namely downsampled REDD [1] at $1min$ and REFIT [2, 3] $\sim 8sec$.

Tsai and Lin *et al.* [63] have proposed an Adaptive Non-Intrusive Appliance Load Monitoring (ANIALM) system using energising and de-energising transient features for each appliance, derived by the current waveform. They have used as classifiers both *k-Nearest Neighbour Rule* (k-NNR) and *Back-Propagation Artificial Neural Network* (BP-ANN), together with an *Artificial Immune Algorithm* (AIA) with the *Fisher criterion* in order to improve identification performance, by adaptively adjusting the feature parameters, when new appliance types are added in the monitoring system. They favoured the k-NNR method, due to its computation and implementation simplicity, as only parameter k requires optimisation using an exhaustive search, and does not require any training or re-training. The features used in the ANIALM system are extracted using the current waveform, which is filtered using a *low-pass filter* (LPF) at $500Hz$, while this thesis uses low resolution data, as already discussed, and obtains the appliance signatures using the active power time series signal.

More recently, Azaza and Wallin [107] have proposed a supervised NILM method based on *dynamic fuzzy c-means clustering* (dFCM) and k-NN for label matching and pattern recognition. Data filtering has been performed using four different filters, namely *median filter*, *mean filter*, *median mean* and a *variant of total variation denoising*, with the latter being selected for implementation, as according to the authors, it was able to outperform the rest of the filters. The dynamic fuzzy c-means clustering was used to create appliance signatures library using active and reactive power together with the time of day usage. This dynamic approach does not require an a priori definition of

the number of clusters, as in the case of the regular fuzzy c-means method. Both *Akaike information criterion* (AIC) and *Bayesian information criterion* (BIC) are computed iteratively, and the final number of clusters is selected when a low value is achieved for both criterions. Finally, k-NN algorithm is used for matching the events to the most similar signature, using the distance norm.

The authors were able to obtain good accuracy for high consumption appliances and cold appliances (e.g. fridge, freezer, stove, dishwasher), which was also similar for the methods of this thesis (see Chapters 4 and 5 for more details), but much lower accuracy for the case of low consumption appliances. Although both the method in [107] and the methods in this thesis were focused on low resolution disaggregation, the authors in [107] obtained active and reactive power and the time of the day usage, while this thesis presents disaggregation approaches using only active power.

In [108], a approximation *Probabilistic K-Nearest Neighbour* (PKNN) method was proposed using as features sensor data, such as temperature and humidity, together with power features (steady state) event detection data equal to the number of appliances, all sampled at $1Hz$. Generally, PKNN is a classification method which calculates the nearest neighbour by its probability, but as a traditional PKNN is computational expensive, as it calculates the probability from the total sample, the proposed approximation PKNN in [108] uses only part of the total sample. The authors have presented a comparison of traditional PKNN method (with and without sensor data) to some well known classification methods, such as SVM, Naive Bayes and J48 (a java implementation of the C4.5 decision tree algorithm). PKNN with sensor data was able to outperform the other classification methods, when the number of appliances was more than four.

Furthermore, the proposed approximation PKNN method, when compared to the full traditional combination method (with and without sensor data) and the independently factorised method with sensor data, was able to obtain accuracy almost at 100%. During the optimisation attempt for the parameter k, it has been noted that the accuracy for the proposed approximation PKNN method was on average $> 99\%$, but for all experiments the authors decided to use $k = 7$. In terms of comparison with the methods presented in this thesis, all methods have performed disaggregation using low sampling rate, but as the approximation PKNN method in [108] uses a variety of

measures, it would be difficult to compare it with the work in this thesis, where only active power has been used for disaggregation purposes.

2.6.1.4 SVM Methods

Other researchers have adopted *Support Vector Machines* (SVM) as classification approaches, as in [81, 109–120]. In [109], three different types of SVM have been implemented using different kernel, namely polynomial kernels, *Radial Basis function* (RBF) and *sigmoid kernels*. All proposed methods were able to achieve almost the same error rates for *"handwritten digit recognition"*, but different error rates for the case of state estimation. According to the authors, when the complexity of data is high, RBF is more useful that the other proposed methods. Regarding the features extracted during this work, the authors have obtained various odd-order current signal harmonics and the corresponding phase angles, whereas this thesis as already discussed have selected to obtain features using the aggregate active power in the time domain.

Lin and Tsai [110] have proposed a *Hierarchical SVM* (HSVM) for load classification using transient features of the current waveform. SVMs are nonlinear classifiers and can optimise decision hyper-plane, but they are binary classifiers, thus can only deal with a two-class problem. The proposed HSVM method was able to overcome this problem by decomposing the problem into a series of two-class sub-problems by arranging multiple binary SVMs into a *hierarchical tree*. They were able to successfully identify both single and multi-state appliances with identification rates higher than 93%. The authors in [110] have extracting energising and de-energising transient features using the current waveform at $500Hz$ sampling rated due to LPF, much higher that the $1min$ and $\sim 8sec$ resolution used for aggregate active power in this thesis.

A NILM method using *Multiple class*-SVM (M-SVM) using Gaussian kernel function has been proposed in [111]. A harmonic feature analysis of the current signal is initially performed and the obtained data are used in the proposed M-SVM for appliance identification. According to the authors, the classification complexity is not affected by the dimension of the feature space and the number of data points, thus it can provide a scalable solution for a larger set of patterns. The authors were able to obtain classification accuracy rates higher than 90%, both when using the unprocessed current data and the relevant harmonics, with the latter being able to

show the highest accuracy rates. As the features, used in [111], were derived from the current waveform, unprocessed or harmonics, it would be difficult to compare this method with the methods proposed in this thesis, which were developed in order to disaggregate using active power.

Kato *et al.* [81] have proposed the use of SVM for appliance classification and one-class SVM for detection. They have compared the proposed method with the results obtained using 100 original dimensional vectors or 5 typical features, such as average value of the electric current, peak to average ratio, etc., and were able to obtain 99.9% accuracy for 16 appliances, and 95.8% for 25 appliances, much higher compared to the typical features performance and equivalent performance to the 100 original dimensional vectors. Detection of unregistered appliances was also investigated by using 24 appliances to train a one-class SVM, and the remaining one to perform as an unregistered appliance. The *Receiver Operating Characteristic* (ROC) was used for assessing the detection accuracy, and the authors have claimed accuracy $> 97.7\%$. The method proposed in [81] offers a NILM solution using multiple features, either the dimensional vectors or the 5 typical features, whereas this thesis presents approaches based on only one feature for disaggregation, namely active power. Multiple features, in general, can provide more parameters in order to classify appliances with similar power consumption.

Similarly, Zoha *et al.* [114] have investigated the use of SVM with two kernel functions using different combinations of features and also correlation of acoustic activity within the test environment with device activity. They have used SVM with polynomial kernel with exponential value 2 and RBF with Gamma value 0.01 and compared their results with 1-NN and 10-NN classifiers. In terms of device recognition, SVM with polynomial kernel was able to outperform SVM with RBF for all feature combinations and k-NN classifiers for feature combinations 2 and 3, whereas for feature combination 1, the performance was comparable. The authors have tested the SVM for higher cost functions and though the classification has improved, a lower cost function is generally desirable. Similarly, for acoustic detection, three sets of features have been used for classification, and SVM and GMM have been implemented for classification and it has been shown that SVM performed better for all feature sets. In terms of traditional energy features, the authors have obtained six steady-state load signatures, such as P,

Q, frequency, V_{rms}, I_{rms} and phase angle, and have used, as already discussed, three combinations of them, with the most successful being the last one where all features but frequency have been incorporated. As this thesis focuses on disaggregation using only P, it is difficult of fairly evaluate the method in [114], which uses various features, with the proposed methods in Chapters 4 and 5.

In [115], a hybrid *Support Vector Machine/Gaussian Mixture Model* (SVM/GMM) was proposed for load disaggregation, where GMMs were used to describe the wave distribution and SVM for classification using the power features. The obtained results have shown that polynomial SVM was able to achieve higher accuracy 90.72% compared to the linear, RBF and sigmoid implementations, and furthermore, the hybrid SVM/GMM was able to obtain the best performance compared to SVM without GMM and *Vector Quantisation* (VQ). As the SVM/GMM method in [115] uses a variety of features in order to obtain the final appliance recognition, while the methods of this thesis, even at the case of the combined DTW and k-means method, it only requires the active power in order to disaggregate the load.

Duarte *et al.* [116] have investigated the use of *continuous wavelet transform* (CWT) based on complex Morlet wavelet, in order to obtain the switching voltage transients of appliances, and compared it with the *short-time Fourier transform* (STFT) using RBF SVM. They have shown that the use of CWT improves the classification and can reduce the vector size by $20x$ times. The main difference of this method, compared to other SVM based methods, was the use of CWT in order obtain appliance features, in the specific scenario from three low-current appliances, such as desk fan. The difference between the features used in [116] and this thesis, does not allow a direct comparison of those methods as the active power feature is an important part of how the DTW-based methods work.

Altrabalsi *et al.* [117, 118] have proposed a hybrid NILM method based on a combination of SVM and k-means. k-means is used in order to perform supervised clustering using minimum distance classification, and essentially it reduces the SVM training size, which is used for classification, while k-means in the combined DTW and k-means methods, is used in order to preform the appliance classification. According to the authors, the proposed hybrid-method was able to achieve comparable classification accuracy to that of the HMM method proposed in [35] over REDD, REFIT and

GREEND datasets.

In [119], a NILM method based on SVM for *Air Conditioning* (AC) load monitoring was proposed. This method consists of an initial recognition in order to perform event detection and extract the features that can assist to the correct classification of AC and other appliances. Four features have been proposed: duration of possible event, average value of the waveform, variance of the waveform and maximum value of the waveform and have been used as inputs to an SVM classifier for a final load classification. According to the authors, their proposed method was able to identify successfully the switching off the different statuses of AC. Moreover, this method was focused on the classification of only one appliance and have not provided any information regarding its applicability on different appliances, while the methods of this thesis were able to disaggregate a much wider range of appliances.

Rao *et al.* [120] have considered a variety of methods for both classification and time-series forecasting, such as SVM, DT, Multi-layer Perceptron and *Autoregressive Moving Average Model* (ARMA), to name a few. According to the authors, SVM with linear kernel was able to outperform the comparable classification methods with accuracy of $\sim 30\%$. In order to improve performance, they have incorporated timestamps as an additional parameter and the performance of the SVM algorithm has increased significantly $\sim 70\%$. In terms of current and future prediction, the ARMA method performed significantly better compared to the other methods, and it has obtained accuracy of 90%. For both device identification and consumption prediction, the authors have incorporated demographic data, such as time of day, or day of week, and household information (number of occupants, age, etc.), together with aggregate active power, with the latter being the only feature type used in the methods of this thesis.

Furthermore, in [121], a *Particle Swarm Optimisation* (PSO) Algorithm with Support Vector Machines (PSO-SVM) method has been proposed for load classification. PSO has been used in order to optimise the parameters of SVM and the optimised SVM was then used appliance identification and the identification rate obtained using this method was higher than 95%The authors have extracted features using time and frequency domain analysis of the current waveform, which were obtained by filtering using an LPF with cut-off of $500Hz$ over the original sampling frequency of $5kHz$,

which is still much higher than the sampling rates used in this thesis were only active power in the time domain was obtained.

More recently, Chui *et al.* [122] have proposed a hybrid *genetic algorithm support vector machine multiple kernel learning* NILM approach (GA-SVM-MKL) for load disaggregation. The genetic algorithm is used in order to find the optimal kernel function, using a variety of traditional kernel function, and therefore provide an optimised classifier that can be used load classification. According to the authors, the proposed GA-SVM-MKL method was able to disaggregate 20 commonly used appliances, and obtained overall accuracy $> 91\%$, outperforming traditional SVM methods, that use only one kernel function. Furthermore, the proposed method in [122] uses seven features obtained from the current and voltage waveform:I_{peak}, I_{rms}, *average current* I_{avg}, P, S, Q, and pf, as according to the authors all of these features are in general essential for identification of the different appliances, while this thesis supports the idea that only P could be used for efficient appliance identification.

2.6.1.5 DT Methods

Another classification method, that has been used for load disaggregation, is the use of *Decision Tree* (DT) classifiers. The use of a decision-tree classifier for pattern matching have been investigated in [66], together with a variety of other classification methods (see Section 2.6.1.3), but according to the authors 1-NN with the Fourier regression coefficients was able to achieve the best performance. Furthermore, Liao *et al.* [32] have proposed a low complexity DT-based method and evaluated classification performance using three houses from REDD dataset and three from REFIT dataset. This method requires a very small dataset of aggregate active power and thus requires the least storage and computational resources compared to HMM [35] and the DTW-based method, which is used in this thesis. DT was able to obtain higher accuracy, but both DT and proposed DTW-based method of this thesis were able to outperform HMM.

Alshareef and Morsi [123] have implemented a NILM method that uses the *Discrete Wavelet Transform* (DWT) of the changes of the current during transient state and an *ensemble decision tree classifier* for load classification. Their proposed method is a combination of DT induction algorithm and AdaBoost algorithm. Their DT

implementation uses Gini index for measuring impurities in order to identify the best split between the attributes. The predictions obtained by the multiple DT classifiers is then used as input to AdaBoost, which assigns weights to the predictions of each of the DT classifiers using the error of the classifier. The authors have investigated the use of different order Daubechies (DB) and they showed that DB3 had the best performance with classification accuracy of 95.83% and 93.06% for training and testing respectively. They have also shown that a higher number of trees can improve accuracy, but will increase the computational cost. As the NILM approach in [123] uses wavelet decomposition in order to obtain the relevant features, it is difficult to compare it with the methods presented in this thesis, which use the time series active power signal measured at the smart meter in order to obtain the relevant appliance signatures.

In [124], a multi-class DT classification method was proposed, that uses the greedy Hunt's algorithm, similarly with the DT implemented in [123]. The authors have investigated the use of both actual power (active, reactive and apparent) and changes in power using *Discrete Fourier Transform* (DFT), and have concluded the use of change in power feature can increase classification accuracy by an absolute difference of 26.26%. The change in the power represent the actual transients/edges which occur while switching on/off appliances or when the appliances change cycles. Even though the proposed methods of this thesis use edge detection to obtain the changes in power, the whole consumption during the duration of each cycle is required in order to obtain an event that could be classified as a specific appliance.

In [125], the authors have proposed a new wavelet design and have performed load classification using DT. *Procrustes Analysis* was used in order to determine the wavelet coefficients of length-6 filter, which were used in order to construct the new wavelets. According to the authors, the proposed wavelet design was able to improve the classification accuracy compared DB3. Similar to what have been discussed for the method in [123], the proposed method in [125] uses as features the wavelet coefficients in order to perform classification using DT, while this thesis performs feature extraction using the active power.

2.6.1.6 ANN Methods

Many researchers have investigated the usage of *Artificial Neural Networks* (ANN) and as an extension *Deep Neural Networks* (DNN), as classification methods for NILM. DNN-based methods have gained a lot of interest in recent years, therefore they will be discussed separately in the following section.

In [112], the authors have proposed the use of harmonic current signatures using a variety of Neural Networks methods and SVM, which is different from the approach used in this thesis, that requires low resolution active power for feature extraction. They have investigated the *Multilayer Perceptron* (MLP), RBF based network and SVM using linear, polynomial and RBF kernel functions. All the proposed methods were able to perform sufficiently enough in the experimental tests, though the MLP based method seemed to perform slightly better compared to the rest.

Tsai and Lin *et al.* [63], as already discussed in Section 2.6.1.3, have investigated the use of both BP-ANN and k-NRR, with the latter being selected as more appropriate due to low computational complexity. According to the authors, their BP-ANN method was able to obtain classification accuracy $> 95\%$, but constructing the network is more complicated, as it requires a range of factors, the optimal network has to be found through trial and error and training is required at the initial stage and re-training whenever a new appliance is added. It is important here to note that the methods proposed in this thesis are not affected by the introduction of a new appliance, as any new signature can be compared with the signatures available in the library, and if it does not match to any of the existing signatures, it can be added in the library.

In [41], the *Real Time Recognition and Profiling of Appliances* (RECAP) system has been proposed for monitoring and power disaggregation. The authors have used ANN and were able to obtain accuracy higher than 84% for all experiments in a prototype kitchen scenario. According to the authors, ANN can provide easy extensibility, does not require understanding of appliance behaviour before the implementation, it can be reinforced by using a user feedback input, can handle multiple simultaneous appliance states and the learning process can be automated. The main drawback of their proposed ANN approach was the computational time required for the training process, especially when there are more than 15 appliances, or for the case of appliances with *"long*

signatures", e.g a washing machine. Although the RECAP system have used low resolution data (1 sample per minute), which is comparable to the sampling rates used in this thesis, the authors were able to extract a variety of features, namely P, pf, I_{peak}, I_{rms}, V_{peak}, and V_{rms}, which allows a better classification of the appliances, especially those that may present similarities in some of the features. This was not the case for the methods in this thesis, as only active power was used for classification, which resulted in misclassification of some appliances, that exhibit similar power consumption (see Chapters 4 and 5 for more details).

Chang [28] has investigated the use of the *turn-on transient energy* U_T for feature extraction and *genetic programming* (GP) for obtaining the optimal features. This approach has been implemented using a range of classifiers for ANNs, such as BP, *Probabilistic Neural Network* (PNN), and *Learning Vector Quantisation* (LVQ) classifiers. Both BP and PNN have outperformed LVQ in terms of accuracy (training and testing) and computational time. Although, PNN was the fastest classifier compared to BP, the authors favoured the use of BP, which demonstrated higher accuracy and smaller number of weights and biases for the BP network. With regards to the extracted features, the author claimed 100% recognition accuracy when using the U_T, while the relevant accuracy when using PQ features was much lower.

In [29, 84], the authors have expanded the work published in [28] and proposed a NILM method based on a BP classifier for *multi-layer feedforward neural network* (MFNN). The authors claimed 100% accuracy during both training and testing when real power, reactive power and startup transient (PQU_T) were used for all their case studies, whereas when only real and reactive power was used (PQ), the training and testing accuracy for multiple operations ranged from $51.28\% - 100\%$ and $39.47\% - 100\%$ respectively. All the different methods proposed in [28, 29, 84] are supervised, while the methods proposed in this thesis provide an unsupervised solution, as it will be discussed further in Chapters 3 and 5. Furthermore, this thesis reports the performance of the proposed methods using low resolution data and only active power, while the methods in [28, 29, 84] used sampling rate of $15kHz$ and PQ and U_T or a combination of them as features.

In [126, 127], a NILM method based on BP-ANN and PSO has been proposed. PSO is used in order to minimise training time and improve recognition, which ac-

cording to the authors performed better compared to GP and simple BP-ANN. Furthermore, Chang *et al.* [128] have presented a NILM algorithm which uses *Wavelet Multi-resolutional Analysis* (WMRA) technique, *Parseval's Theorem* and BP-ANN. The authors have used the *Wavelet Transform Coefficients* (WTC) of the transient signals in order to obtain their energy spectra, which they claimed have shown better accuracy and lower computational complexity compared to the use of PQ and PQU_T. The methods proposed in both [126, 127] and [128] have used a variety of features obtained using sampling rate of $6Hz$ and of $15kHz$, respectively, therefore it would be difficult to compare them with the method proposed in this thesis, where only active power was used obtained from datasets with much lower sampling rate.

In [129], the authors have used BP-ANN in embedded NILM system measuring low sampling rate ($1Hz$) P and Q, while the research work presented in this thesis was focused on providing a NILM method using only P. They have used four appliances: TV, refrigerator, air conditioner and rice cooker, and were able to obtain accuracy 98% during the simulation test and 95% during the installation test.

2.6.1.7 DNN based methods

In [57], the authors have presented three deep neural networks architectures for load disaggregation, a form of *Recurrent Neural Network* (RNN) called *"long short term memory"* (LSTM), dAEs and a *regression network* that can estimate the start time, the end time and mean power demand of the first appliance activation (*"rectangles"* architecture). They have compared their work with the *combinatorial optimisation* (CO) algorithm [24] and the *Factorial Hidden Markov Model* (FHMM) [1, 75] as in [34]. According to the authors, dAE and the "rectangles" architecture were able to perform well especially for "unseen" houses, and in the majority of the accuracy measurements were able to outperform CO and FHMM. LSTM was able to outperform CO and FHMM on two-state appliances, but did not perform well on multi-state appliances.

Bonfigli *et al.* [130] have extended the work presented in [57], by improving it both in terms of architecture and algorithm, and performed an "exhaustive" performance evaluation with the *Additive Factorial Approximate Maximum A-Posteriori Probability* (AFAMAP) method proposed in [47] (see Section 2.6.2.2 for more details for the AFAMAP method). dAE was improved by conducting a detailed study of the topology

of the network, using *"pooling"* and *"upsampling"* hidden layers and the *rectifier linear unit (ReLU) activation function* [131] in the output layer and combining the network output by using a median filter. They have used two scenarios for all methods, a *denoised scenario*, where the aggregated signal is the sum of the power of the appliances used for disaggregation, and a *noised scenario*, where the aggregated signal includes also measurement noise and any contributions of unknown appliances. All algorithms were tested using UK-DALE [46], AMPds [132], and REDD [1]. According to the authors, their proposed approach was able to outperform both AFAMAP and dAE from [57], for both scenarios.

Garcia *et al.* [133] have proposed a NILM approach based on *Stacked de-noising autoencoders* (sdAE), where the aggregate power was decomposed in order to obtain power consumption signals for all individual appliances. The proposed model was evaluated using data from four appliances, and the authors claim that they were able to receive promising results in terms of mean absolute error and proportion of energy correctly classified.

Le *et al.* [134] have proposed a NILM approach based on RNN using *Gated Recurrent Unit* (GRU) and compared their proposed method (GRU RNN) with the original RNN approach using UK-DALE dataset [46]. Additionally, they compared their method to previous research works that used the same dataset (Bayes, SVM [81], HMM [47, 75], FHMM [135], FHMM variants [75], FHMM using *Maximum a posteriori probability* (MAP) Algorithm [47]). The proposed GRU RNN approach was able to outperform all previous works and original RNN, and accuracy and F-Measure for energy disaggregation were at ranges $[89\% - 98\%]$ and $[81\% - 98\%]$ respectively.

In [136], a supervised NILM method using deep RNN and LSTM was proposed, and according to the authors, it can estimate the power signal of an appliance or any sub-circuit from the aggregated data, disaggregating both on-off and multi-state appliances. The same authors have proposed, in [137], a combined method using HMM and DNN that does not require event detection and can disaggregate multiple and variable loads from the same aggregate signal. The authors claim that their method performed better than FHMM, as it only requires the knowledge of the target load, whereas FHMM needs a priori knowledge of loads still present in the aggregate load.

Furthermore, Nascimento in [138] has investigated the use of RNNs and *Convolu-*

tional Neural Networks (CNNs) for disaggregation purposes. Similarly, He and Chai [139] have proposed two different deep learning approaches, one using a multi-layer feed forward network with convolutional layer and one using RNN with LSTM, which according to the authors was able to obtain good performance. Another approach using RNN was presented in [140], where a LSTM based RNN was proposed for power disaggregation together with a novel signature which obtains variants of the original signal, which is suitable for multi state appliances, and according to the authors have improved the performance of their proposed LSTM-RNN for those appliances.

In [141], a sequence-to-point learning was proposed, which uses a window from the aggregate signal and obtains a single point of the appliance of interest. The proposed method is using CNN that is able to learn the appliance signatures of a given appliance inherently, which increases the performance of the method in terms of identification. More recently, Murray *et al.* [142] have proposed two types of deep learning networks, one being a GRU architecture and the other a CNN, by using a two-branch layout for both cases. The secondary branch is related to the state estimation and is used for helping the main branch with the consumption estimation. According to the authors, both methods were able to obtain similar performance, with the GRU method being less complex.

Additionally, Barsim *et al.* [143] have proposed a semi-supervised NILM approach, namely UniversalNILM, that utilises modelling appliances using a number of training houses, which obtains detailed at appliance level information and then performs disaggregation to unknown houses, using the obtained training knowledge.

One common factor between the NILM methods using DNN is that they are supervised methods, as they require training of the neural network using known/seen appliances or houses, whereas the DTW-based method proposed in this thesis can be categorised as unsupervised, at it will be further explained in Chapters 3 and 5. Furthermore, the majority of the DNN methods discussed above have demonstrated the ability to use the trained network on "unseen"/unknown houses, which allows the generalisation of the methods and therefore the potential suitability of those methods for an online implementation. The proposed DTW-methods obtain appliance signatures for each house and have not been tested in "unseen" houses in order to assess the potential of the methods in a more generalised scenario.

2.6.1.8 GSP-based Methods

Another approach that have recently been investigated for both supervised and unsupervised NILM is the use of GSP, which actually represents any obtained features using graph signals. Stankovic *et al.* [144] have proposed a supervised event-based GSP method, that uses regularisation on graph approach by maximising smoothness of the graph signal. According to the authors, the obtained results showed GSP can be potentially used to solve the NILM problem, as the proposed method improved the performance compared to the state-of-the-art HMM-based method in the majority of REDD dataset appliances. In [4], the authors have extended the use of GSP for both robust event detection, clustering and feature matching, using a train-less unsupervised approach, see 2.6.2.1 for more details.

He *et al.* [33] have proposed two GSP-based NILM approaches, one based on *"total graph variation minimisation"* and one based on a combination of total graph variation minimisation and *simulated annealing* for further refinement. The proposed methods were able to demonstrate comparable performance compared to the DT-based approach in [32] and HMM-based approach in [35]. In terms of computational cost, the proposed methods were able to outperform HMM-based method.

The GSP-based methods presented in this section will be used for benchmarking purposes of the DTW-based methods proposed in this thesis, therefore a performance comparison will be discussed in details in Chapter 4.

.

2.6.2 Unsupervised Methods

Unsupervised NILM methods, according to [99], can be distinguished in *load classification methods* and *source separation methods*. HMM-based methods are a part of the *load classification methods*, but as these methods have been widely used in NILM research, they will be presented on a separate section, while the rest will be included in the *load classification methods* section. This categorisation can be seen in Figure 2.3.

2.6.2.1 Load Classification Methods

This section presents unsupervised techniques that are not HMM-based or source separation related.

Zhao *et al.* [4] have proposed an unsupervised disaggregation method of the aggregated active power signal using *Graph Signal Processing* (GSP). The proposed GSP-based method does not require training and has been used initially for robust event detection, clustering and finally for feature matching. The authors have used both REDD and REFIT datasets and compared their methods with HMM-based methods presented in [53] and [145], with comparable performance. An important advantage of this approach is the low complexity, simple operation and minimal intervention, as it does not require any labelling or training.

More recently, Zhao *et al.* [5] have proposed GSP filters for pre-processing data (see Section 2.3.2), and a NILM refinement method, discussed in Section 2.7, that can be applied to any event-based NILM methods, and can increase their disaggregation performance. The methods in [4] and [5] have been used for benchmarking purposes and a performance comparison between the proposed DTW-based methods can be found in Chapter 4.

The proposed *Dynamic Time Warping*-based method, that is presented in this thesis, has been published in [146], [147] and [32], where, as it will be explained in the following chapters, appliance signatures were acquired using only aggregate power consumption. DTW performs non-linear mapping between already acquired events and unknown events, and can classify the events by finding the minimum distance between them via dynamic programming.

2.6.2.2 Hidden Markov Model Based Methods

Recently, *Hidden Markov Model* (HMM) and its variations have become popular for (low-rate) NILM, because they are good at modelling the combination of stationary processes, with continuous value data over discrete time (e.g [31, 47, 53, 75, 148]). HMM can probabilistically model sequential data incorporating, in the learning process, the time-dependency in running appliances as well as the transition of the appliance through different states during its operation. According to [31], state-of-the-art HMM-based

methods [47, 53, 75] achieve accuracy of $75-95\%$, work offline, can be supervised or unsupervised, and are not scalable. Furthermore, the complexity of HMM exponentially increases with the number of appliances modelled and the whole model needs to be retrained whenever a new appliance is added [31].

Kim *et al.* [75] have proposed four different methods for low-rate power disaggregation using (conditional) factorial HMM and Hidden semi-Markov models. Using active power measurements collected every $3sec$ together with duration and time of use of appliances, accuracy ranging from 72% to 99% was obtained for seven different houses with up to 10 appliances with an average accuracy of 83%. This method cannot disaggregate base load, such as, for example, refrigerator, and is prone to converge to a local minimum. Similar with the other NILM methods discussed throughout this chapter, the use of multiple features provide better information regarding the appliance usage, allowing the extraction of more distinctive characteristics. Therefore a method like that could not compare with methods, such as the methods of this thesis, that use only one feature (active power).

In [35], a factorial HMM is used for disaggregation of active power load. Although it is presented as an unsupervised NILM method that uses expert knowledge to set initial models for states of known appliances, the models' operation for reliable results strongly depends on correctly setting the a priori-values for each state for each appliance, which in turn is strongly dependent on the particular aggregate dataset on which NILM is being performed. Indeed, a similar factorial HMM-based approach is tested in [1], where it is shown that the disaggregation accuracy drops for up to 25% when different houses are used to set the initial models compared to the case when the same house is used for building the models and testing. Results are reported for REDD dataset [1] with sampling rates of $1min$ and $3sec$.

This implies that the training of the model is carried out on a labelled dataset with unique active power values for each state and each appliance. FHMM is also unnecessarily complex by allowing multiple states per appliance when usually most of the times an on-off state transition is all that is needed for accurate detection. The accuracy with the sampling rate of $1min$ is 53% for microwave and 69% for tumble dryer. The HMM method proposed in [35] will be used for benchmarking purposes in this thesis, therefore a performance comparison can be found in Chapter 4.

More recently, in [53], *Hierarchical Dirichlet Process Hidden Semi-Markov Model* (HDP-HSMM) factorial structure is used for unsupervised NILM removing some downsides of previous approaches, such as the need for training and modelling different power modes per device (as in [75]) instead of binary values only. The results are reported for five devices using $20sec$ resolution with 18 24-hour segments across four houses from the REDD dataset [1] obtaining disaggregation accuracy of 81% outperforming the EM-based method of [1].

In [47], an unsupervised AFAMAP inference algorithm is proposed using differential FHMMs. First, all snippets of active power data are extracted using a threshold and modelled by an HMM; next the k-NN graph is used to build nine motifs that are treated as HMMs over which AFAMAP is run. The results show average accuracy of 87.2% using 7 appliances and sampling rate of $60Hz$.

Guo *et al.* [149] have proposed a model named *Explicit-Duration Hidden Markov Model with differential observations* (EDHMM-diff), in order to detect and estimate individual appliance loads using aggregated data, together with a *specialised forward-backward algorithm*. The proposed model was able to model the state durations and to overcome overlaps on aggregated data, when multiple appliances are on simultaneously.

Aiad and Lee [145] have proposed an FHMM that uses device interactions in order to enhance load disaggregation. The authors have compared their proposed method with FHMM method without device interactions and they were able to acquire better or comparable disaggregation accuracy in all cases.

In [150], the authors have presented a disaggregation algorithm using *super-state HMM* and a new Viterbi algorithm with sparse transitions, which was able to compute efficiently sparse matrices with large number of super-states. For testing purposes, AMPds [132] and REDD [1] datasets have been used and the proposed algorithm was able to obtain significantly higher accuracy compared to the methods proposed in [1] and [53]. This method can be applied for real-time disaggregation using low sampling rates, while the DTW-based methods proposed in this thesis would not be appropriate solutions for this scenario due to their computational complexity, which will be further discussed in Chapter 5.

Cominola *et al.* [151] have proposed a novel *Hybrid Signature-based Iterative Disaggregation* (HSID) algorithm that combines FHMMs and *Iterative Subsequence Dynamic*

Time Warping (ISDTW). The FHMM module performs initial disaggregation of the aggregated data into *"2-state single-appliance piece-wise constant trajectories"*. ISDTW is applied to this initial disaggregation in order to reshape according to typical consumption pattern together with appliance power range and usage duration. The authors have presented both a supervised version, which requires appliance-level measurements and a semi-supervised version, which uses aggregate data in order to retrieve appliance-level measurements. This work will be further discussed in Chapter 3, for comparison purposes with the proposed DTW-based methods used throughout this thesis.

In [152], a NILM approach was proposed using a *bivariate Hidden Markov Model* for appliance modelling, which uses joint active-reactive power signals, and an alternative formulation of the AFAMAP algorithm used in [47] for disaggregation. The authors have used the AMPds dataset (in noised and de-noised condition) and compared their proposed NILM method with the AFAMAP [47], Hart's algorithm [24] and Hart's algorithm using *Maximum A Posteriori* (MAP) techniques. The proposed method was able to outperform the above methods by +14.9%, +21.8%, and +2.5% (respectively) for de-noised conditions and by +25.5%, +51.1%, and +6.7% for noised conditions. As the proposed method in [152] use both active and reactive power in order to disaggregate the load, it would be difficult to compare it with the work in this thesis, where all methods utilise only active power.

One common factor between the majority of the HMM-based methods, discussed in this section, is that they offer an unsupervised NILM solution, similar to the methods proposed in this thesis (see Chapters 3 and 5 for more details). Furthermore, most of these methods focused on low resolution disaggregation, as were the proposed DTW-based methods. HMM generally approaches the NILM problem as a probabilistic problem, while the DTW-based methods provide a solution based on time series signal processing.

2.6.2.3 Source Separation Based Methods

Source Separation based methods solve the load disaggregation problem by either using unsupervised or *"blind"* source separation techniques, where there is no a-priori information about the sources, or try to solve a convex optimisation problem using appropriate basis functions. In both cases, the aim is to use aggregate data in order to reconstruct all source signal and specify each appliance [99, 153].

Figueiredo *et al.* [50] have examined energy disaggregation as a single-channel source separation based on *Non-negative Matrix Factorisation* (NMF). For further performance improvement, the authors have incorporated additional information about dependencies between appliance usage, and have created a technique called *Source Separation Via Tensor and Matrix Factorisations* (STMF). This method assumes linear mixing of the source power signal, and uses a tensor representation and PARAFAC factorisation in order to extract any dependencies between the sources. A comparison between the proposed STMF method and the DDSC method proposed in [74] have shown a much smaller disaggregation error for the case of the STMF method.

Wytock and Kolter [153] have proposed a single-channel source separation approach for load disaggregation using convex optimisation methods in order to identify the correlation between the input features of each source signal and the *"unobserved signal decomposition"*. The authors have used a representation of each individual signal as a linear function of some component bases, a *contextually supervised method* based on temperature time series and the non-linear dependence represented with *radial-basis functions* (RBFs), is defined as basis for the energy consumption. The theoretical analysis has shown that accurate separation can only occur when the features are linear independent for different signals.

Gonçalves *et al.* [154] have proposed a linear *Blind Source Separation* (BSS) approach in order to disaggregate each appliance from the aggregated data, using steady state real and reactive power features. The authors have investigated both *Genetic k-means* and *Hierarchical Agglomerative* clustering methods, in order to obtain automatically all representative appliance clusters. Following the clustering process, the *matching pursuit* (MP) is employed for source reconstruction, where iterations are performed by the MP algorithm in order to minimise the distance between unknown

events and possible clusters. The study's output has shown that the genetic k-means approach had better performance compared to the agglomerative based clustering method. According to the authors, source reconstruction was successful for large appliances, as they were able to form clusters on their own, whereas small appliances, or appliances with very similar consumption were wrongly grouped in the same cluster. Additionally, multi-states appliances have been grouped in different clusters depending on the state that was active, something that can be avoided by modelling the appliances as a set of two-states appliances [31, 99].

One of the main difference between the DTW-based methods of this thesis and the NILM methods that use source separation is that the former treat the load disaggregation as a classification problem, while the latter as a source separation problem. Source separation methods provide unsupervised solution for the the NILM problem, similarly with the DTW-based method, as it will be discussed further in Chapters 3 and 5.

2.7 Data Post-processing

More recently, many researchers have proposed various *post-processing* techniques for further refining the NILM classification results, in order to improve their performance, as seen in [5, 33, 47, 52, 58]. According to [58], a refinement process could be applied by iterating over the disaggregation algorithm and using previous knowledge in order to identify operational sequences, such as a washer dryer is following the washing machine operation.

Kolter *et al.* [47] have refined their NILM results, using inference approximation by incorporating AFAMAP in their proposed FHMM-based NILM method, a method that was able to outperfom alternative inference methods, such as *structured mean field* (SMF). Furthermore, a residual analysis has been employed in [52], using the quantisation error waveform, which is calculated during the pre-processing stage, as already mention in Section 2.3.2. The residual analysis can provide an indication of cases of inaccurate disaggregation, allowing repetition of the previous steps.

Simulated Annealing (SA) has been incorporated in the GSP-based method proposed in [33, 155], in order to further refine two-state appliance identification, by optimising the difference between actual power and initial power estimation. Zhao

et al. [5] have presented a graph-based refinement approach, that can address the misclassification of events from appliances with similar operational range. This method is generic and can be applied to any event-based NILM methods, similarly with the proposed pre-processing techniques, as seen in Section 2.3.2. Based on the reported results in [5], the proposed method both pre-processing and refinement, was able to improve the performance of several NILM algorithms, such as DT [32], supervised GSP [33], unsupervised GSP [4] and FHMM (*Factorial Hidden Markov Model*) [34].

2.8 Performance Evaluation

As with every new research field, evaluation of the performance of the different NILM methods is critical and can provide the basis for the further development and benchmarking. According to Kim *et al.* in [75], Zeifman and Roth [30] and Makonin in [156], evaluation metrics are not consistent through NILM related bibliography. Makonin and Popowich in [44] presented a unified solution for this problem by summarising all the different performance evaluation methods that have been used in the past, proposing some additional methods and research guidelines that can assist to more accurate assessment of NILM-related projects. More recently, Pereira *et al.* [40] have reviewed the most recent performance evaluation metrics, and their work can provide a good reference for NILM researchers in order to decide how to evaluate their proposed methods.

The different methods that can be used for accuracy evaluation of a NILM method can be summarised in the following: *basic accuracy*, *classification accuracy*, *estimation accuracy*, *data noise* and *ground truth and bias*, as found in [44]. *Basic accuracy* has been used in [29, 63, 132], but according to Kim *et al.* in [75], the accuracy results were affected by appliances that the on-event or specific state is a rare event. *Classification accuracy* and *estimation accuracy* will be discussed in Sections 2.8.1 and 2.8.2, with a focus on the metrics that will be used for performance evaluation through out this thesis.

Makonin and Popowich in [44] proposed the use of a *Data noise* measure in order to evaluate the performance of a load disaggregation algorithm. In terms of *ground truth and bias*, Makonin and Popowich in [44] suggested detailed description of the

data used (eg. publicly available datasets, such as REDD [1]), description of the data cleaning process used, explanation of whether the data were de-noised or not, and a *10-fold cross validation* as in [157], which is commonly used in data mining community (see [44] for more details). Furthermore, the authors have proposed the *percent-noisy measure* $(\% - NM)$, as seen in [44], while Zhao *et al.* [5] have extended the concept of NM by proposing the *Disaggregation Error Measure* (DEM), which can provide a better understanding of the performance of a NILM method in correlation to the levels of noise present in the used dataset.

2.8.1 Classification Accuracy

Classification accuracy measures, such as *F-Measure* (a.k.a. *F-score* or F_1-*score*), *confusion matrix*, have been applied into NILM research, as in Figueiredo *et al.* [106]; Berges *et al.* [27]; Kim *et al.* [75], in order to assess how accurately NILM algorithms can correctly identify appliances ON/OFF states and multiple states. Thus *F-Measure*, often used in information retrieval and text/document classification ([158]), is currently being used as one of the main evaluation measures in recent NILM publications, as seen in [40, 44, 159].

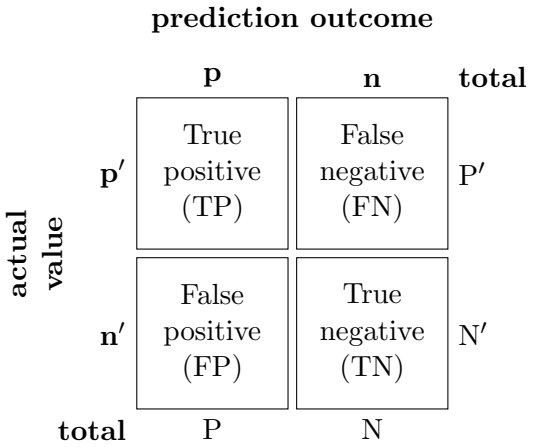
Confusion matrix is a common classification method in machine learning and is regarded as a suitable measure for the NILM problem [34]. Figure 2.4 explains graphically the four categories generated when using a confusion matrix. *True positive* (TP) presents the correct claim that an appliance was used, *false positive* (FP) represents an incorrect claim that an appliance was used, and *false negative* (FN) indicates that the appliance used was not identified.

The evaluation measures, commonly used in NILM, are derived from the confusion matrix found in Figure 2.4 and include *Precision* (PR), *Recall* (RE) and *F-Measure* (F_M). *Precision* is the positive predictive values as in 2.2 and *recall* or *sensitivity* is the true positive rate as in 2.3. *F-Measure* is the harmonic mean of *precision* and *recall* and can be found in equation 2.4:

$$PR = \frac{TP}{TP + FP} \tag{2.2}$$

$$RE = \frac{TP}{TP + FN} \tag{2.3}$$

Figure 2.4: Confusion Matrix.



$$F_M = 2 * \frac{PR * RE}{PR + RE}, \tag{2.4}$$

Normally, these metrics are applied for evaluation of the classification accuracy per appliance. Furthermore, F_M can also be expressed directly as a function of TP, FP, FN, as seen in equation 2.5:

$$F_M = \frac{2 * TP}{2 * TP + FP + FN} \tag{2.5}$$

A revised method of F_M with variants has been proposed in [75], where TP is divided into the *accurate true positive* (ATP) and the *inaccurate true positive* (ITP), which has beed adapted in [4, 5, 44]. Further details can be found in [4, 5, 40, 44, 75].

Another variation of F_M is the use of either the *micro-average* or *macro-average* F-Measure, $F1_{micro}$ and $F1_{macro}$, respectively, which provide a measure of the overall classification performance. *Micro-average* F_M, as found in equation 2.6, is calculated using the summation of all TP, FP and FN values, and *macro-average* F_M is calculated by averaging across the F_M per appliance (see equation 2.7) [40].

$$F1_{micro} = F1\left(\sum_{i=1}^{M} TP_i, \sum_{i=1}^{M} FP_i, \sum_{i=1}^{M} FN_i\right) \tag{2.6}$$

$$F1_{macro} = \frac{1}{M} \sum_{i=1}^{M} F1(TP_i, FP_i, FN_i) \tag{2.7}$$

, where i represents the appliance i from a set of appliances $\mathcal{M}$, and M is the total

number of appliances in the set.

$F1_{micro}$ has been used as an overall classification accuracy in [32, 146, 147], in order to evaluate the performance of DTW-based methods, presented in both Chapter 4 and 5, a DT-based [32, 160] method and HMM from [35].

Batra *et al.* [34] have proposed *Hamming Loss* as a classification measure, which measures the total information lost, when the appliances are incorrectly classified over the dataset used for disaggregation. Details about this measure can be found in [34, 40].

2.8.2 Estimation Accuracy

Another important method for the evaluation of a NILM algorithm are accuracies based on power estimation, which can assess the algorithm in terms of how accurately it can estimate the consumed power when compared to the actual consumption. *Root mean square error* (RMSE) is one of the proposed estimation accuracy measures and has been used in [34, 35, 161, 162]. Furthermore, a normalised measure of the difference between the estimated and actual power consumed from the ith appliance, known as *normalised disaggregation error* (NDE), has been used in [35, 47, 163].

Estimation Accuracy or *Total Energy Correctly Assigned* (*TECA*) has been first introduced by Kolter and Johnson in [1], and used by Johnson and Willsky in [53], Kelly and Knottenbelt in [57] and Makonin and Popowich in [44]. *Acc* can be described mathematically as follows:

$$Acc = 1 - \frac{\sum_{t=1}^{N}\sum_{i=1}^{M}|\hat{p}_t^i - p_t^i|}{2\sum_{t=1}^{N}\sum_{i=1}^{M}p_t^i} \tag{2.8}$$

, where $\hat{p}_t^i$ represents the algorithm's prediction of power consumption for the ith appliance at time instance t, p_t^i is the power consumed at time instance t from the i th appliance, M is the number of appliances in the set of appliances $\mathcal{M}$ and N the number of samples.The factor 2 in the denominator, according to [47], comes from the fact that the absolute value in the numerator will *"double count errors"*, since $\sum_{t=1}^{N} p_t^i = \sum_{i=1}^{M} \hat{p}_t^i$.

The total estimation accuracy, as defined in equation 2.8, can be adapted in order to evaluate the performance at appliance level, as proposed in [44]:

$$Acc_i = 1 - \frac{\sum_{t=1}^{N}|\hat{p}_t^i - p_t^i|}{2\sum_{t=1}^{N}p_t^i} \tag{2.9}$$

, which represents the estimation accuracy of appliance i. This measure has been used in most recent works, found in [4, 5, 33, 155].

Furthermore, Zhao *et al.* [5] have proposed a new metric in order to assess the disaggregation performance of their proposed algorithms which shows the error rate of the estimated total consumption per appliance with regards to the ground truth total consumption per appliance. The *error rate of the total power consumption* (TER) can be expressed mathematically using the equation 2.10:

$$TER = \frac{\left|\sum_t^N p_t^i - \sum_t^N \hat{p}_t^i\right|}{\sum_t^N p_t^i} \tag{2.10}$$

, where similarly $\hat{p}_t^i$ is the estimated power consumed at time instance t from appliance i, p_t^i is the ground truth power consumed at time instance t from appliance i and N the number of samples.

Piga *et al.* [164] have introduced the measure of *estimated energy fraction index* (EEFI), which provides the fraction of the energy assigned to the ith appliance and *actual energy fraction index* (AEFI),which provides the fraction of the actual energy consumed by the ith appliance [40, 159]. More details regarding the specific metrics can be found in [40, 159, 164].

2.9 Summary

The aim of this chapter was to provide to the reader an extensive background for NILM and a better understanding of both the problem statement and the different building blocks relevant to the NILM process. Sampling rate affects both the monitoring techniques and sensors that can be used for data acquisition and the features that can be acquired and used for appliance classification. NILM researchers have used a wide range of sampling rates from Hz to MHz, but more recently, there is a focus on low sampling rates, as the vast majority of "off-the-shelf" smart meters can acquire data using low sampling rates.

Furthermore, common pre-processing techniques (e.g. interpolation, data filling, and filtering) and their NILM applications have been reviewed, as NILM data often require cleaning, smoothing and de-noising. Following the pre-processing stage, every

NILM method performs event detection, either in the form of edge detection or probabilistic edge detection. Edge detection represents the most simplified form of event detection, as it can detect occurring events by using the rising and falling edges in the power data. The edge detection method used for the implementation of the proposed NILM-methods in this thesis is based on this simplified version and will be further analysed in Chapter 3.

Feature extraction, as already discussed, can be affected by the granularity of the acquired data. Data acquired using low sampling rates, can provide researchers with features that are based on the steady-state of each appliance, e.g. active-reactive power, and V-I and lower harmonics, while those acquired using higher sampling rates can provide transient features, such as start-up current, higher harmonics, etc. The research contributions of this thesis focus on real/active power features, acquired from aggregated data with low granularity.

An extended review of the different classification methods that have been used in recent bibliography has been presented throughout this chapter, in order to show the different approaches taken by researchers, eg. machine learning based or signal processing based, supervised or unsupervised. A basic categorisation of NILM classification methods has been presented, in addition to information about methods based on optimisation and pattern recognition, such as SVM, Neural Networks, etc., and methods based on HMM and source separation. Post-processing has also been discussed, as more recent research works have proposed various methods for refining the disaggregation results, in order to increase the overall performance of a NILM method.

Finally, this chapter has provided a review of the different measures used for performance evaluation among NILM researchers, with a focus on the metrics used in this thesis. In terms of evaluation measures, it has been decided to use *Precision* (PR), *Recall* (RE), *F-Measure* (F_M) and $F1_{micro}$ for classification performance and, in terms of estimation accuracy, the estimation accuracy per appliance Acc_i and the newly proposed TER measure, as found in [5]. The classification accuracy measures will be able to provide information regarding the correct/incorrect classification of the events and the estimation accuracy Acc_i will be able to assess the performance of the proposed NILM method in terms of how accurately it can estimate the consumed power per appliance, when compared to the actual consumption.

Chapter 3 will present the algorithmic background for the proposed DTW-based method, and comparison of the proposed technique with similar DTW-based works in the NILM field.

.

Chapter 3

Disaggregation using Dynamic Time Warping

3.1 Introduction

Dynamic Time Warping (DTW) is a time-series based approach that uses *dynamic programming* (DP) in order to find best alignment between two time series or sequences. By minimising the distance between them, the algorithm finds the optimal mapping path and the sequences are *"warped"* non-linearly to match each other [165]. For this purpose, DTW is an ideal algorithm for comparing vectors of different lengths and non-identical values.

Let $\mathbf{p}$ and $\mathbf{q}$ be two time series with length n and m respectively. The time series can be represented by the following vectors:

$$\mathbf{p} = p(1), p(2), \ldots, p(n) \tag{3.1}$$

$$\mathbf{q} = q(1), q(2), \ldots, q(m) \tag{3.2}$$

The DTW algorithm finds the best mapping path using the following constraints and formulas, as found in [165] and [166]

- For all $i, j > 0$, $D(i, 0) = D(0, j) = \inf$ and $D(0, 0) = 0$.
- For any given node (i, j) in the mapping path, the possible fan-in nodes are restricted to $(i-1, j)$, $(i, j-1)$, $(i-1, j-1)$.
- As a boundary constraint, the mapping path starts at $(1, 1)$ and ends at (n, m).

- $d(i,j) = \sqrt{(p(i) - q(j))^2}$ is the distance between points $p(i)$ and $q(j)$, where $(1 \leq i \leq n)$ and $(1 \leq j \leq m)$.

- $D(i,j)$ is the accumulated DTW distance between points $p(1)$ to $p(i)$ and points $q(1)$ to $q(j)$ and $D(n,m)$ is the final accumulated distance between the two vectors.

$$D(i,j) = d(i,j) + min\{D(i-1,j), D(i-1,j-1), D(i,j-1)\} \tag{3.3}$$

Originally, DTW has been used in speech recognition in order to compare different speech patterns, see [167] and [168]. More recently, DTW has been successfully applied in data mining, information retrieval and pattern recognition [165]. In [169] and [166], DTW has been used in a hybrid algorithm combined with HMM in order to categorise residential water end use events, which is highly relatable with the potential use of DTW as a tool for NILM disaggregation, where the different appliance usage events can be classified. During the specific implementation, DTW was used after HMM only to detect the so-called mechanical end-use categories (clothes washer and dish washer), which were not detected by HMM. According to Nguyen *et al.* [169] and [166], DTW requires longer processing time but was able to classify unclassified events, and eventually increased the overall accuracy of their disaggregation.

The aim of this chapter is to propose an unsupervised NILM method that is based on DTW algorithm. First, a formulation of the NILM problem will be defined, together with the motivation behind the use of DTW and its potential suitability for solving the NILM problem. A short review and comparison of the proposed method to other DTW-based NILM works will be able to assist in a better understanding of any differences in terms of approach and implementation of the algorithm and furthermore to clearly present the novel contribution of the proposed method. This proposed DTW-based method will be explained based on the NILM steps, that have been described already in Figure 1.3, describing data pre-processing, event detection, feature extraction and finally classification using pattern matching.

3.2 NILM Problem Formulation

NILM methods are focused on recognising the operation of individual appliances and estimating the power contribution of these appliances using the measured aggregate load. Let p_t be the measured aggregate active power at time instance t. Without loss of generality, this measured aggregate power p_t can be expressed by the following formula:

$$p_t = \sum_{i=1}^{M} p_t^i + n_t \tag{3.4}$$

, where p_t^i is the estimated power consumption of appliance i at time instance t and n_t is the noise at time instance t, where $t = 1, \ldots, N$. The noise term includes the measurement noise due to acquisition or communications errors, the base load and the consumption of all unknown appliances at time instance t. Note that equation 3.4 can be generalised for other power signals, such as reactive power, etc.

The operation of an appliance i, either during the ON state for ON/OFF appliances, or during the different states for multi-state appliances, can vary in terms of duration and power consumption, depending on the user's specifications and appliance settings. This operation can be expressed as a time series, found in equation 3.5:

$$\mathbf{p^i} = p_1^i, p_2^i, \ldots, p_T^i \tag{3.5}$$

, where each element represents the active power consumed by the appliance i at any given time instance of the duration of the current appliance operation. T corresponds to the number of samples that the appliance was operating either at the ON state, or at a specific state of a multi-state appliance. For different operations of the appliance i both the power consumed and the T can be different, as already discussed.

The different duration and non-identical values of appliance operations motivated the investigation of DTW for addressing the NILM problem, as the algorithm is very efficient into comparing time series of different lengths and non-identical values, as described in 3.1. Therefore, any two appliance operations could be compared and by minimising the distance between them, DTW would be able able to find the best alignment and consequently classify any operation to the relevant appliance.

3.3 DTW in the NILM Problem

DTW has been proven an effective method for pattern matching, as already mentioned in Section 3.1, in speech recognition, data mining, information retrieval and even for the case of residential water disaggregation [166, 169]. It is an efficient approach for comparing data sequences with different length, thus it can be easily adapted for the case of load disaggregation, where there is a variety of appliances, with different operational times and power consumption, thus time-dependent sequences, with different lengths and non-identical values.

In the context of smart grids, it has been used in [170] for clustering load profiles of electricity customers. The authors have implemented DTW as a distance measure and performed clustering using k-means and a simple *top-down partition* algorithm (TS-Part). The obtained results have been compared with the use of the Euclidean distance, as distance measure, using the same clustering algorithms. According to the authors, DTW outperformed Euclidean distance and provided better clustering solutions, with the best performance when used in combination with the TS-Part algorithm.

The potential application of DTW for disaggregation has been shown in [171], where the authors have used energy consumption, rising edge count and DFT as diagnostic features calculated over a time window. They were able to show that DFT is not required for load disaggregation, as the load variations within a window are very small, compared to similar variations of features used for speech related applications. The algorithm has been tested using one-day's data with sampling rate of 10sec in a laboratory environment, which makes it difficult to evaluate the actual performance using only these limited results. Similarly, Zaidi *et al.* [172] have investigated the performance of DTW and HMM algorithms, using as features, the average consumption, edge count and the percentage energy consumption. Both algorithms were tested using a limited experimental dataset obtained in a laboratory environment similarly to [171]. According to the authors, in contrast to HMM, DTW tends to be affected by the usage behaviour of the appliance.

More recently, Basu *et al.* [173] proposed a *time series distance-based approach* for power disaggregations. DTW was only incorporated as a distance metric, similarly to [170], along with other distance measures, such as Euclidean distance and *temporal*

correlation (TC). For classification purposes, a k-NN classifier was used and the performance for all measures was compared to the results obtained using HMM. Time series distance-based k-NN was able to outperform in the majority of the cases HMM, and the best performance was obtained when k-NN was combined with DTW. The DTW-based method, proposed in this chapter, is using DTW for classification via pattern matching and not solely as a distance measure. More recently, Teeraratkul *et al.* [174] have proposed a shape-based clustering approach using DTW, in order to identify energy consumption patterns with regards to the regular consumer behaviour.

Cominola *et al.* [151] proposed a *Hybrid Signature-based Iterative Disaggregation* (HSID) method, which combines the use of FHMM and *Iterative Subsequence Dynamic Time Warping* (ISDTW). FHMM is used for the initial approximation of the end-use trajectories and ISDTW then processes the initial results, in order to perform appliance classification by matching using the typical power consumption pattern of each appliance. The authors have implemented a supervised and a semi-supervised version of the proposed HSID method. The supervised HSID method requires appliance-level measurements for calibration purposes and the semi-supervised version is retrieving appliance-level information using the aggregate signal obtained from the smart meters. According to the authors, the supervised HSID was able to disaggregate multiple appliance operating simultaneously and to outperform the state-of-the-art FHMM in [34], without increasing computational complexity. In terms of accuracy, the supervised method was able to accurately reproduce the consumption trajectory. For the case of the semi-supervised HSID, the authors claim that a further research is required, as the algorithm had difficulty identifying appliances with variable signatures, such as washing machine (different programs can create different signatures), using only a singe-event signature extraction.

As already mentioned above, the proposed DTW-based method is based on the concept of using solely DTW for disaggregation purposes, whereas Cominola *et al.* [151] have used an ISDTW combined with FHMM.

Another interesting implementation of DTW for Non-Intrusive Appliance Monitoring was proposed in [175], where the authors proposed a DTW-based non-intrusive load transient identification. DTW was used in order to measure the similarity between the *Transient Power Waveform* (TPW) signatures and combined with k-NN identifier, the

method was able to obtain improved accuracy. Three DTW-based integrated distances were used, combining both active and reactive TPW signatures, which were fed to the k-NN algorithm. The authors were able to better identify different operating states of appliances. The DTW-method proposed in this thesis focuses on performing disaggregation for low sampling rates using only active power consumption, which does not allow the use of transient. For this purpose, it would be impossible to compare it with the method proposed in [175], where the authors use transient signatures, both active and reactive, and additionally incorporate k-NN for classification.

3.4 Proposed Unsupervised DTW-based NILM Method

As with every NILM method, it is important to discuss and explain the different steps that are required for the implementation and evaluation of the proposed DTW-based method. Therefore, we next discuss data pre-processing, event detection, feature extraction, classification via pattern matching and performance evaluation, in the order they occur as presented in the block diagram in Figure 2.1.

3.4.1 Data Pre-processing

Data pre-processing is instrumental for removing noisy and corrupted data from any datasets and generally for preparing them for any further processing in order to extract meaningful information. There are various techniques that can be used for this purpose, and it is up to the researcher to identify and apply the most suitable approach for implementing their research, as seen in Section 2.3.

As already mentioned, the NILM datasets used for disaggregation purposes in the presented research work is the REDD dataset [1], that includes smart metering data from US households and the REFIT dataset [2, 3], that includes data obtained from UK households. For the case of the REDD dataset, the data have been downsampled to 1min resolution, but there was no further processing for removing noise or base load. For the UK households, there was no pre-processing, as Murray *et al.* [2, 3] have published a clean version of the dataset. More details with regards to the pre-processing of both datasets can be found in the following Chapter in Section 4.2.

3.4.2 Event Detection

Following data acquisition and data pre-processing, according to Figure 2.1, a NILM approach requires an *event detection* step. Event detection refers to the detection of any changes in the aggregate power consumption time series, that occur during the operation of one or multiple appliances.

In terms of mathematically formulating event detection, we can state that if

$$|p_t^i - p_{t-1}^i| \geq W, \tag{3.6}$$

, we say that the appliance i has *changed state* at time instant t, where W is an adaptive threshold. For single-state (On-Off) appliances, this means that the appliance was switched on if

$$p_t^i - p_{t-1}^i \geq W \tag{3.7}$$

or off if

$$p_{t-1}^i - p_t^i \geq W. \tag{3.8}$$

For multi-state appliances, such as washing machine,

$$|p_t^i - p_{t-1}^i| \geq W \tag{3.9}$$

indicates that the appliance has transited from one state to another.

Threshold W depends on the appliance set $\mathcal{M}$ and must be set low enough so that for all i, if $|p_t^i - p_{t-1}^i| \leq W$, Appliance i did not change its state and, otherwise, it did change its state. All appliances that operate below threshold W will not be detected. W is set based on the minimum state transition that needs to be detected and the maximum variation of the active power within one appliance state across all appliances' states; that is

$$W = \max\{\min_{i \in \mathcal{M}} \mathbf{p^i}, \max_{i \in \mathcal{M}} |\max(\mathbf{p^i}) - \min(\mathbf{p^i})|\}, \tag{3.10}$$

, where $\mathbf{p^i}$ is a vector of readings of appliance i. Note that the value of W depends on the set of available appliances, and since the disaggregation process works iteratively by detecting one appliance at a time and removing it from $\mathcal{M}$, W will be adaptively changed whenever one appliance is detected and removed. W is automatically obtained

using edges in each iteration. It is important here to state that for each W, appliances with similar consumption (rising and falling edges) can be obtained during the same iteration and therefore those appliances estimated consumption will be removed from the aggregate load before the next iteration.

An *event* occurs whenever an appliance changes its state. Edge detection is used to detect events by comparing $|p_t - p_{t-1}|$ with W. An event started at time l_s and ended at l_e if an appliance changed its state at l_s and l_e, and

$$|[p_{t=l_s} - p_{t=l_s-1}] + [p_{t=l_e} - p_{t=l_e-1}]| \leq C, \tag{3.11}$$

, where C is a parameter that is upper limited by $\max_{i \in \mathcal{M}} |\max(\mathbf{p^i}) - \min(\mathbf{p^i})|$ and hence is always smaller or equal to W.

For $t = 2, \ldots, N$, let

$$\Delta p_t = p_t - p_{t-1} \tag{3.12}$$

The event detection process is summarised in Algorithm 3.1.

Algorithm 3.1 Event Detection: For any aggregate sample window **p** return a matrix of detected events **EVENT** where each event is defined by the starting and ending time.

```
function EVENT DETECTION(p(StartTime : EndTime), W)
    k = 1
    for l_s = StartTime : EndTime do
        if |Δp| ≥ W then
            i = l_s
            C = |Δp(l_s)|
            repeat
                i++
            until |Δp_t| ≥ W ∧ (3.11) ∨ t == EndTime
            if i == EndTime then
                exit('No end of the event found')
            end if
        else
            C = W
        end if
        EVENT(k, 1) = l_s, EVENT(k, 2) = t, k++
    end for
    return EVENT
end function
```

Note that parameter C is set as 20% or 5% of the detected rising edge. The selection between the two options depends on the value of the detected edge $|\Delta p_t|$. If $|\Delta p_t| \leq$

300, thus the possibility of a lower consuming appliance operation, parameter C is set $C = 0.2|\Delta p_t|$, in order to capture the variations of this appliance in the aggregate data. Else if $|\Delta p_t| \geq 300$, is set $C = 0.05|\Delta p_t|$, because if C is set higher, the starting and closing edges of higher consuming appliances will create inaccurate events. It is important here to explain that the edge detection step can identify as separate events, events that occur for one instance.

The function returns a $R \times 2$ matrix of events, **EVENT**, where row i contains the i-th detected event characterised by its starting time $EVENT(i, 1)$ and ending time $EVENT(i, 2)$.

After obtaining the available events, extracting the features, as described in the following section, and removing the relevant power loads from the aggregate data, the algorithm repeats the edge detection step by allowing the incorporation of the closest instants $l_e \pm 1$ and if $|\Delta p(l_e - 1)| + |\Delta p(l_e)| \leq C$ or $|\Delta p(l_e)| + |\Delta p(l_e + 1)| \leq C$ allow the closure of the event either at l_e or at $l_e + 1$ respectively.

Figure 3.1 shows an example of rising and falling edges, that are used during event detection. The specific edges correspond to operations of microwave and kettle, that occur simultaneously. The rising edges, are represented with blue, and the falling edges with red. The circle on the two first edges, explains graphically the search and inclusion of the previous instances of a rising edge, in order to identify the corresponding start and end event. This operation closes using the falling edge, that occurs exactly after the rising edge of the kettle at around 3000 watt.

3.4.3 Feature Extraction

After detecting all events, it is necessary to extract features from each event for further processing. The proposed DTW-based method uses as features the entire active power time-series signature of each event/appliance operation, which is stored into a database.

Let R be the number of rows in the matrix **EVENT** returned by Algorithm 3.1. Then, feature extraction for the proposed DTW-based method is summarised in Algorithm 3.2. The algorithm takes the matrix of detected events as inputs and performs feature extraction, and returns a matrix of extracted features **E**, which contains R rows. As the actual length of each extracted feature varies depending on the number of power samples in each detected event, in every iteration the length of the rows is set

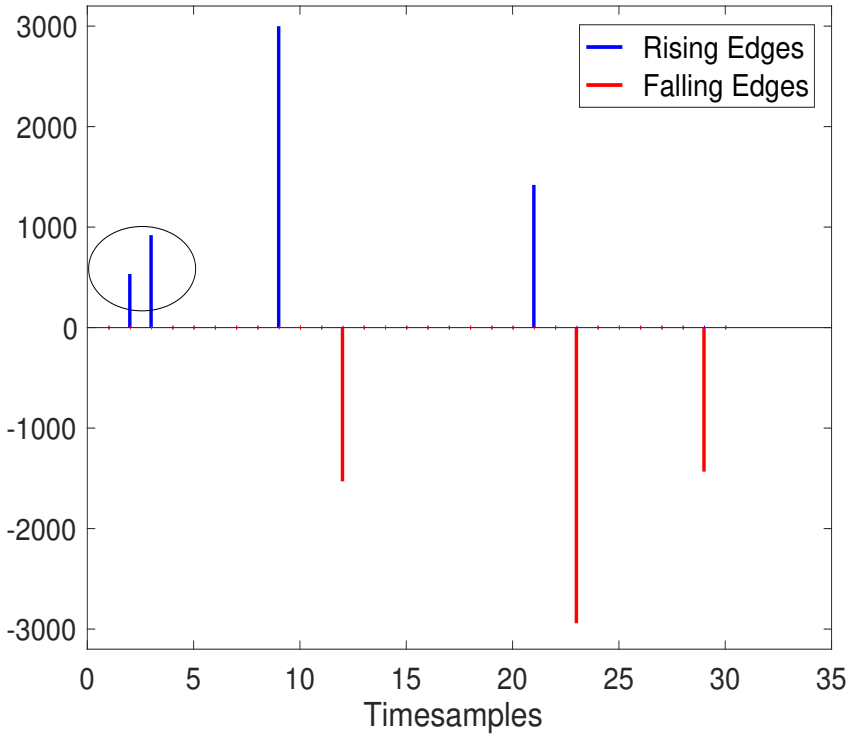


Figure 3.1: Example of the rising and falling edges used during edge detection.

equal to the length of the largest event, and all empty fields are replaced with 0 . In order to recover the intermediate points between the start and end time of an event, the algorithm uses the aggregate data and the variations that occur between the two points, as long as there are $|\Delta p_t| < C$, where $i = l_s + 1, \ldots, l_e - 1$. This could also be performed by averaging the power consumption between the start and end time of the event.

Algorithm 3.2 Feature Extraction: For a given set of events and proposed DTW-based method, return extracted features.

function FEATURE EXTRACTION(**EVENT**, **p**)
  for $k = 1 : R$ **do**
    $E(k,:) = |\Delta p(EVENT(k,1) : EVENT(k,2))|$
  end for
  return **E**
end function

3.4.4 Disaggregation using DTW

The proposed DTW-based method performs disaggregation into two different phases: *appliance library creation* and *pattern matching of the disaggregated load*, which will be described in the following sections.

3.4.4.1 Appliance Signatures Library

During *appliance library creation*, a set of unique features $E(k,:)$, called *signatures*, is stored in the library. That is, when a new event is detected, it is checked against all signatures currently present in the library and if a match is found, the new signature is not included. Therefore, a library of unique signatures (class representatives) is formed. Comparison between different signatures is performed using DTW pattern matching explained which was explained in Section 3.1. Note that the library of signatures is obtained using historical unlabelled aggregate load measurements. Algorithm 3.4 summarises the appliance library creation phase of the proposed DTW-based method.

Algorithm 3.3 shows how DTW performs pattern matching between two vectors, based on the constraints and formulas introduced in Section 3.1.

Algorithm 3.3 DTW pattern matching: Find a distance between two vectors of possibly unequal lengths.

function DTW($\mathbf{p}, \mathbf{q}$)
  $n = length(\mathbf{p})$, $m = length(\mathbf{q})$
  for $i = 1, \ldots, n$ **do**
    for $j = 1, \ldots, m$ **do**
      $d(i,j) = |p(i) - q(j)|$
      $D(i,j) = d(i,j) + \min\{D(i-1,j), D(i-1,j-1), D(i,j-1)\}$
    end for
  end for
  Output $D(n,m)$
end function

Algorithm 3.4 returns **DTWLIB** as a library of signatures, where each row in **DTWLIB** contains one unique signature. Let Q be the total number of rows (signatures) in **DTWLIB**. Using expert knowledge (daily diary entries and IAMs data), each library signature is mapped to one appliance. Let **tag** be a Q-length vector containing an appliance list associated to the library signatures.

Algorithm 3.4 Appliance Library Creation: Using the extracted features **E** create the library of appliance signatures.

function LIBRARYCREATION(**E**) ▷ Where **E** found in Algorithm 3.2
  DTWLIB = []
  for $k = 1 : R$ **do**
    $i = 0$
    for $l = 0 : |\mathbf{DTWLIB}|$ **do**
      if $DTW(E(k,:), DTWLIB(i,:)) \geq DTW_{Thr}$ **then**
        $i++$
      end if
    end for
    if $i == |\mathbf{DTWLIB}|$ **then**
      DTWLIB = **DTWLIB** $\cup E(k,:)$
    end if
  end for
  return **DTWLIB**
end function

DTW_{Thr} is a threshold used to remove similar signatures that come from the same appliance. It trades off performance and complexity: larger DTW_{Thr} would allow for more signatures (i.e., multiple signatures for the same appliance due to variations in operation) to be kept. For this purpose, this threshold is set $DTW_{Thr} < C$.

3.4.4.2 Pattern matching of the Disaggregated Load

The proposed DTW-based method is performing pattern matching in order classify any operating appliances. That is, each identified event is compared to each signature in the library using Algorithm 3.3, and the best match is found in terms of minimising $D(n,m)$ distance measure over all library entries. Note that the output of the DTW classification is a soft value - $D(n,m)$ - the minimum distance between the testing event and the best matching signature from the library. The DTW classification via pattern matching is summarised in Algorithm 3.5, where the new event **q** is matched to the m-th signature in **DTWLIB**, which represents the signature with the minimum distance from the new event **q**. Therefore, the new event **q** is classified as appliance $tag(m)$.

During this phase, the collected measurements $p(TestStart : TestEnd)$ are first input to the Edge Detection function which returns detected events that are then passed to Feature Extraction. For each extracted event, the proposed DTW-based method compares the extracted events one by one with all library signatures available in **DTWLIB**.

Algorithm 3.5 DTW classification via pattern matching: Compare the new event q with all signatures in the library.

function DTWCLASS(T,**DTWLIB**)
  $D = DTW(\mathbf{q}, \mathbf{DTWLIB}(\mathbf{1}, :))$, $m = 1$
  for $i = 2 : Q$ **do**
    if $DTW(\mathbf{q}, \mathbf{DTWLIB}(\mathbf{i}, :)) \leq \mathbf{D}$ **then**
      $m = i$
    end if
  end for
  return m
end function

Disaggregation is performed iteratively, where in each iteration the threshold W is changed starting from the highest value detected. This way, in each iteration one appliance, starting from the highest consumers, are detected and removed. Set of appliances is updated, and W is recalculated using (3.10). Decreasing W in each iteration facilitates detecting small loads after high loads are detected and removed. Note that W is detected from the aggregate load in every iteration and appliances with similar power consumption in the aggregate load will be detected and therefore removed in the same iteration.

Suppose that set $\mathcal{M} = \{1, \ldots, M\}$ is ordered such that $p^1 \geq p^2 \geq \cdots \geq p^M$. DTW-based algorithm performs pattern of any aggregate load measurements using Algorithm 3.6, where the output is stored in a vector **classified** that contains all detected appliances. The threshold is adaptably changed based on the remaining appliances in the test dataset using (3.10), as described above.

Algorithm 3.6 Test: Perform pattern matching on the new collected aggregate active power measurements **p**.

function TEST(p(TestStart:TestEnd),**DTWLIB**, $\mathcal{M}$))
  Mk = []
  for $i = 1 : M$ **do**
    Mk = [$\mathbf{Mk}_i$]
    Calculate $W(i)$ with (3.10) using **Mk**
    $\mathbf{Event_{Test}}$=EventDetection($p(TestStart : TestEnd)$,$W(i)$)
    $\mathbf{E_{Test}}$=FeatureExtraction($Event_{Test}$, $p(TestStart : TestEnd)$)
    for $k = 1 : NumberOfRowsIn\mathbf{E_{Test}}$ **do**
      $D = tag$(DTWClass($E_{Test}(k, :)$, **DTWLIB**))
      $classified(k) = D$
    end for
    Output **classified** and remove events of Appliance i from testing dataset $p(TestStart : TestEnd)$
  end for
end function

A correction step is applied for separating appliances that might have not been distinguished correctly through the DTW pattern matching. This step performs a check on the feature values at the instant before and after the detected event window. The distance between these values and the relevant start and end values of the window are compared and the minimum distance is compared to the thresholds defined for each used appliance. The final corrected identified appliances, together with the associated time-stamps of the start and end of each event are the final output of the algorithm.

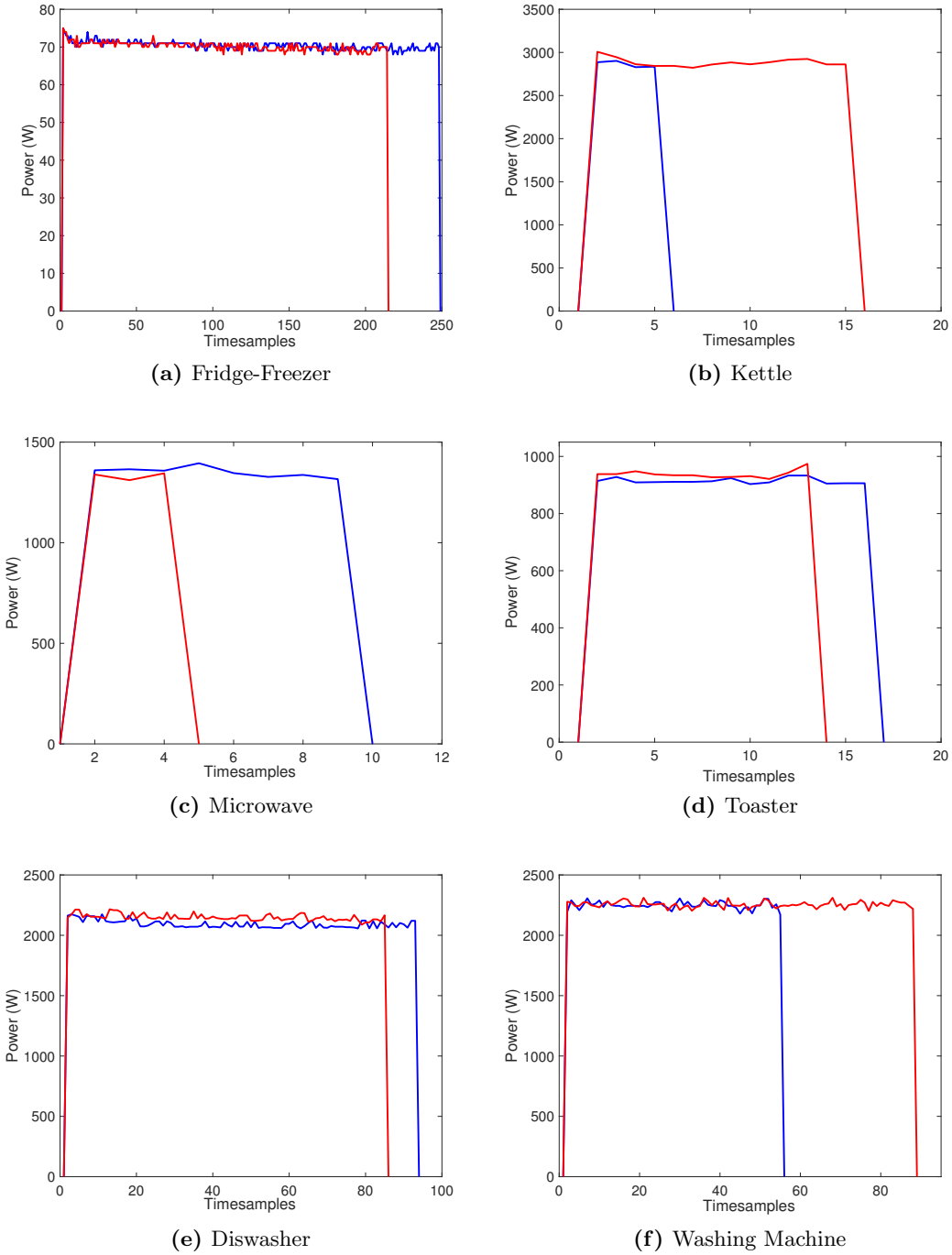
The proposed DTW-based method is an *unsupervised* NILM approach, as it performs pattern matching using unlabelled aggregate load measurements. The database of appliance signatures, known as *library of appliance signatures* is populated using unlabelled historical aggregate load measurements and post labelling is performed using the data from the IAMs. The proposed NILM method performs pattern matching using DTW against the database of signatures via a greedy search (full database).

Figure 3.2 shows an example of DTW pattern matching for the main appliances disaggregated during the implementation of the proposed method using house 2 from REFIT dataset [2, 3]. The blue signatures represent signatures available in the **DTWLIB** after the initial creation of the appliance library and the red signatures represent the classified events for each appliance. Note that dishwasher and washing machine have very similar signatures, which may result in misclassification between the two appliances, which will be discussed further in Section 4.4.1, where the disaggregation case study for the specific house is presented.

3.4.5 Performance Evaluation

Using the output vector **classified**, that contains all detected appliances, which is obtained using Algorithm 3.6, the accuracy of the proposed DTW-based method can be evaluated using both *classification accuracy* as defined in 2.8.1 and *estimation accuracy* as defined in 2.8.2.

For *classification accuracy*, we formally define TP, FN, and FP in the following way. Let p_t^i and $\hat{p}_t^i$ be the ground-truth value of the active power for Appliance i at time instance t and the disaggregation result, respectively, where the ground-truth is obtained either from the activity log and/or from the IAMs. Then, for each identified event starting at time instance t_s and ending at time instance t_e:



(a) Fridge-Freezer

(b) Kettle

(c) Microwave

(d) Toaster

(e) Diswasher

(f) Washing Machine

Figure 3.2: Classification using the proposed method of the main appliances disaggregated in House 2 from REFIT dataset [2, 3], where blue are the signatures available in the **DTWlib** and red the classified appliances.

- If for all $t \in [t_s, \ldots, t_e]$, $p_t = \hat{p}_t = 0$, then the disaggregation is *true negative* (TN).

- If there exists $t \in [t_s, \ldots, t_e]$, such that $p_t = 0$ and $\hat{p}_t > 0$, then the disaggregation is *false positive* (FP).

- If there exists $t \in [t_s, \ldots, t_e]$, such that $p_t > 0$ and $\hat{p}_t = 0$, then the disaggregation is *false negative* (FN).

- If there exists $t \in [t_s, \ldots, t_e]$, such that $p_t > 0$ and $\hat{p}_t > 0$, and for all $t \in [t_s - 1, \ldots, t_e + 1]$, $|p_t - \hat{p}_t| \leq C$, then the disaggregation is *true positive* (TP).

Note that as the DTW-based method is event based, the algorithm allocates one TN, FN, FP or TP with respect to the whole event duration $[t_s, \ldots, t_e]$, thus the algorithm was allowed to search both $t_s \pm 1$ and $t_e \pm 1$, in order to address any shifting between the aggregate and IAMs data due to synchronisation issues. If for any of these instances $p_t > 0$ or $\hat{p}_t > 0$ and for the rest of the time instances both $p_t > 0$ and $\hat{p}_t > 0$ and $|p_t - \hat{p}_t| \leq C$, the total event is classified as TP.

Therefore, *Precision*, *Recall* and *F-measure* can be calculated using equations 2.2, 2.3 and 2.4, respectively. Furthermore, for *estimation accuracy*, it has been decided to use the estimation accuracy Acc_i per appliance i and *TER* per appliance, as discussed in Chapter 2, by using the formulas 2.9 and 2.10. The selection of the specific metrics will facilitate to the better comparison of the proposed DTW-based NILM method with the methods presented in [4, 5, 33] for benchmarking purposes.

3.5 Summary

The aim of this chapter was to present a novel unsupervised NILM classification method using *Dynamic Time Warping*, that can potentially provide an efficient NILM solution. DTW algorithm has been defined using the relevant formulas and constraints. While DTW originates from speech recognition, its use has been expanded to data mining and information retrieval. A short review of successful applications of the algorithm

in different research fields has been included, in order to show the diverse use of DTW algorithm.

Furthermore, a formulation of the NILM problem regarding appliance disaggregation has been defined, together with the motivation behind the use of the DTW algorithm and its suitability for the NILM problem. More recently, some researchers have considered the use of DTW either as a standalone algorithm or combined with other methods for load disaggregation. These research works have been discussed in more detail, in order to outline the differences and thus the novelty of the proposed DTW-based method.

The proposed method follows the steps as defined in [24], namely pre-processing, event detection, feature extraction and classification. The implementation of each step has been described in detail using the relevant mathematical formulation and algorithmic representation, that allows the reader to better understand the proposed NILM method.

Chapter 4 will report the disaggregation performance of the proposed method for three US houses from REDD dataset [1] and two UK houses from REFIT dataset [2, 3]. This performance will be evaluated using a variety of classification and estimation accuracy metrics with benchmarks the performance of four NILM approaches: DT [32], supervised GSP, [33], unsupervised GSP [4] and their variations and FHMM available in NILMTK toolbox [34], as presented in [5].

Chapter 4

Analysis of proposed DTW-based Methodology

4.1 Introduction

In order to provide a rigorous analysis of the proposed DTW-based NILM methodology of Chapter 3, two public and commonly used electrical measurements datasets in the NILM literature are used for understanding the performance for a range of appliances and households with different appliance patterns of use. Metrics, described in Chapter 2 are used for evaluation and discussed in relation to understanding appliance detection and disaggregation.

For benchmarking purposes, the performance of the proposed DTW-based method will be compared to a variety of state-of-the-art NILM methods, such as HMM algorithm [35], DT [32], supervised GSP [33], unsupervised GSP [4] and FHMM available in NILMTK toolbox [34], and their variations, as found in [5, 33].

4.2 Data Pre-Processing and Experimental Setup

This section provides the relevant details regarding the implementation of the proposed DTW-based method with respect to the pre-processing of the datasets and the experimental setup.

4.2.1 REDD Data Pre-Processing Method

The sampling rates, as already discussed in Section 2.2.1, can vary from Hz to MHz and although they provide higher granularity, and more features that can be extracted (see Section 2.5), occasionally *downsampling* can be beneficial as it can reduce the size/length of the extracted features.

Batra *et al.* [34] have included a downsample preprocessor in NILMTK, which allows downsampling of any dataset to a specified frequency using aggregation function, such as mean, mode and median. Furthermore, downsampling has been performed in various NILM works, see [4–6, 32, 33, 35, 117, 118, 146, 160], where the authors downsampled REDD dataset to $1min$ resolution and [58] to $10sec$. Similarly with these works, it has been decided to downsample REDD dataset to $1min$ for the implementation of all the proposed methods.

This implementation can provide an understanding of how the proposed algorithm performs in lower resolution, as many appliances may operate for small durations, such as microwave, which can be set to work only for couple of minutes or so, depending on the resident's needs. Consequently, for the case of $1min$ resolution data, there will be on/off events that may occur for only one sample, which the proposed algorithm can separate and use them as appliance signatures, thus disaggregate and classify them during testing.

Furthermore, all Not-a-Number (NaN) values have been removed and forward filled. Any gaps of less than $2min$ have been forward filled with previous values or filled with zeros, for any gaps larger than $2min$, as proposed in [3]. This technique has been used previously for pre-processing UK-DALE in [46], where the authors identified the $2min$ threshold, by assuming the gaps longer for more than $2min$ are the result of the appliance (and the monitor) being switched off from the mains.

4.2.2 REFIT Data Pre-Processing Method

Throughout this thesis the REFIT dataset used for implementation was the cleaned version (labelled as CLEAN REFIT), which have been pre-processed by the authors in [3]. Murray *et al.* [3] have implemented a corrective step in order to adapt time/date for the case of British Summer Time (BST). Any values greater than $4,000W$ present in the IAMs data-streams have been removed and replaced with zeros, as these values are above the maximum possible power draw. Furthermore, if an appliance had been switched between plugs, the data streams have been moved, in order to provide continuous record for each appliance. Any NaN values have been removed and forward filled by the authors in [3], using the method already described in Section 4.2.1.

4.2.3 Dataset Description

From the available houses in REDD dataset, the houses selected for disaggregation were House 1, 2 and 6, as they have been used in various NILM research woks, and more specifically in [4, 5, 32, 33, 33–35], which will be used as benchmarks for the evaluation of the proposed algorithm. Table 4.1 provides the details of the time periods from each dataset, that have been selected for creating the library of appliance signatures and for disaggregation, including start and end data, and duration (number of days) of the dataset used.

Similarly, it has been decided to use House 2 and 17 from the REFIT dataset for the time period of *01/10/2014 - 31/10/2014*. This selection was based on the decision to evaluate the performance of the proposed DTW-based method towards the works presented in [4, 5, 33, 33], where the authors have implemented their disaggregation approach using the specific houses for the same time period. For both houses, the time periods used for creating the library of appliance signatures and testing, can be found in more details in Table 4.1.

Table 4.1: REDD and REFIT Datasets Time Periods used for creation of the Appliance Library and Disaggregation.

Houses	*Library Start Date*	*Library End Date*	*Duration*	*Testing Start Date*	*Testing End Date*	*Duration*
House 1 REDD	18/04/2011	24/04/2011	*7 days*	25/04/2011	24/05/2011	*16 days*
House 2 REDD	18/04/2011	21/04/2011	*4 days*	22/04/2011	22/05/2011	*12 days*
House 6 REDD	22/05/2011	27/05/2011	*7 days*	28/05/2011	14/06/2011	*11 days*
House 2 REFIT	01/10/2014	06/10/2014	*6 days*	07/10/2014	31/10/2014	*25 days*
House 17 REFIT	01/10/2014	06/10/2014	*6 days*	07/10/2014	31/10/2014	*25 days*

4.3 Down-sampled REDD Dataset Disaggregation

This section presents the disaggregation cases studies using the proposed DTW-based algorithm for houses 1, 2 and 6 from the downsampled REDD dataset.

4.3.1 REDD House 1 Disaggregation Case Study

According to the relevant information obtained from REDD dataset [1], the appliances/devices monitored in house 1 were: dishwasher, bathroom GFI, electric heat, fridge, various kitchen outlets, lighting, microwave, oven and washer dryer. Note that the proposed algorithm was not able to detect the events for bathroom GFI, electric heat, lighting and oven, although all the appliances were present in the aggregate load. For this reason, no appliance signatures have been extracted as features, and the proposed method was not able to disaggregate and classify these appliances.

The obtained classification results, in the form of TP, FP, FN, PR, RE and F_M, for house 1 are presented in Table 4.2.

Table 4.2: Classification Accuracy for REDD House 1.

Apps	*TP*	*FP*	*FN*	*PR*	*RE*	F_M
Dishwasher	6	0	26	1	0.19	0.32
Fridge	159	24	53	0.87	0.75	0.80
Kitchen Outlet	65	9	11	0.88	0.85	0.86
Microwave	61	13	21	0.82	0.74	0.78
Washer Dryer	46	2	22	0.96	0.67	0.79

Note: Apps=Appliances. TP=True Positive, FP=False Positive, FN= False Negative, PR=Precision, RE=Recall, F_M=F-Measure.

The proposed method was able to obtain high precision $> 80\%$ for all the disaggregated appliances, which means that the classification did not suffer from high numbers of false positives that an appliance was operating. On the other hand, dishwasher and washer dryer have shown an increased number of not identified appliance operations FN, which resulted to 19% and 67% of recall, respectively, but the algorithm was still able to obtain $> 70\%$ recall for the rest of the appliances. For both dishwasher and washer

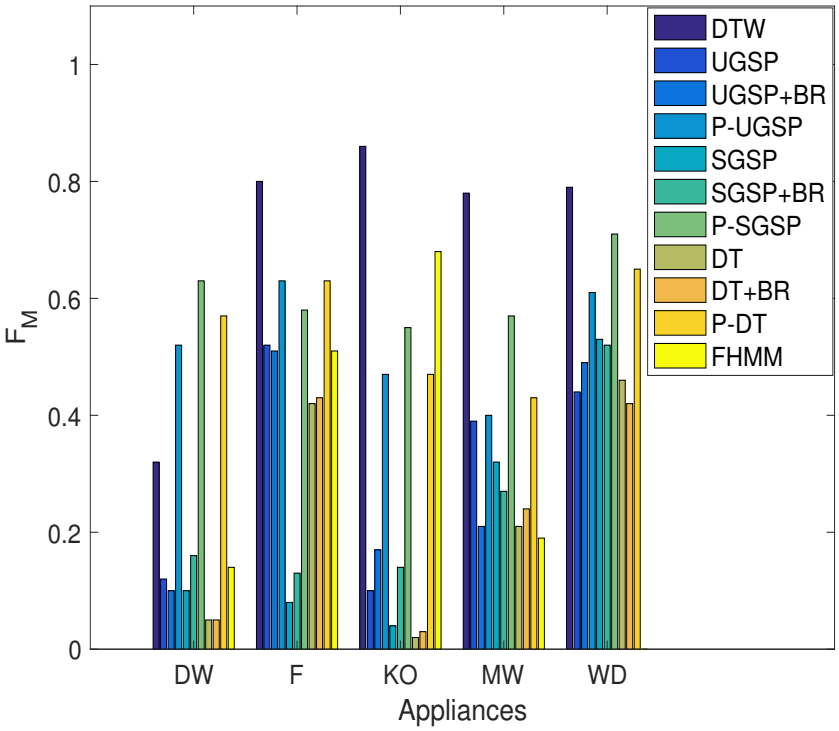


Figure 4.1: Performance evaluation using F_M for the proposed DTW-based NILM method (Table 4.3) with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REDD House 1, as presented in [5].

dryer, the DTW-based method can disaggregate the washing state of each appliance, therefore the rest of the states were evaluated as FN, thus not identified. In terms of F_M, the classification accuracy was on average $> 78\%$ for most of the disaggregated appliances, with the exemption of dishwasher, were the algorithm performed poorly at 32%.

In order to further analyse the obtained results, Tables 4.3-4.4 present the obtained classification accuracy F_M and estimation accuracy Acc_i per appliance in comparison to the algorithms presented in [5]. In addition to the tables, a graphical representation of the obtained results can be found in Figures 4.1-4.2.

Although the proposed algorithm, as already discussed, obtained low classification accuracy for the dishwasher, compared to the rest of the disaggregated appliances, it was able to outperform the majority of the methods presented in [5] by at least 15%. All the methods using the pre-processing and refinement method P [5] were able to surpass the DTW-based approach, with best F_M reported by the P-SGSP method (0.63) [5, 33].

Table 4.3: Classification Accuracy using F_M for the proposed DTW-based NILM method, with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REDD House 1, as presented in [5].

		UGSP[4]			*SGSP*[33]			*DT*[32]			
Apps	*DTW*	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.32	0.12	0.10	0.52	0.10	0.16	**0.63**	0.05	0.05	0.57	0.14
F	**0.80**	0.52	0.51	0.63	0.08	0.13	0.58	0.42	0.43	0.63	0.51
KO	**0.86**	0.10	0.17	0.47	0.04	0.14	0.55	0.02	0.03	0.47	0.68
MW	**0.78**	0.39	0.21	0.40	0.32	0.27	0.57	0.21	0.24	0.43	0.19
WD	**0.79**	0.44	0.49	0.61	0.53	0.52	0.71	0.46	0.42	0.65	0

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, WD=Washer Dryer. **Bold** values represent the method/methods with the highest obtained classification accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Table 4.4: Estimation Accuracy using Acc_i for the proposed DTW-based NILM method, with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REDD House 1, as presented in [5].

		UGSP[4]			*SGSP*[33]			*DT*[32]			
Apps	*DTW*	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.64	0.23	0.21	0.66	0.16	0.24	**0.72**	0.55	0.51	0.58	0.21
F	0.80	0.68	0.65	0.91	0.48	0.51	**0.93**	0.47	0.47	0.88	0.59
KO	0.74	0.16	0.25	0.83	0.09	0.11	0.84	0.24	0.27	0.80	**0.92**
MW	0.63	0.40	0.26	0.66	0.37	0.46	**0.69**	0.58	0.61	0.62	0.40
WD	0.52	0.58	0.61	0.95	0.64	0.61	**0.96**	0.76	0.73	0.89	−1.99

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, WD=Washer Dryer. **Bold** values represent the method/methods with the highest obtained estimation accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Furthermore, the proposed algorithm was able to obtain the best classification performance for the fridge with F_M of 0.8, outperforming the algorithms in [5] by as high as 72% and as low as 17%. Similarly, the proposed method has shown the best classification accuracy for both kitchen outlet, microwave and washer dryer with 0.86, 0.78 and 0.79 respectively. For the case of the kitchen outlet, the proposed method was able to obtain better classification accuracy F_M ranging from $18-84\%$, compared to the methods presented in [5]. For both microwave and washer dryer, the highest classification accuracy in [5] was shown for P-SGSP [5, 33] with 0.57 and 0.71 respectively, which is less than 21% and 8% compared to the accuracy acquired using the proposed method.

In summary, the proposed method was able to classify more accurately the majority of the disaggregated appliances, with the exception of the dishwasher, compared to the methods presented in [5]. Although classification accuracy is an important measure for assessing the performance of a NILM method, it is not enough for claiming that a NILM method has successfully disaggregated the appliance consumption from the aggregate load. Therefore, the proposed method is going to be evaluated using the estimation accuracy Acc_i per appliance, in order to identify how accurate is the estimation of the power consumption compared to the actual consumed power, which is available through individual monitors.

In terms of estimation accuracy, the best performing method, according to [5], was the P-SGSP [5, 33], with the exception of the kitchen outlet, where FHMM [34] obtained the highest accuracy of 0.92, which is 18% better than the performance of the proposed method. For the case of the dishwasher, the proposed method was able to outperform the majority of the presented NILM methods by $6-48\%$, but P-UGSP [4, 5] and P-SGSP [5, 33] obtained better estimation accuracy by 2% and 8% respectively. Furthermore, the NILM approaches using the P method from [5] were able to estimate more accurately the power consumed by the fridge by $8-13\%$, whereas the proposed method obtained better estimation by $12-33\%$ compared to the rest of the methods.

The algorithms using the P method from [5] were able to outperform the proposed method for the kitchen outlet by $6-10\%$, but the DTW-based method performed better by $47-65\%$ compared to the rest of the methods. For the case of the microwave, the proposed method obtained higher estimation accuracy (0.62) than the majority of the

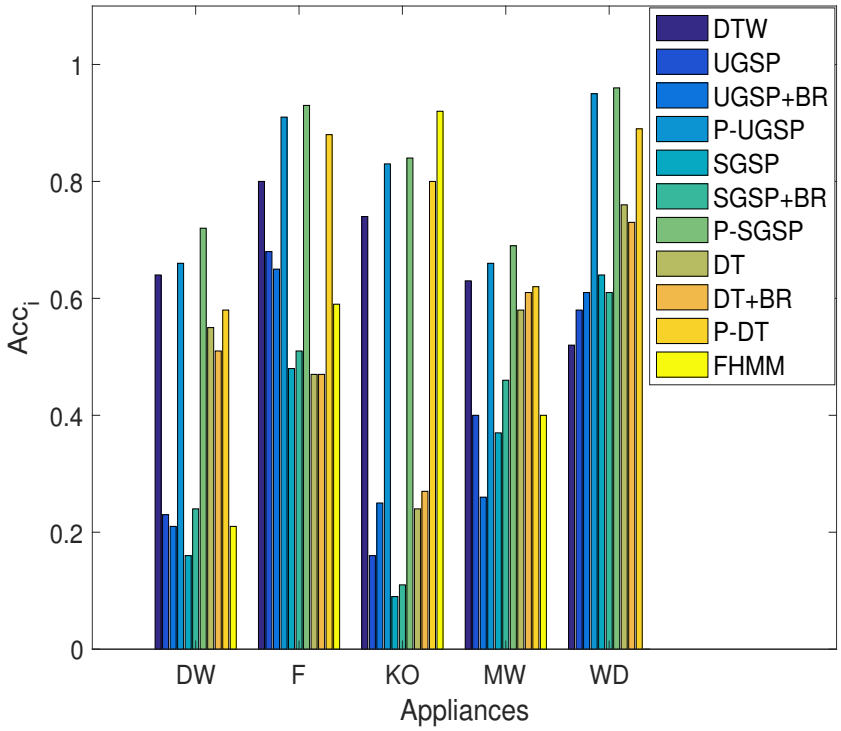


Figure 4.2: Performance evaluation using Acc_i for the proposed DTW-based NILM method (Table 4.4) with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REDD House 1, as presented in [5].

methods presented in [5] by $1 - 37\%$, and its performance was comparable to the best performing methods, using the P method.

Furthermore, the proposed method obtained the poorest estimation accuracy for the washer dryer with an Acc_i of 0.52 and was only able to perform better than FHMM [34], which obtained a negative estimation accuracy of 1.99. The rest of the methods presented in [5] were able to outperform the DTW-based method by as low as 6% for UGSP [5] and as high as 44% for P-SGSP [5, 33]. One of the reasons that could have potentially caused this performance is the fact the proposed algorithm seems to overestimate the consumption of the washer dryer.

In addition to classification and estimation accuracy, it has been decided to evaluate the performance of the proposed algorithm using the normalised total power consumption estimation error (TER), which has been proposed in [5]. According to the results presented in Table 4.5, it is apparent the P-UGSP [4, 5] had obtained the best performance for most appliances in terms of the TER error, as it was able to

show on average an error of ~ 0.1, with the highest identified error for the dishwasher and the smallest for the microwave. The proposed method was able to outperform both methods for the washer dryer, where the proposed method was able to obtain error of 0.07. Furthermore, the proposed method was able to outperform UGSP [4] for microwave, but obtained higher error for the rest of the appliances, with the highest at 0.72 for dishwasher.

Table 4.5: Normalised total power consumption estimation error (TER) for the proposed DTW-based NILM method, with benchmarks UGSP [4] and P-UGSP [5] for REDD House 1, as presented in [5].

Apps	*DW*	*F*	*KO*	*MW*	*WD*
DTW	0.72	0.39	0.50	0.42	**0.07**
UGSP	0.53	0.26	0.41	0.53	0.29
P-UGSP	**0.22**	**0.10**	**0.15**	**0.03**	0.11

Note: Apps=Appliances, B=Bathroom GFI, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, L=Light, MW=Microwave, O=Oven, WM=Washing Machine. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Figure 4.3 present the disaggregation results using the proposed DTW-based NILM method as percentage of load contribution per appliance with regards to the aggregate load and Figure 4.4, the relevant ground truth consumption per appliances, using the data from the individual monitors available in the REDD dataset [1]. The load contribution to the aggregate load of any appliances, that have not been disaggregated using the proposed DTW-based method, has been included as "*unknown*" in both Figure 4.3 and 4.4. Note that the "*unknown*" includes also the base load and measurement noise, similarly to the noise term in equation 3.4.

It is apparent that for the washer dryer, the proposed method overestimated the consumption, especially when taking into account the fact that the algorithm was only able to disaggregate the washing state of the appliance. Furthermore, the proposed method was only able to disaggregated 8% of the fridge consumption, whereas the actual consumption was 14%, and underestimated the dishwasher consumption by 5%, as similarly with the washer dryer, the DTW-based method was only able to obtain the washing cycle of the appliance.

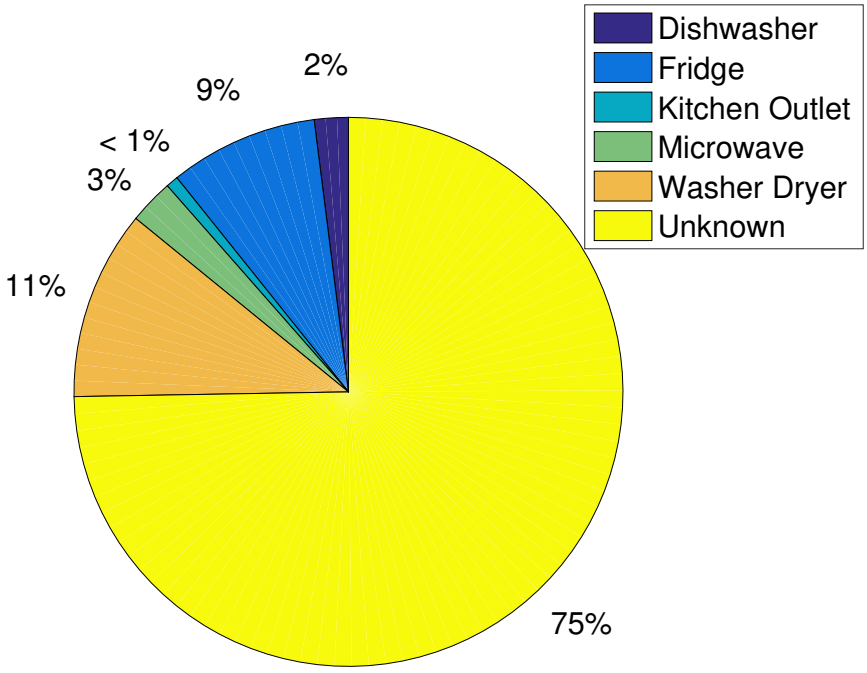


Figure 4.3: Disaggregation load contribution using proposed DTW-based NILM method, as percentage of load contribution per appliance relative to the aggregate load, for House 1 from REDD dataset.

For both kitchen outlet and microwave, the obtained consumption results were closer to the ground truth consumption with 2% instead of 4% and $< 1\%$ instead of 1%, respectively. Finally, the load classified as unknown for the disaggregated results was 75%, more than 10% from the actual unknown load, which is mainly the result of non-identified appliance operations and standby states.

In general, it has been apparent from the obtained results that the proposed method was able to obtain better classification accuracy in the context of F_M for the majority of the appliances, with the exception of dishwasher, compared to the methods presented in [5]. Furthermore, the estimation accuracy of the proposed method was more than 0.6 for all appliances, apart from the washer dryer, where it had the lowest accuracy only surpassing FHMM [34]. The DTW method was mainly outperformed by the methods using the pre-processing and refinement method P [5] and for the case of the kitchen outlet, FHMM, which obtained the highest accuracy.

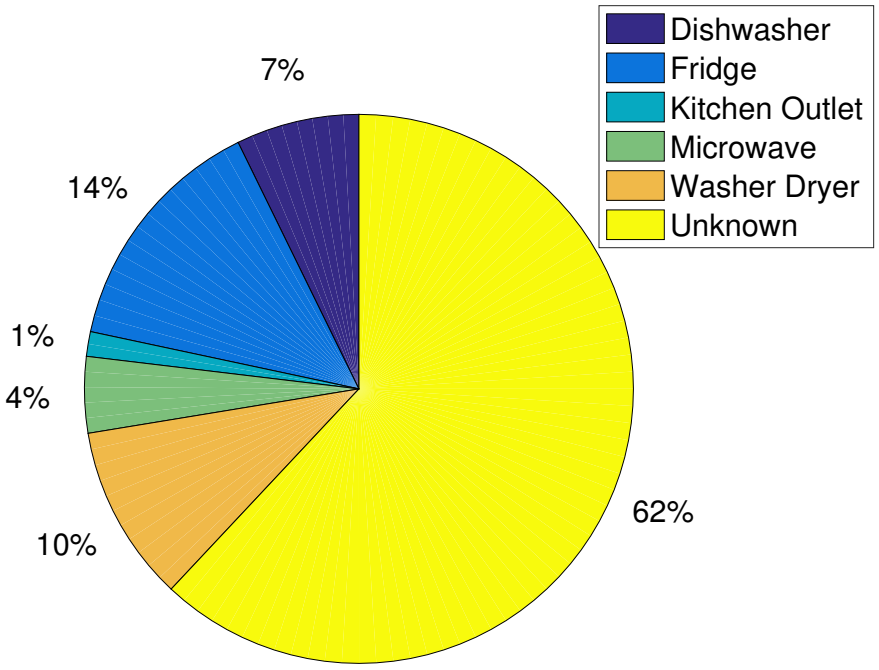


Figure 4.4: Ground truth load contribution, as percentage of load contribution per appliance relative to the aggregate load, for House 1 from REDD dataset.

4.3.2 REDD House 2 Disaggregation Case Study

For the case of house 2 from REDD dataset [1], the monitored appliances/devices apart from the mains, are: dishwasher, fridge, two kitchen outlets, lighting, microwave, stove and washer dryer. The proposed algorithm was not able to extract appliance signatures for the washer dryer and lighting, and therefor disaggregate the appliances from the aggregate load. Therefore the presented results include only the appliances disaggregated using the proposed DTW-based method for both classification and estimation accuracy, with benchmarks the various methods presented in [5]. TER error was not used for performance evaluation, as the authors in [5] have not presented the relevant results for the specific house.

Table 4.6 presents the classification accuracy obtained using confusion matrix TP, FP and FN, together with the resulting PR, RE and F_M.

Similarly with the results presented for house 1, the proposed algorithm was able to obtain high precision for the majority of the disaggregated appliances on average

Table 4.6: Classification Accuracy for REDD House 2.

Apps	*TP*	*FP*	*FN*	*PR*	*RE*	F_M
Dishwasher	2	0	7	1	0.22	0.36
Fridge	213	58	36	0.79	0.85	0.82
Kitchen Outlet 1	13	0	5	1	0.72	0.84
Kitchen Outlet 2	56	9	20	0.86	0.74	0.79
Microwave	31	0	7	1	0.82	0.90
Stove	10	18	0	0.36	1	0.53

Note: Apps=Appliances. TP=True Positive, FP=False Positive, FN= False Negative, PR=Precision, RE=Recall, F_M=F-Measure.

$> 80\%$, with the exception of the stove, where the algorithm wrongly identified as stove an increased number of events resulting to a precision of 36%. In terms of recall, the proposed method was able to obtain on average more than 70% for most appliances, except for dishwasher, where the reported recall was 33%. As already discussed, the algorithm can only disaggregate successfully the high consuming state of a multi state appliance such as dishwasher, therefore the reported non-identified operations represent the other states. When evaluating the performance using F_M, the proposed method was able to report on average $> 80\%$ for most appliances. Both dishwasher and stove obtained lower F_M of $<\sim 50\%$, which was caused by the increased number of missing operations and wrongly positively identified operations for each appliance respectively.

Tables 4.7 and 4.8 present the obtained classification and estimation accuracy per appliance using F_M and Acc_i, with respect to the reported accuracies, as presented in [5]. Figures 4.5 and 4.6 show a graphical representation of results.

The proposed method was able to outperform the majority of algorithms presented in [5] for most disaggregated appliances, with the exception of dishwasher and kitchen outlet 1. For dishwasher, the proposed method was outperformed by the approaches using the P method by more than 40%, but was able surpass the rest of the methods on average by $> 15\%$, with F_M of 0.36. The best performing method with F_M of 0.82 was the P-UGSP [5].

Table 4.7: Classification Accuracy using F_M for the proposed DTW-based NILM method, with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REDD House 2, as presented in [5].

		UGSP[4]			*SGSP*[33]			*DT*[32]			
Apps	*DTW*	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.36	0	0.21	**0.82**	0.12	0.21	0.8	0.08	0.16	0.77	0
F	**0.82**	0.12	0.11	0.24	0.17	0.16	0.31	0.30	0.28	0.29	**0.82**
KO1	0.84	0.07	0.13	0.82	0.01	0.09	**0.90**	0.18	0.23	0.85	0.52
KO2	**0.79**	0.33	0.37	0.60	0.45	0.51	0.53	0.24	0.33	0.43	0.38
MW	**0.90**	0.31	0.34	0.69	0	0.01	0.83	0	0.05	0.78	0.24
S	**0.53**	0	0	0.26	0	0	0.23	0.01	0.01	0.33	0.01

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, S=Stove. **Bold** values represent the method/methods with the highest obtained classification accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Table 4.8: Estimation Accuracy using Acc_i for the proposed DTW-based NILM method, with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REDD House 2, as presented in [5].

		UGSP[4]			*SGSP*[33]			*DT*[32]			
Apps	*DTW*	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.61	0.50	0.52	**0.78**	0.32	0.33	0.77	0.15	0.13	0.64	−2.13
F	0.69	0.04	0.01	0.55	0.07	0.07	0.64	0.09	0.07	**0.72**	0.69
KO1	0.41	0.17	0.24	**0.89**	0.09	0.15	0.86	0.32	0.34	0.86	−0.40
KO2	**0.81**	0.58	0.06	0.72	0.55	0.49	0.73	0.59	0.54	0.66	0.12
MW	0.64	0.54	0.56	0.82	0.40	0.21	**0.85**	0.65	0.63	0.76	0.11
S	**0.75**	0.02	0	0.45	0.02	0.04	0.44	0.24	0.19	0.43	−4.49

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, S=Stove. **Bold** values represent the method/methods with the highest obtained estimation accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

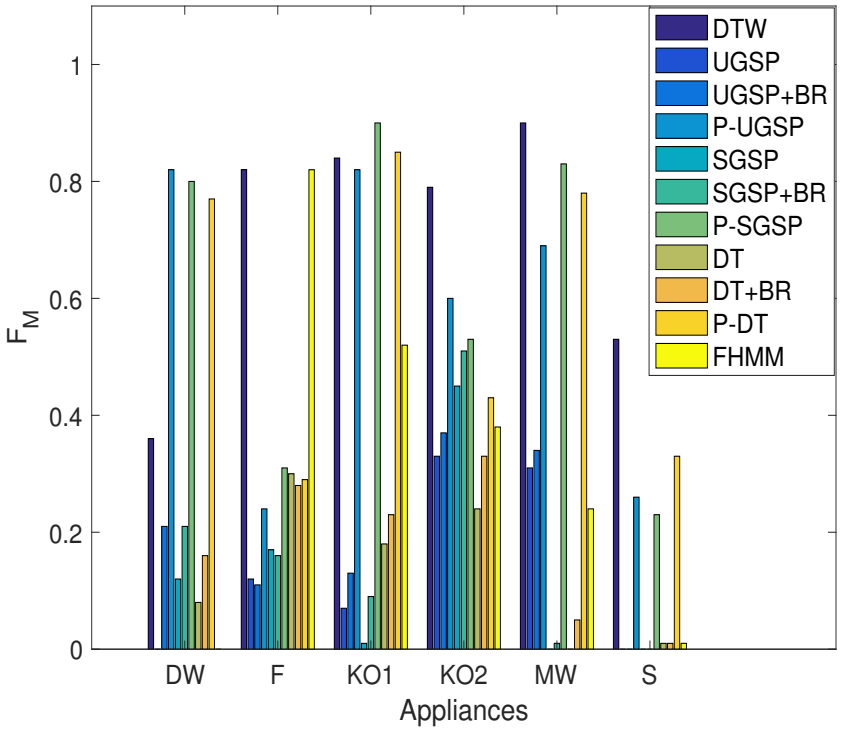


Figure 4.5: Performance evaluation using F_M for the proposed DTW-based NILM method (Table 4.7) with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REDD House 2, as presented in [5].

For the case of the fridge, the proposed method obtained the same result with FHMM, accuracy of 0.82, which is more than 50% better that the rest of the methods presented in [5]. P-SGSP obtained the highest classification accuracy for kitchen outlet 1 at 0.9, according to the authors in [5]. The proposed method was able to obtain an accuracy of 0.84, which is comparable to the accuracy reported by rest of the P refined method in [5], and outperformed the other methods.

Furthermore, the proposed method was able to outperform all methods from [5] for kitchen outlet 2, by at least 19%. Similarly, the obtained accuracy for microwave was 0.9, which is higher than the best performing P-SGSP as seen in [5], where the accuracy was 0.83. Finally, in terms of the stove, although the accuracy acquired using the proposed DTW method was only 0.53, this is still higher by 30%, compared to the reported accuracies in [5].

In terms of estimation accuracy per disaggregated appliance, as shown in Table 4.8 and Figure 4.6, the proposed method was able to outperform all methods presented in

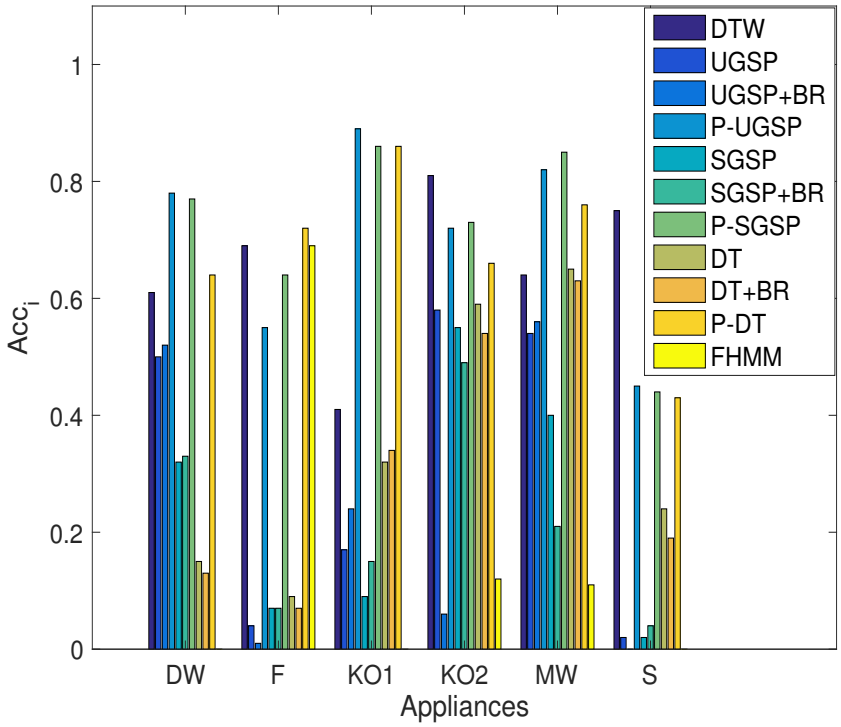


Figure 4.6: Performance evaluation using Acc_i for the proposed DTW-based NILM method (Table 4.8) with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REDD House 2, as presented in [5].

[5] for the case of kitchen outlet 2, by acquiring an estimation accuracy of 0.81 and for stove, where the obtained accuracy was 0.75, which is at least $\sim 30\%$ higher compared to the accuracies shown in [5].

Furthermore for dishwasher, the proposed method obtained accuracy of 0.61, which is higher than most of the methods in [5] with the exception of the P refined approaches. For the case of fridge, the proposed method reported the same estimation accuracy as FHMM, namely 0.69, and was only outperformed by P-DT, which obtained a comparable accuracy of 0.72. Kitchen outlet 2 reported the best estimation accuracy at 0.81 for the proposed method, which was higher than the accuracies presented in [5].

Finally, for microwave, the proposed method was able to obtain estimation accuracy of 0.64, which was comparable or better from the majority of the NILM methods, apart from the case fo the P refined methods, which reported the best performance with 0.85 for P-SGSP [5].

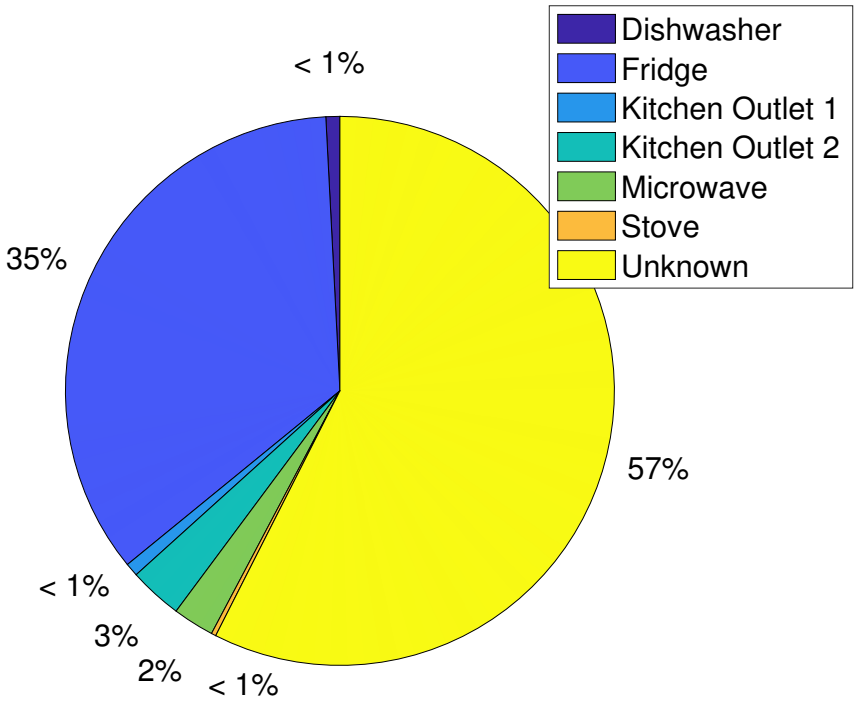


Figure 4.7: Disaggregation load contribution using proposed DTW-based NILM method, as percentage load contribution per appliance relative to the aggregate load, for House 2 from REDD dataset.

Similarly with the results presented for house 1, Figure 4.7 shows the disaggregation results of the proposed DTW-based NILM method as percentage of load contribution per appliance with regards to the aggregate load and Figure 4.8, the relevant ground truth consumption per appliances, using the data from the IAMs as presented in the REDD dataset [1]. The load contribution to the aggregate load of any appliances, that have not been disaggregated using the proposed DTW-based method, has been included as "*unknown*" in both Figure 4.7 and 4.8.

The dishwasher was underestimated by more than 2%, which is mainly caused by the inability of the proposed method to disaggregate other than the washing cycle of the dishwasher. Fridge reports disaggregated results of $\leq 3\%$ compared to the ground truth, as according to the results presented in Table 4.6, the proposed method was not able to recognise and therefore estimate $\sim 15\%$ of the the actual fridge operations.

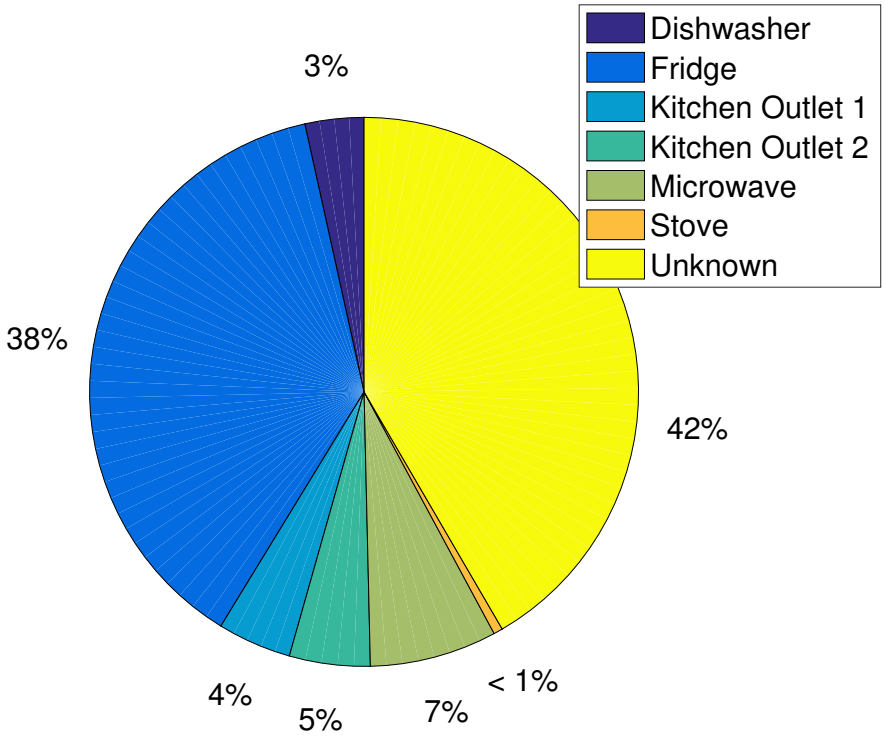


Figure 4.8: Ground truth load contribution, as percentage load contribution per appliance relative to the aggregate load, for House 2 from REDD dataset.

Furthermore, kitchen outlet 1 contribution according to the disaggregation obtained using the proposed method was less than 1%, compared to the actual consumption contribution of 4%, which is the result of the missing operations that account for almost 40% of the total operation of the kitchen outlet 1 during the disaggregation period. In the same context, kitchen outlet 2 have shown 3% contribution to the total aggregate load, whereas the actual contribution based on the individual monitors was 5%. This difference is the outcome of the not identified use of kitchen outlet 2 during classification, which was $\sim$ 25% of the actual appliance use. The contribution of stove was similar in both disaggregated and ground truth data as it only accounts for $<$ 1%. Finally, microwave reported the highest contribution difference of 5% compared to the actual contribution, which can be explained using the results presented in Table 4.6, where it is apparent that $\sim$ 20% of the microwave's operations were not classified during disaggregation. Therefore, the reported unknown load for the disaggregation results is 16% higher than the ground truth contribution of the unknown load.

In summary, the proposed method has shown better classification accuracy for most appliances compared to the methods in [5] and the same classification accuracy with FHMM for fridge. For the case of kitchen outlet 1, the obtained F_M was comparable to the relevant accuracy reported in [5] using the P refined methods. As with dishwasher, the proposed method was only outperformed by the NILM approaches in [5], when using the P refinement. In terms of estimation accuracy, the DTW-based method, was able on average to obtain better accuracy for most appliances, but was outperformed by the same algorithms when using the P method. The proposed method was proven more accurate from all methods available in [5] for estimating the use of kitchen outlet 2 and stove.

4.3.3 REDD House 6 Disaggregation Case Study

The published data for house 6, available in REDD dataset [1], include data for the mains, air conditioning (AC), bathroom GFI (B), dishwasher, electric heater, electronics, fridge, various kitchen outlets, lighting, stove and washer dryer. The performance of the proposed DTW-based method for house 6 will be evaluated using the results reported in both [4] and [33]. He *et al.* [33] have reported disaggregation of the AC, electronics, fridge, kitchen outlet and stove, whereas the authors in [5] were able to further classify electric heater and lighting. During the implementation of the proposed method for house 6 both dishwasher and washer dryer were ignored, as there is no operation of those appliances. Furthermore, the microwave operation reported in [4, 33] is found in the kitchen outlet 1 from the REDD dataset, therefore the reported accuracy in this thesis will be stated as KO1. The proposed algorithm was able to disaggregate the bathroom GFI, whereas was not able to obtain signatures and therefore classify electronics, electric heater, lighting and the reported kitchen outlet in [4, 33], which were all present in the aggregate load.

Table 4.9 presents the obtained classification accuracy, in the form of confusion matrix TP, FP, FN, precision, recall and F_M, and the estimation accuracy per appliance Acc_i. In terms of classification accuracy, the proposed algorithm, similarly with the results presented for house 1 and 2 from REDD dataset, was able to obtain high precision, more than 80% for most appliances, except for the bathroom GFI, where the reported precision was 50%.

Table 4.9: Classification and Estimation Accuracy per appliance for REDD House 6.

Apps	*TP*	*FP*	*FN*	*PR*	*RE*	F_M	Acc_i
Air Conditioning	27	0	4	1	0.87	0.93	0.92
Bathroom GFI	3	3	0	0.50	1	0.67	0.57
Fridge	127	19	48	0.87	0.79	0.73	0.83
Kitchen Outlet 1	2	0	0	1	1	1	0.54
Stove	5	1	0	0.83	1	0.91	0.75

Note: Apps=Appliances. TP=True Positive, FP=False Positive, FN= False Negative, PR=Precision, RE=Recall, F_M=F-Measure.

For recall, the proposed method acquired more than 70% for all appliances and achieved 100% for air conditioning, bathroom GFI, stove and kitchen outlet 1, which means that the proposed method did not suffer from increased number of missing appliance operations. The obtained F-measure was the lowest for bathroom GFI and fridge, which was at 67% and 79% respectively, while for the rest of the appliances, the proposed method obtained F_M higher than 90%. Furthermore, the DTW-based method obtained estimation accuracy of $\geq$ 75% for most appliances, with the exception of bathroom GFI and kitchen outlet 1, where the estimation accuracy was 57% and 54% respectively. Both appliances, according to the REDD dataset individual monitors, have continuous measurements of their standby state, which the proposed method is not able to estimate, but is still taken into account when calculating the estimation accuracy.

The performance of the proposed method was evaluated using F_M as a classification accuracy metric, with benchmarks GSP [144], GSP+FS [33],GSP+SA [33], SGSP [33], UGSP [4], DT [32], HMM [35], for REDD House 6, as presented in [4, 33]. The obtained results are presented numerically in Table 4.10 and graphically, for better understanding, in Figure 4.9. By reviewing the results, it is apparent that the performance of the proposed method was on average more than 0.8 for most appliances and comparable to the accuracies reported in [4, 33].

Table 4.10: Classification Accuracy using F_M for the proposed DTW-based NILM method, with benchmarks GSP [144], GSP+FS [33],GSP+SA [33], SGSP [33], UGSP [4], DT [32], HMM [35], for REDD House 6, as presented in [4, 33].

		GSP						
Apps	*DTW*	*GSP*	*GSP+FS*	*GSP+SA*	*SGSP*	*UGSP*	*DT*	*HMM*
AC	**0.93**	0.49	0.74	0.73	0.89	0.88	0.89	0.12
F	0.79	0.54	0.77	0.77	0.77	0.82	**0.99**	0.88
KO1	**1**	0.92	0.77	0.77	0.91	0.87	0	0
S	0.91	**1**	0.75	0.75	0.92	0.54	0.67	0

Note: Apps=Appliances, AC= Air Conditioning, F=Fridge, KO1=Kitchen Outlet 1, S=Stove. **Bold** values represent the method/methods with the highest obtained classification accuracy per appliance. The KO1, as found in REDD dataset [1], represents the microwave appliance presented in [4, 33].

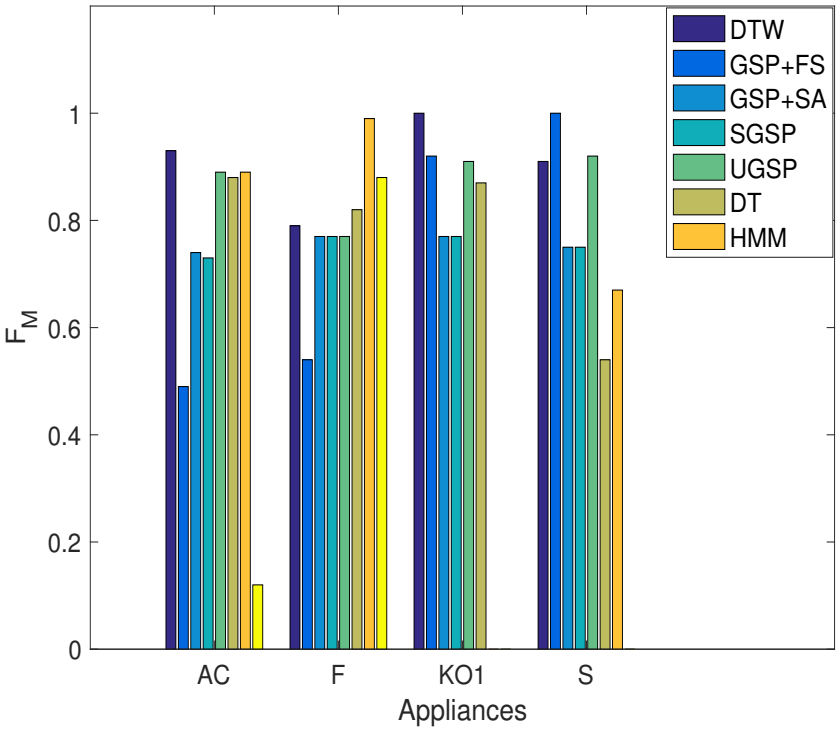


Figure 4.9: Performance evaluation using F_M for the proposed DTW-based NILM method, as presented in Table 4.10, with benchmarks GSP [144], GSP+FS [33],GSP+SA [33], SGSP [33], UGSP [4], DT [32], HMM [35], for REDD House 6, as presented in [4, 33].

More specifically, the proposed method was able to outperform all methods presented in [4, 33] for both air conditioning and kitchen outlet 1, where the reported accuracy of the proposed method was 0.93 and 1 respectively. For fridge, the best performance was reported for DT [32], with classification accuracy at 0.99, while HMM [35] achieved F_M of 0.88. The proposed algorithm was able to obtain 0.79, which is comparable or higher than the rest of the methods.

Similarly, regarding the disaggregation of the stove, the DTW-based method was able to achieve higher or comparable classification accuracy than most of the methods presented in [4, 33], with F_M of 0.91. The best performance was reported for GSP+FS [33], which obtained perfect classification with F_M equal to 1.

Figure 4.10 presents the disaggregation results of the proposed DTW-based NILM method as percentage of load contribution per appliance with regards to the aggregate load and Figure 4.11, the relevant ground truth consumption per appliances, using the

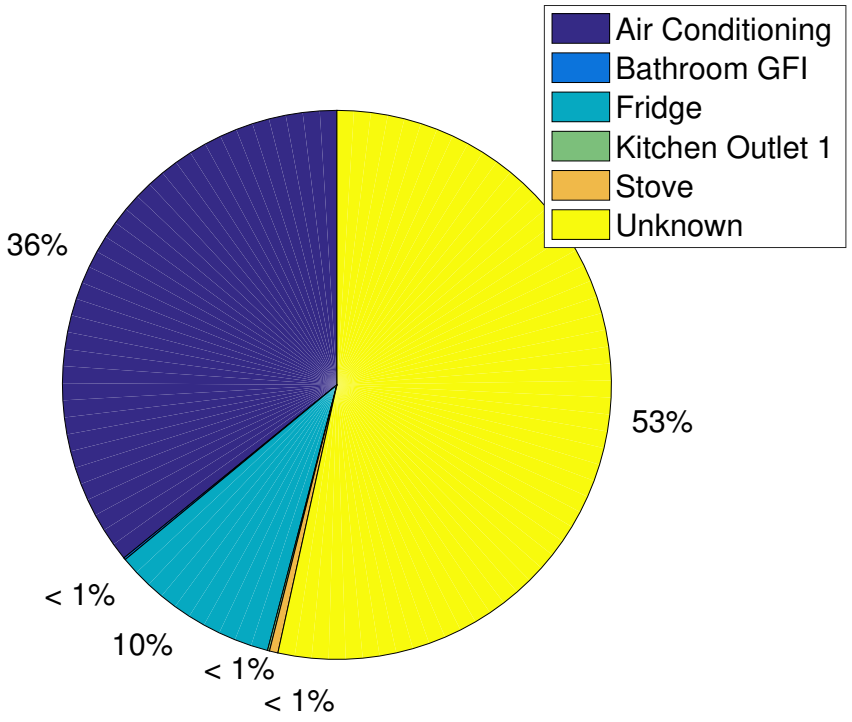


Figure 4.10: Disaggregation load contribution using the proposed DTW-based NILM method, as percentage load contribution per appliance relative to the aggregate load, for House 6 from REDD dataset.

data from the IAMs as presented in the REDD dataset [1]. The load contribution to the aggregate load of any appliances, that have not been disaggregated using the proposed DTW-based method, has been included as "*unknown*" in both Figure 4.10 and 4.11.

It is apparent, that the proposed method overestimated the consumption of the air conditioning by 3% compared to the actual AC consumption. As the base load is not removed from the mains aggregate data, it is possible that the reported power consumption of the AC and furthermore other appliances, may be estimated higher than the ground truth.

Although, both bathroom GFI, kitchen outlet 1 and stove contribute $\leq 1\%$ to the total disaggregation and ground truth consumption, Figure 4.10 shows that the actual contribution of these appliances, is slightly higher than the one estimated using the proposed method. Furthermore, fridge, for the proposed method, seems to contribute 10% to the total consumption, whereas the actual contribution according to the IAMs

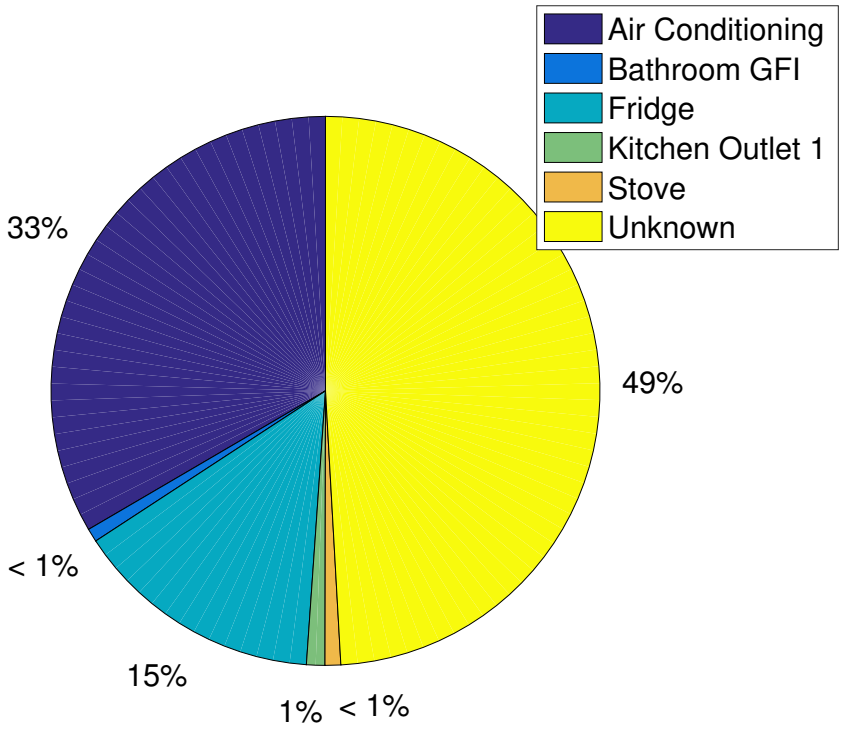


Figure 4.11: Ground truth load contribution, as percentage load contribution per appliance relative to the aggregate load, for House 6 from REDD dataset.

was 15%, which could be caused by the fridge operations that were not classified using the DTW-based method. Finally, the unknown load for the disaggregation results using DTW contributes 53% to the total load, which is only 4% higher than the ground truth.

In conclusion, the proposed method was able to obtain classification accuracy with respect to F-Measure of more than 0.9 for most appliances, with the exception of bathroom GFI and fridge, where the obtained accuracy was 0.67 and 0.79 respectively. Estimation accuracy was more than 0.75 in most cases, apart from kitchen outlet 1 and bathroom GFI, which obtained on average an accuracy of ~ 0.55. When evaluating the performance of the proposed method with benchmarks the methods presented in [4, 33], the DTW-based algorithm was able to achieve either comparable or better classification accuracies for most appliances, besides fridge, where HMM [35] and DT [32] reported higher accuracy by 10% and 30% respectively.

4.4 REFIT Dataset Disaggregation

This sections provides the performance analysis of the proposed DTW-based method using houses 2 and 17 from the REFIT dataset [2, 3].

4.4.1 REFIT House 2 Disaggregation Case Study

The focus of the disaggregation for the case of house 2 from the REFIT dataset [2, 3], similarly with the work presented in [5], was to attempt to disaggregate the appliances that are responsible for most of the power consumption present in the aggregate data, that were monitored using IAMs. These appliances include dishwasher, fridge-freezer, kettle, microwave, toaster and washing machine, which actually represent some of the most commonly used appliances in a UK household.

Table 4.11 shows the obtained classification accuracy with regards to TP, FP, FN from confusion matrix, precision PR, recall RE and F-measure F_M. The proposed method was able to achieve high precision during classification of more than 85% for most appliances, with the exception of microwave, dishwasher and washing machine. Microwave reported precision of 79%, as there were events caused by another appliance with similar signature that were falsely identified as microwave. Dishwasher and washing machine in the specific household have a similar signature in terms of consumption and duration during the washing cycle, which is the state that can be identified using the proposed method, thus in both cases many operations of one were misclassified as the other and vice versa. This can be seen clearly through Figure 3.2 in Section 3.4.4, where various examples of classification were presented. Therefore the obtained PR for dishwasher and washing machine was 41% and 59% respectively. The best performance was obtained during the classification of kettle with 99%, followed by fridge-freezer with 98%.

By assessing the performance of the proposed method using recall, which is a measure of understanding the amount of appliance operations that were not classified during the disaggregation, it is apparent that the proposed method was able to acquire recall more than 80% for most appliances, besides dishwasher and fridge-freezer. The best performances were reported for the kettle with 94% and the microwave with 93%. Dishwasher, as it was falsely identified in many cases as washing machine, was only

Table 4.11: Classification Accuracy for REFIT House 2.

Apps	*TP*	*FP*	*FN*	*PR*	*RE*	F_M
Dishwasher	9	5	35	0.41	0.20	0.27
Fridge-Freezer	209	4	270	0.98	0.44	0.61
Kettle	255	2	17	0.99	0.94	0.96
Microwave	239	65	19	0.79	0.93	0.85
Toaster	56	10	12	0.85	0.82	0.83
Washing Machine	64	45	12	0.59	0.84	0.69

Note: Apps=Appliances. TP=True Positive, FP=False Positive, FN= False Negative, PR=Precision, RE=Recall, F_M=F-Measure.

able to obtain 20% recall. Although the classification of the fridge-freezer, was accurate in terms of precision, it suffered from increased number of missing operations, which resulted to a recall of 44%.

With regards to F_M, the proposed method obtained classification accuracy of more than 80% for most appliances, but reported the lowest accuracy of 27% for the dishwasher. Fridge-freezer and washing machine reported classification accuracy of 61% and 69% respectively, while kettle obtained the best results with F-measure at 96%.

Similarly with the disaggregation of the REDD dataset, the performance evaluation of the proposed DTW-based method will be performed using for classification accuracy, F_M, for estimation accuracy, the Acc_i per appliance, and the normalised total power consumption estimation error (TER), as proposed in [5]. Tables 4.12 and 4.13 show the comparison between the F_M and Acc_i accuracies obtained using the proposed method with benchmarks the algorithms presented in [5]. Figures 4.12 and 4.13 capture graphically the same results, for visual comparison.

The proposed method was able to outperform all methods for all appliances, except for dishwasher, where it obtained the lowest classification accuracy 0.27, comparable only to UGSP and FHMM, where the authors in [5] reported accuracies of 0.32 and 0.23 respectively. The proposed method acquired F_M of 0.61 for fridge-freezer, which was on average 10% more than most of the methods in [5].

Table 4.12: Classification Accuracy using F_M for the proposed DTW-based NILM method, with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REFIT House 2, as presented in [5].

		UGSP[4]			*SGSP*[33]			*DT*[32]			
Apps	*DTW*	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.27	0.32	0.41	**0.79**	0.54	0.52	0.73	0.64	0.66	0.73	0.23
FFZ	**0.61**	0.47	0.52	0.42	0.32	0.33	0.59	0.30	0.54	0.33	0.50
K	**0.96**	0.66	0.68	0.88	0.77	0.67	0.90	0.45	0.42	0.76	0.06
MW	**0.85**	0.42	0.48	0.73	0.35	0.44	0.84	0.32	0.31	0.80	0.09
T	**0.83**	0.40	0.25	0.54	0.52	0.53	0.64	0.60	0.54	0.55	–
WM	**0.69**	0.24	0.21	0.23	0.32	0.29	0.43	0.37	0.32	0.48	0.09

Note: Apps=Appliances, DW=Dishwasher, FFZ=Fridge-Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine. **Bold** values represent the method/methods with the highest obtained classification accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Table 4.13: Estimation Accuracy using Acc_i for the proposed DTW-based NILM method, with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REFIT House 2, as presented in [5].

		UGSP[4]			*SGSP*[33]			*DT*[32]			
Apps	*DTW*	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.34	0.33	0.40	0.42	0.40	0.43	**0.67**	0.28	0.31	0.61	0.30
FFZ	0.69	0.42	0.34	0.77	0.58	0.56	**0.80**	0.37	0.32	0.73	0.24
K	**0.91**	0.49	0.68	0.83	0.51	0.51	0.85	0.41	0.39	0.76	−0.34
MW	**0.79**	0.51	0.48	0.64	0.55	0.53	0.65	0.33	0.42	0.64	−3.17
T	**0.85**	0.37	0.48	0.62	0.31	0.36	0.66	0.26	0.22	0.58	–
WM	**0.73**	0.46	−1.59	0.43	0.47	0.19	0.48	0.44	0.17	0.35	−1.84

Note:Apps=Appliances, DW=Dishwasher, FFZ=Fridge-Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine. **Bold** values represent the method/methods with the highest obtained estimation accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

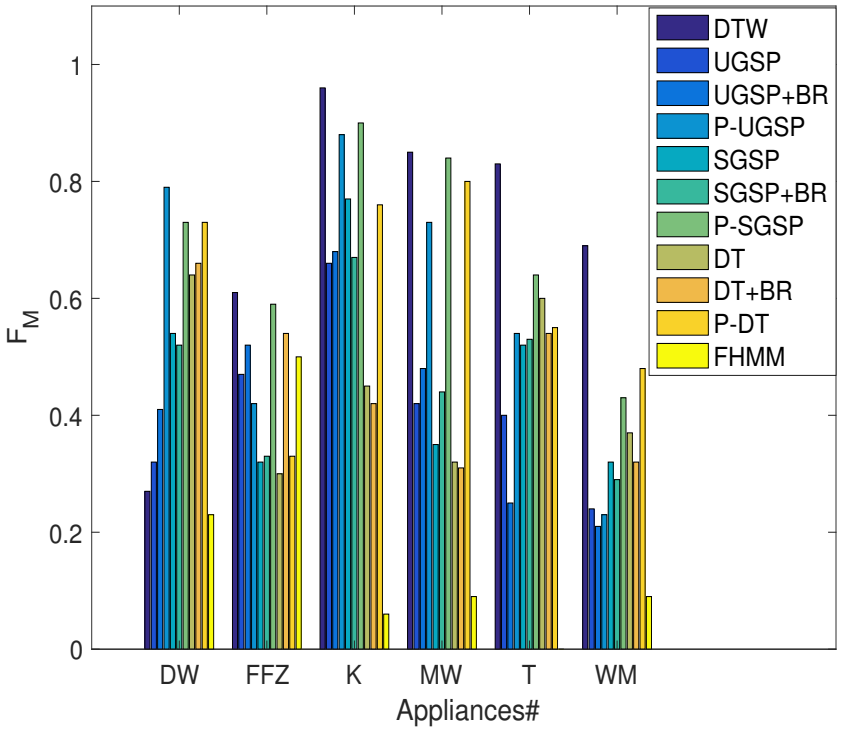


Figure 4.12: Performance evaluation using F_M for the proposed DTW-based NILM method (Tables 4.12) with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REFIT House 2, as presented in [5].

For the case of the kettle, the proposed method outperformed all method presented in [5], by more than 20% for most cases and as much as 90% for FHMM. Similarly, microwave reported the best classification accuracy at 0.85 for the proposed method, outperforming the methods in [5], by $12\% - 76\%$ in most cases.

Furthermore, the proposed method was able to obtain higher classification accuracy for the toaster compared to the methods in [5], by on average more than 20%. Similarly, for the washing machine, even though there was a misclassification between the washing machine and the dishwasher, the proposed method was able to report F_M of 0.69, higher by more than 20% from the NILM methods in [5].

Similarly with the classification accuracy, the proposed method achieved higher values of estimation accuracy for most appliances, compared to the methods used in [5], as seen in Table 4.13 and Figure 4.13. Estimation accuracy for kettle, was higher using the proposed method by at least 15% compared with the methods in [5].

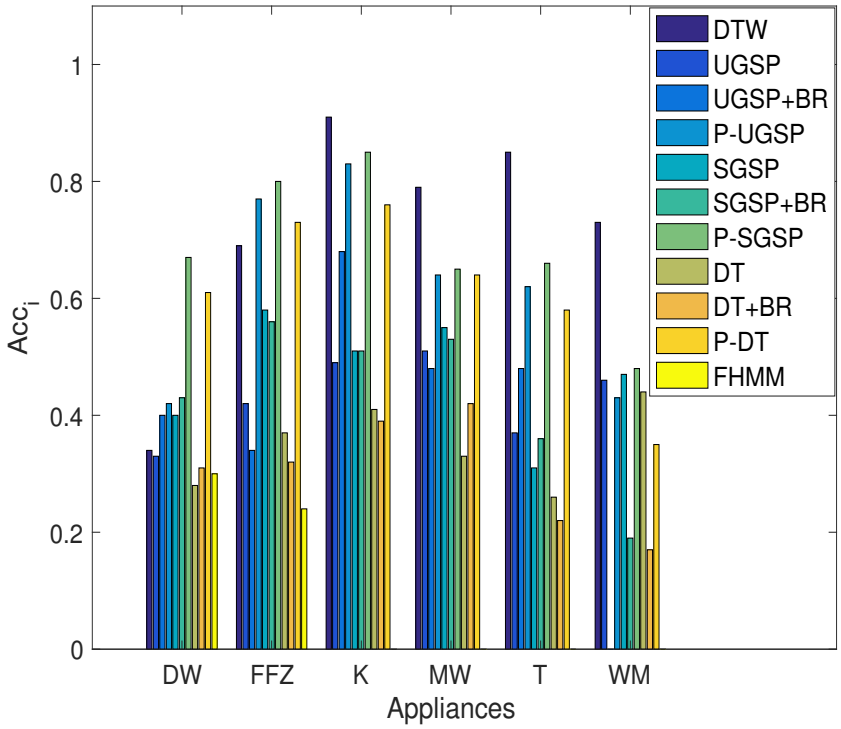


Figure 4.13: Performance evaluation using Acc_i for the proposed DTW-based NILM method (Table 4.13) with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REFIT House 2, as presented in [5].

Furthermore, microwave consumption was estimated more accurately using the proposed method, with Acc_i at 0.79, which was higher by more than 10% compared to NILM methods presented in [5]. The proposed method was able to outperform all methods for the toaster case by obtaining an estimation accuracy of 0.85, which is more than $\sim$ 20% than the methods used for disaggregation presented in [5]. Similarly, the best estimation accuracy for the washing machine, reported at 0.73, was obtained using the DTW-based method, which was better by more than 25% compared to the methods in [5].

The appliances that the proposed method acquired lower Acc_i and was outperformed by a variety of the methods implemented in [5], were the dishwasher and the fridge-freezer. For the dishwasher, the obtained Acc_i was 0.34, mainly caused by the missing dishwasher operations and the inability of the proposed method to identify the other states of the dishwasher. Although, this accuracy is one of the lowest compared to the methods in [5], the proposed method was able to report comparable results with

most of the methods used for benchmarking, with less than 10% difference from those methods, that reported lowest estimation accuracy but were more successful than the proposed method. The best Acc_i was achieved using P-SGSP [5, 33] at 0.67, followed by P-DT [5, 32] at 0.61.

Finally, for fridge-freezer, the proposed method obtained estimation accuracy of 0.69 and was able to outperform by at least 10% most of the methods in [5], with the exception of the methods using the P pre-processing and refinement method.

For further evaluating the performance of the proposed method, Table 4.14 presents the TER estimation error. The proposed method was able to obtain the smallest estimation error for the toaster case, where the obtained TER was 0.31, more than 40% less than most of the methods in [5], with the exception of P-DT, which reported a TER error of 0.42. For the dishwasher, the DTW-based method was able to outperform, with an error of 0.44, the majority of the methods in [5], apart from the approaches using the BR methods and P-SGSP [5, 33].

The proposed method obtained the highest TER error at 0.7, while estimating the consumption of the fridge-freezer, only comparable to the 0.67 received using DT with BR. The smallest error in [5] was obtained using UGSP, with only 0.1, more than 60% lower than the TER reported for the proposed method. The estimation error for kettle was 0.019 using the proposed method, which was the lowest error compared to the methods proposed in [5].

For microwave, the proposed method was able to obtain the lowest TER error at 0.09 compared to the methods in [5]. For the toaster, the proposed method reported TER error of 0.21, which was the lowest compared to the methods in [5] by at least 20%. The proposed method reported one of the highest errors for the washing machine at 0.42 and was only able to outperform SGSP+BR and FHMM. Finally, for the dishwasher the proposed method obtained total estimation error of 0.43, only higher than the methods using the BR method and P-SGSP.

Table 4.14: Normalised total power consumption estimation error (TER) for the proposed DTW-based NILM method, with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REFIT House 2, as presented in [5].

		UGSP[4]			*SGSP*[33]			*DT*[32]			
Apps	*DTW*	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.43	0.73	**0.08**	0.66	0.47	0.13	0.33	0.64	0.35	0.49	0.75
FFZ	0.7	**0.10**	0.53	0.31	0.19	0.42	0.27	0.47	0.67	0.31	0.38
K	**0.019**	0.34	0.06	0.05	0.27	0.11	0.04	0.31	0.25	0.15	0.41
MW	**0.09**	0.43	0.17	0.34	0.52	0.16	0.28	0.66	0.59	0.63	13.63
T	**0.21**	0.83	0.81	0.82	0.90	0.88	0.73	0.78	0.71	0.42	7.53
WM	0.42	0.17	0.40	0.09	0.32	0.52	**0.10**	0.28	0.33	0.22	3

Note: Apps=Appliances, DW=Dishwasher, FFZ=Fridge-Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine. **Bold** values represent the method/methods with the lowest TER error per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

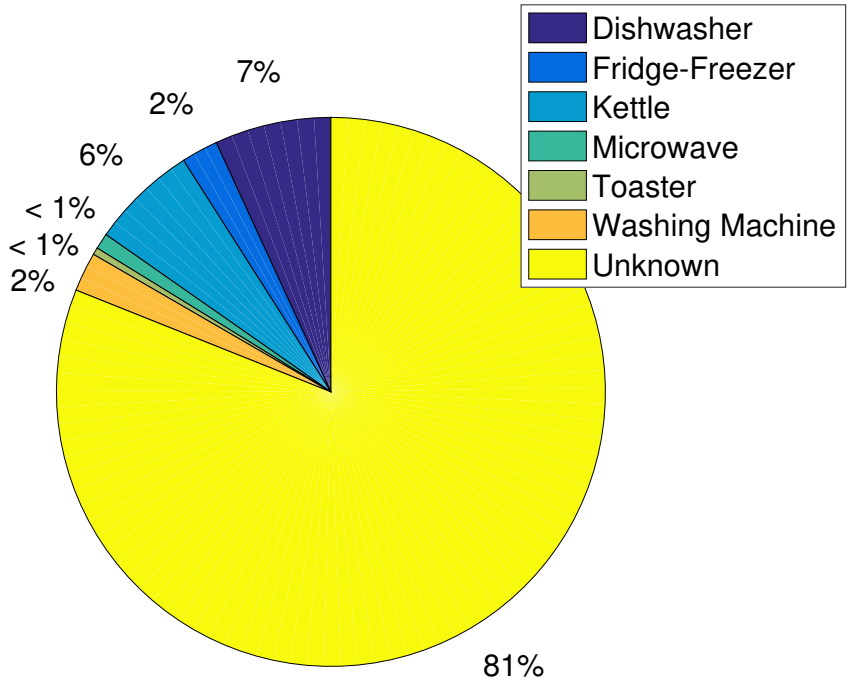


Figure 4.14: Disaggregation load contribution using the proposed DTW-based NILM method, as percentage load contribution per appliance relative to the aggregate load, for House 2 from REFIT dataset.

Figure 4.14 shows the disaggregation results of the proposed DTW-based NILM method as percentage of load contribution per appliance with regards to the aggregate load and Figure 4.15, the relevant ground truth consumption per appliances, available from the IAMs of the REFIT dataset [2, 3]. The term "*unknown*" corresponds to any appliances that were present in the aggregate load but were not disaggregated using the DTW-based method.

As expected, due to the missing operations of dishwasher during classification, the proposed method only estimated 7% contribution of the dishwasher to the total consumption, whereas the actual contribution was 12%. Similarly, for fridge-freezer, the contribution was estimated at 2%, while the contribution using the IAMs information was accounted for 7%, which again is the result of the increased number of not identified fridge-freezer operations, as seen in Table 4.11.

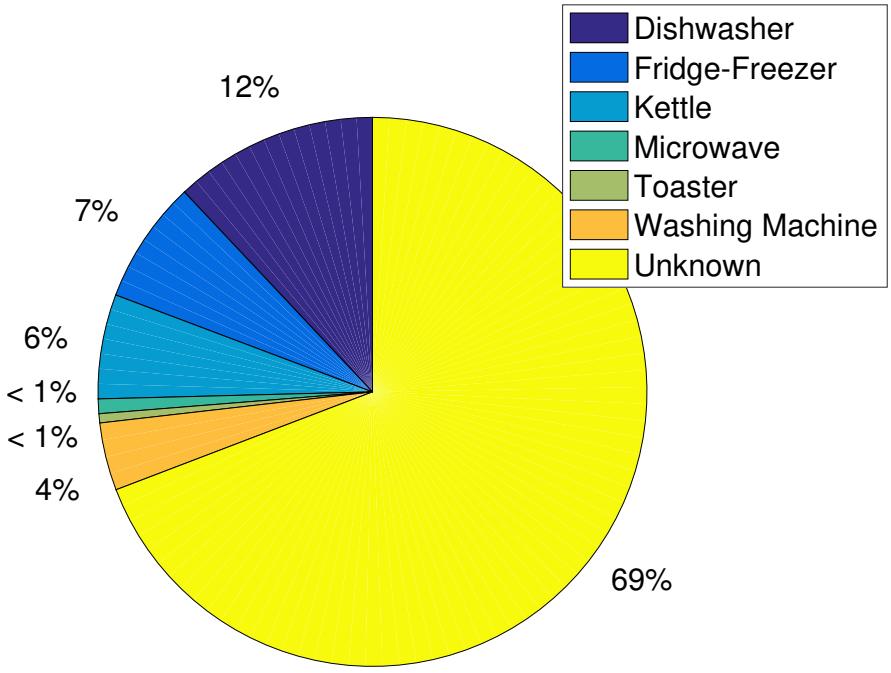


Figure 4.15: Ground truth load contribution, as percentage load contribution per appliance relative to the aggregate load, for House 2 from REFIT dataset.

The proposed method accurately estimated the contribution for the case of the kettle, the microwave and the toaster and underestimated the contribution of the washing machine by 2%, as expected due to both the missing operations of the washing cycle and the fact that the proposed method is not able to recognise the other states of the washing machine.

In conclusion, the proposed method was able to obtained the best performance for classification and estimation accuracy for the kettle, the microwave, the toaster, the washing machine and the fridge-freezer, with the exception of the latter for the case of estimation accuracy, where it was only outperformed by the P refined methods, as presented in [5]. During classification, the proposed method had a difficulty to separate the operation of the dishwasher and the washing machine, which resulted to missing and wrongly identified events corresponding to the operation of those appliances. Therefore the performance of the DTW-based method for dishwasher was among the lowest compared to the methods in [5] used for benchmarking purposes. The normalised

total consumption estimation error was either lower or comparable for most cases, with the exception of fridge-freezer, where the proposed algorithm obtained the highest error.

4.4.2 REFIT House 17 Disaggregation Case Study

The final case study, presented in this chapter, is for house 17 from the REFIT dataset [2, 3], for which the authors in [5] have investigated the performance of the presented methods for the following appliances: fridge-freezer, freezer, kettle, microwave, toaster and washing machine. The proposed DTW-method will be evaluated using the obtained disaggregation results for freezer, kettle, microwave and toaster, with benchmarks the methods proposed in [5]. Similarly with the previous case studies, the disaggregation was obtained using the aggregate load, where all appliances were present.

Unfortunately, the proposed method was not able to obtain a signature for the washing machine, as during the data used for the creation of the library there was only one operation of the appliance, which consequently means that the appliance was not able to be classified during the testing period. Similarly, for the case of the fridge-freezer, the DTW-based was not able to obtain appliance signatures successfully, even though it is an appliance that was visible multiple times during each day. The main reason for that was that according to the IAMs the fridge-freezer consumes on average less than 80 watt, which in the aggregate data it was estimated with much smaller start and end edges, which created a challenge for the proposed method for obtaining signatures and disaggregate during the testing period.

Table 4.15 presents the classification accuracy results for the proposed method, using TP, FP, FN from confusion matrix, PR, RE and more importantly F-measure F_M. The DTW-based method was able to obtain precision accuracy of more than 70% for freezer, kettle, and toaster, with the highest at 99% occurring during the disaggregation of freezer, which only reported two wrongly positively identified freezer operations. For the microwave, the reported precision was only 58%, due the increased number of FP, as there was an appliance operation with similar consumption and duration, with the microwave signature.

In terms of recall, the proposed method obtained accuracy of more than 80% for most disaggregated appliances, with the exception of freezer, where the algorithm reported a high number of not identified operations, which account for almost 65%

of the total number of freezer operations, resulting to a recall of only 35%. With regards to F_M, the proposed method reported accuracy of more than 80% for both kettle and toaster, and 52% and 68% for freezer and microwave, respectively, due to the low precision and recall observed in each case.

Table 4.15: Classification Accuracy for REFIT House 17.

Apps	*TP*	*FP*	*FN*	*PR*	*RE*	F_M
Freezer	183	2	339	0.99	0.35	0.52
Kettle	209	80	12	0.72	0.95	0.82
Microwave	73	52	15	0.58	0.83	0.68
Toaster	92	14	9	0.87	0.91	0.88

Note: Apps=Appliances. TP=True Positive, FP=False Positive, FN= False Negative, PR=Precision, RE=Recall, F_M=F-Measure.

Similarly with the case study for the house 2 from REFIT dataset, the performance evaluation of the proposed DTW-based method will be performed using for classification accuracy F_M and for estimation accuracy, the Acc_i per appliance, with benchmarks the algorithms presented in [5]. Tables 4.16 and 4.17 show the F_M and Acc_i accuracies obtained using the proposed method versus those obtained using the benchmarking algorithms from [5]. Figures 4.16 and 4.17 show a graphical representation of those results, for better understanding.

The best classification accuracy for the proposed method, according to Table 4.16, was obtained for the toaster, which was reported at 0.88, outperforming all the methods presented in [5] by more than 10% for most cases. For the freezer, the proposed method reported 0.52, and was outperformed by all implementations of DT [32], FHMM [34] and the rest of the P refined methods [5].

For the kettle, P-SGSP [4, 33] and P-DT [5, 32] obtained the highest classification accuracy, at 0.96 and 0.95, whereas the proposed method was able to achieve an accuracy of 0.82. The reported accuracy was comparable or higher to the rest of the approaches in [5].

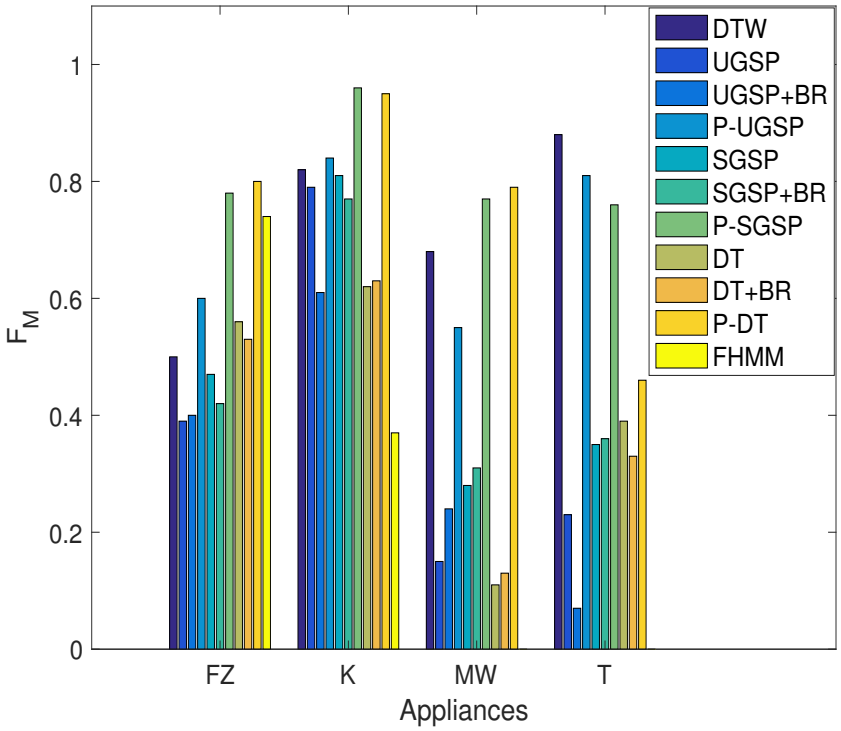


Figure 4.16: Performance evaluation using F_M for the proposed DTW-based NILM method (Tables 4.16) with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REFIT House 17, as presented in [5].

Finally, the proposed method reported F-measure 0.68 for the microwave disaggregation and was able to obtain better accuracy, compared to most of the methods implemented in [5], by more than 30%, with the exception the P implementations of DT and SGSP which obtained 0.79 and 0.77, respectively.

Furthermore, according to Table 4.17 and Figure 4.17, the DTW-based method was able to achieve the highest estimation accuracy Acc_i for most appliances, besides freezer, where P-DT [5, 32] and FHMM [34] reported 0.68 and 0.71 respectively, outperforming the proposed algorithm, which obtained estimation accuracy of 0.65. The proposed method was able to outperform the rest of the methods by more than 20%.

Table 4.16: Classification Accuracy using F_M for the proposed DTW-based NILM method, with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REFIT House 17, as presented in [5].

		UGSP[4]			*SGSP*[33]			*DT*[32]			
Apps	*DTW*	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
FZ	0.52	0.39	0.40	0.6	0.47	0.42	0.78	0.56	0.53	**0.80**	0.74
K	0.82	0.79	0.61	0.84	0.81	0.77	**0.96**	0.62	0.63	0.95	0.37
MW	0.68	0.15	0.24	0.55	0.28	0.31	0.77	0.11	0.13	**0.79**	–
T	**0.88**	0.23	0.07	0.81	0.35	0.36	0.76	0.39	0.33	0.46	–

Note: Apps=Appliances, FZ=Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine. **Bold** values represent the method/methods with the highest obtained classification accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the pre-processing and refinement NILM method proposed in [5].

Table 4.17: Estimation Accuracy using Acc_i for the proposed DTW-based NILM method, with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REFIT House 17, as presented in [5].

		UGSP[4]			*SGSP*[33]			*DT*[32]			
Apps	*DTW*	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
FZ	0.65	0.04	0.43	0.64	0.21	0.26	0.63	0.26	0.21	0.68	**0.71**
K	**0.90**	0.12	0.66	0.79	0.32	0.31	0.80	0.43	0.43	0.77	0.20
MW	**0.68**	0.47	0.23	0.55	0.47	0.54	0.55	0.59	0.57	0.61	–
T	**0.88**	0.37	0.23	0.80	0.42	0.39	0.81	0.54	0.50	0.71	–

Note: Apps=Appliances, FZ=Freezer, K=Kettle, MW=Microwave, T=Toaster. **Bold** values represent the method/methods with the highest obtained estimation accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

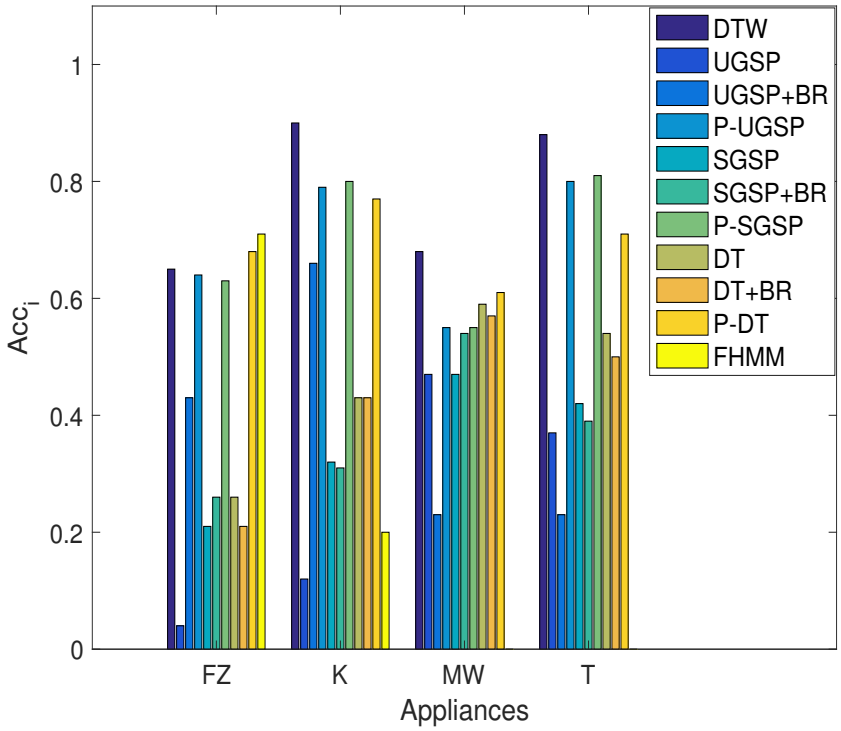


Figure 4.17: Performance evaluation and Acc_i for the proposed DTW-based NILM method (Table 4.17) with benchmarks UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REFIT House 17, as presented in [5].

For estimating the kettle operation, the proposed method was able to obtain accuracy of 0.9, which is more than 10% higher compared to the Acc_i reported in [5]. Similarly, accuracy of 0.68 was reported for the proposed method for the case of the microwave, higher than all the methods presented in [5], but only by a small percentage. Finally, the proposed method was able to outperform all methods in [5] for estimating the toaster's consumption, as it was able to achieve Acc_i of 0.88, which was more than 30% compared to most of the presented methods in [5].

Figure 4.18 presents the disaggregation results as a percentage of contribution to the total aggregate load and Figure 4.19 shows the ground truth load contribution, obtained using the data from the individual monitors, available in the REFIT dataset [2, 3]. Similarly, with what has been discussed in the previous case studies, for any appliances that have not been disaggregated, their contribution to the aggregate load has been included in the "*unknown*".

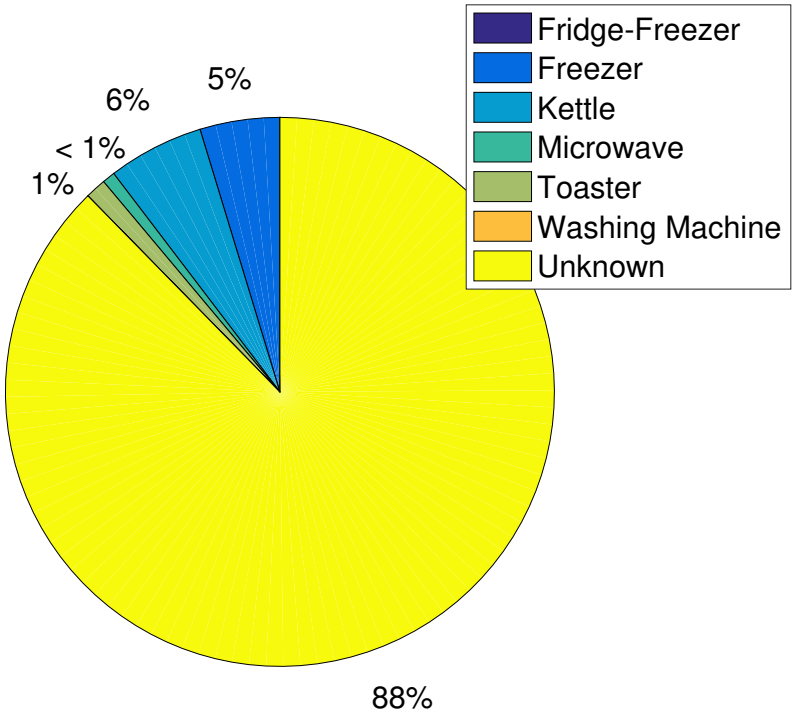


Figure 4.18: Disaggregation load contribution using the proposed DTW-based NILM method, as percentage load contribution per appliance relative to the aggregate load, for House 17 from REFIT dataset.

Both fridge-freezer and washing machine, that were not disaggregated using the proposed method, are included in Figure 4.19, in order to show the actual contribution of the specific appliances with respect to the total consumption, which accounts for 7% and 2% respectively. The DTW disaggregation classified those appliances contribution under the unknown appliances, which can explain partially the 20% difference between the reported unknown contribution using the proposed method and the actual unknown contribution. Another reason, that contributed significantly, was that the freezer contribution was only estimated at 6%, while the actual freezer contribution was 14%. For both kettle and toaster, it was in line with the actual consumption contribution reported from the IAMs, while microwave was estimated at $< 1\%$, compared to the ground truth 1%.

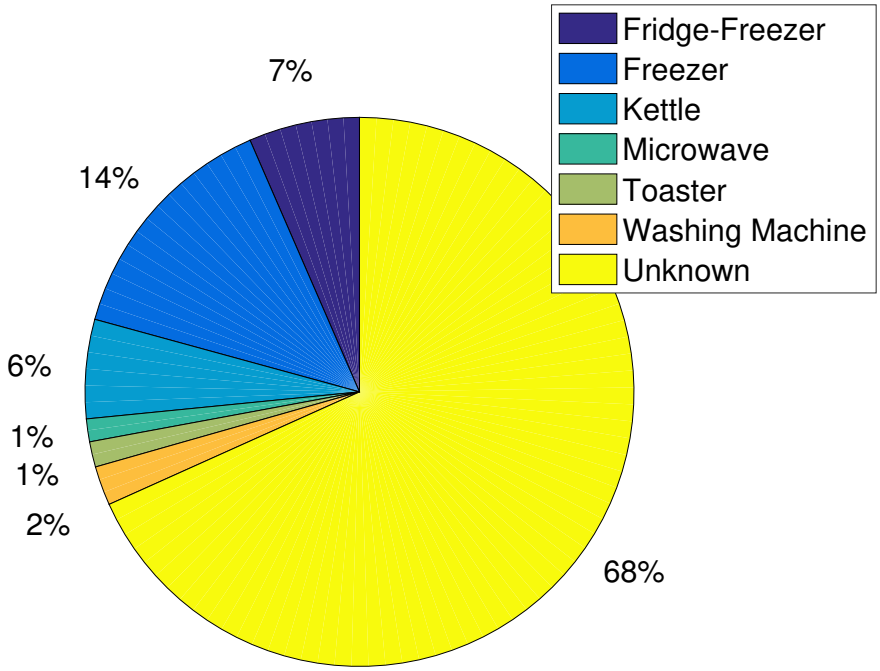


Figure 4.19: Ground truth load contribution, as percentage load contribution per appliance relative to the aggregate load, for House 17 from REFIT dataset.

To conclude the case study for the specific house, the proposed method was able to obtain higher or at least comparable classification and estimation accuracy for most appliances, and was able to outperform most of the methods in [5] for those appliances. In terms of load contribution, the DTW-based method was able to estimate accurately the majority of the appliances, with the exception of the freezer, where the algorithm underestimated the contribution, due to the increased non-classified freezer operations. With respect to the fridge-freezer and washing machine, as already discussed, the proposed algorithm was not able to extract appliance signatures, therefore disaggregate the appliances from the aggregate load.

4.5 Overall Performance Evaluation

This section focuses on evaluating the overall performance of the proposed unsupervised DTW-based method with benchmarks using the methods presented in [5] and any advantages and limitations that were observed during the disaggregation process with the proposed method. Firstly, Table 4.18, presents the overall classification accuracy using $F1_{micro}$, as defined in equation 2.6, which calculates the *micro-average* using the total values of TP, FP and FN for all classified appliances.

Table 4.18: Overall Classification Accuracy using $F1_{micro}$, as defined in equation 2.6, with benchmarks P-SGSP [5, 33], P-UGSP [4, 5] and P-DT [5, 32], as presented in [5]. DTW presents the overall accuracy for only the disaggregated appliances, whereas m-DTW is accounting for the missing operations of the appliances that where not disaggregated using DTW, but were present in the disaggregation using the methods in [5].

Datasets	*REDD*		*REFIT*	
NILM Methods	*House 1*	*House 2*	House 2	House 17
DTW	0.79	0.80	0.77	0.68
m-DTW	0.49	0.80	0.77	0.51
P-SGSP	0.69	0.73	0.61	0.62
P-UGSP	0.67	0.72	0.59	0.59
P-DT	0.70	0.68	0.59	0.60

The proposed algorithm was able to disaggregate the same appliances with the benchmark methods for both house 2 in REDD and REFIT, but, as already discussed, was not able to obtain signatures and therefore classify bathroom GFI, lighting and oven for house 1 in REDD and fridge-freezer and washing machine for the case of house 17 in REFIT. Therefore, in addition to the reported overall classification accuracy obtained by the disaggregated appliances using DTW, Table 4.18 and Figure 4.20, which represents graphically the results presented in the table, include the *m-DTW*, which shows the accuracy by taking into account all the missing operations of the non-disaggregated appliances, that were successfully classified in [5].

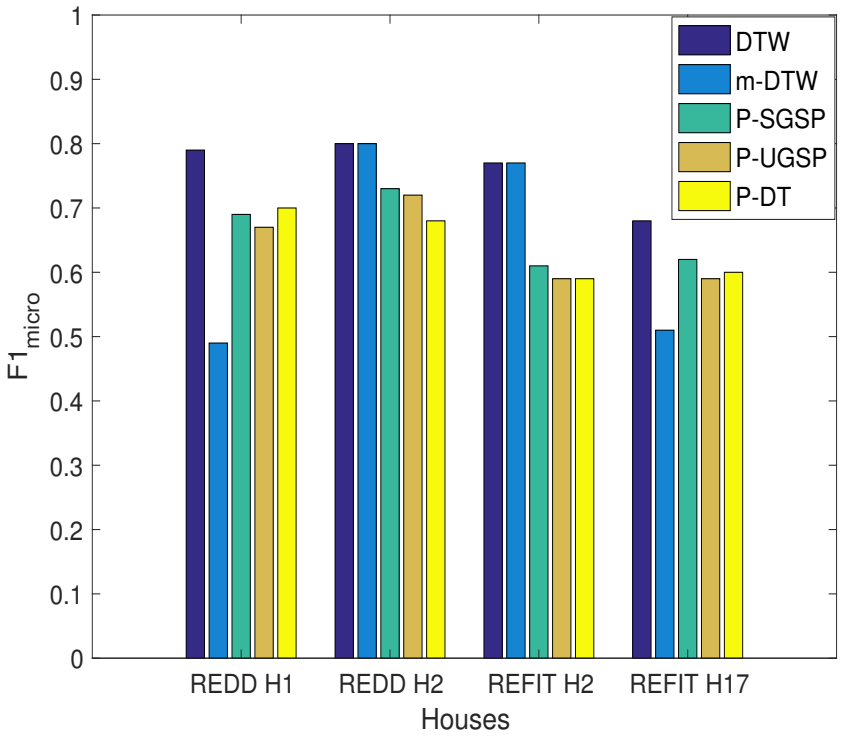


Figure 4.20: Overall classification performance evaluation using $F1_{micro}$ for the proposed DTW-based NILM method, as presented in Table 4.18, with benchmarks P-SGSP [5, 33], P-UGSP [4, 5], P-DT [5, 32], HMM [35], for Houses 1 and 2 from REDD dataset [1] and Houses 2 and 17 from REFIT [2, 3], as presented in [4, 33]. m-DTW presents the overall performance by taking into account the FN of the appliances that were disaggregated in [5] but not using the proposed DTW.

In general, the proposed method was able to obtain the highest overall classification accuracy for house 2 in both REDD and REFIT dataset, with 0.8 and 0.77 respectively. For house 1 from the REDD dataset, when reporting only the classified appliances, DTW was able to outperform the methods in [5], but for the fair comparison, when taking into account the *m-DTW* results, it only obtained 0.49, which is on average less than 20% compared to the other methods. For the case of bathroom GFI, the proposed method was *"confusing"* the events as microwave. Note that the signatures that were corresponding to the bathroom GFI operation, were removed during the automated labelling using the IAMs data.

For house 17 from the REFIT dataset, the proposed unsupervised DTW-method, obtained accuracy of 0.68 and 0.51, when taking into account the missing fridge-freezer and washing machine. The *m-DTW* accuracy is lower compared to the performance of

the methods presented in [5], but only by on average 10%.

Furthermore, the proposed method reported high classification and estimation accuracy for the kitchen appliances, namely kettle, toaster and microwave, for most of the houses in REDD and REFIT dataset. Only the high consuming washing state of both dishwasher and washing machine was obtained using the proposed method, which resulted in higher recall values, affecting the reported F_M accuracy and underestimation of the consumed power. Specifically, for house 2 in the REFIT dataset, the DTW-based method had a difficulty in separating the dishwasher and washing machine operation, as the washing state of both reported similar average consumption and duration.

For the majority of the houses, the proposed method has shown classification accuracy for fridge/fridge-freezer more than 0.65, with the exception of the house 17 in REFIT, were fridge-freezer was not disaggregated and the reported accuracy for the freezer was 0.52, which was still comparable to most of the methods in [5], as discussed in detail throughout Section 4.4.2.

Although the overall performance of the proposed DTW-based method was satisfactory compared to the methods presented in [5], it has been observed that the proposed methods had shown increased computational complexity. This limitation is partly caused by the iterative removal of the recognised load from the aggregated data, in order to perform classification of the lower consuming appliances. Furthermore, the DTW classification performs a point to point comparison in order to minimise the distance, and find the optimal solution. During the testing phase, the proposed method was comparing each of the obtained events with all the available appliance signatures available in the library. On average for the REDD dataset, the algorithm required 3 hours in order to create the library of appliance signatures and testing more than 10 hours, which was higher for the case of the REFIT data, where the data are reported on average every 8sec. Therefore, it is important to investigate ways of addressing the computational complexity and optimise the disaggregation process.

4.6 Summary

The aim of this chapter was to implement the DTW-based algorithm, proposed in Chapter 3, and evaluate its performance with benchmarks to various state-of-the-art NILM algorithms. The algorithm was implemented and tested using REDD [1], downsampled to $1min$, and REFIT [2, 3] datasets, and the obtained results were compared to various state-of-the-art NILM methods, namely DT [32], supervised GSP, [33], unsupervised GSP [4], HMM [35] and FHMM available in NILMTK toolbox [34], as presented in the works of [4, 5, 33].

The performance of the proposed method was evaluated in terms of classification accuracy using F-measure (F_M) and estimation accuracy using Acc_i per appliance. The normalised total power consumption estimation error (TER), proposed in [5], was reported for house 1 from REDD and house 2 from REFIT, in order to compare with the error obtained by the methods used in [5]. Furthermore, the consumption contribution of the obtained disaggregation were presented with regards to the actual appliance contribution, as obtained using the individual monitors.

The proposed method was able to achieve high classification and estimation accuracy for most houses and appliances, with similar or better performance than most of the proposed methods in [5]. For the appliances, where it obtained lower accuracy, the proposed method was in general outperformed by the methods implemented in [5], using the pre-processing and refinement method P proposed by the authors of the same work.

By evaluating the overall classification accuracy, the proposed NILM method was able to obtain better results in two houses (house 2 in both REDD and REFIT), but the reported accuracy was lower for the rest, as the proposed method was not able to disaggregate all the appliances presented in [5].

Furthermore, the proposed algorithm has shown high computational complexity, which for the case of online and real time disaggregation is a great limitation. Therefore, it is essential to identify methods that would optimise the original proposed method in terms of complexity, event detection and classification. For this purpose, Chapter 5 will present a scheme using a variation of the DTW-based method for creating the library of appliance signatures combined with k-means for classification and will evaluate the

performance of the original and the optimised approach in terms of both accuracy and computational complexity.

Chapter 5

Optimised DTW Disaggregation

5.1 Introduction

In Chapter 4, an unsupervised method has been proposed for load disaggregation using solely DTW for classification, after the creation of the appliance signatures library using the same algorithm. The method was able to obtain on average good performance in terms of classification and estimation accuracy, comparable to most of the methods used for benchmarking, but inevitably incurred high computational complexity due to the exhaustive distance comparison between the library signatures and the acquired appliance events. DTW distance comparison is employed to perform classification and the step-by-step removal of the already classified load from the aggregate data, in order to continue disaggregating the remaining load.

For this purpose it has been decided to investigate methods, could be used in conjunction with DTW, to leverage the performance of DTW-based algorithm but reduce its complexity, in particular the classification step. k-means and its variations are one of the most commonly used methods of clustering and has applications in various fields and, as already discussed in Chapter 2, has been successfully used in different NILM approaches for either training or disaggregation. As a method, it offers fast performance and can be used for unsupervised learning.

5.1.1 k-means Background

k-means is commonly used for clustering and classification for various research problems, as it provides both semi-supervised and unsupervised learning. Throughout the NILM background, as seen in Chapter 2, many researchers have incorporating k-means and its variations for various purposes. Hassan *et al.* [87] have applied k-means to the obtained features in order to cluster the features and label the events. In both [117, 118], the

authors have implemented k-means in the form of supervised clustering in order to reduce the training size of the SVM method used for classification. Gonçalves *et al.* [154] have used a genetic k-means in order to acquire the representative appliance clusters automatically.

More recently, Kong *et al.* [176] have applied an iterative k-means for identifying the hidden states of HMM and have performed disaggregation of the household load using segmented integer quadratic constraint programming. In the context of DTW, as already mentioned in 3.3, DTW was combined with k-means as a distance measure in [170], instead of the use of the Euclidean distance, for clustering load profiles of electricity customers.

5.2 Unsupervised Combined DTW and k-means Method

The proposed combined DTW and k-means method, referred as DTW+kM is following the same steps as discussed in Chapter 4, namely data pre-processing, event detection, classification via clustering and performance evaluation. During data pre-processing the same approach, as discussed in Section 3.4.1, has been used. The following steps of event detection, feature extraction and classification via clustering and any differences in their approaches will be presented in more detail in the following sections.

5.2.1 Event Detection and Feature Extraction for the Combined DTW+kM Method

Event detection follows the same procedure as defined in Algorithm 3.1 during the implementation of the original proposed DTW method in Chapter 4. For optimisation of the computational complexity, it has been decided to obtain the rising l_s and falling edges l_e of the possible appliance events, without iterative change of threshold W, defined in equation 3.10, and removal of the already detected events from the aggregate load. Parameter C was set at 20% of the detected rising edge, similarly with Chapter 4, for detected load of ≤ 300 watt, which represents lower consuming appliances, and 5% of the rising edge for the rest of the detected load.

Similarly, feature extraction is performed using the steps presented in Algorithm 5.1, in order to obtain the matrix of extracted features $\mathbf{E}$. The algorithm uses the

aggregate data and the variations that occur between the two points, as long as there are $|\Delta p_t| < C$, where $t = l_s + 1, \ldots, l_e - 1$. As the events are not obtained through the iterative removal of the load, when this condition is not satisfied, the algorithm forward fills the missing event values by interpolating between the last obtained power value and the value of the closing edge. For every feature extracted the algorithm can obtain the average power of each extracted event, in the form of a vector **average**, which will be used during the disaggregation using the k-means clustering method.

Algorithm 5.1 Feature Extraction: For a given set of events and proposed DTW-based method, return extracted features.

function FEATURE EXTRACTION($DTWmethod$, $Event$, p)
  for $k = 1 : P$ **do**
    $E(k,:) = \Delta p(Event(k,1) : Event(k,2))$
    $average(k) = average(E(k,:))$
  end for
  return **E**, **average**
end function

5.2.2 Disaggregation using the Combined DTW+kM Method

The proposed DTW+kM method creates the appliance signatures library using the same process as the DTW method using the **E** features extracted as found in 5.1. Accordingly, DTW pattern matching is following the same process as described in Chapter 3, shown algorithmically in 3.3. In addition to the **DTWLIB**, which represents the library of the appliance signatures, the algorithm returns the number of signatures available in the final library *numcluster* and the mean of each signature in the form of a vector **centroid**, which will be used to initialise k-means during disaggregation of the test aggregate load data in terms of number of clusters and centroids.

DTW is able to obtain a library of signatures, without the previous knowledge of both the number and type of appliances present in the aggregate load and therefore it can automatically provide the number of clusters, that can be used in k-means. Otherwise the user would have to select them manually, based on the number of appliances, obtained through the IAMs or house surveys, resulting in a supervised implementation of k-means, which was not the desired outcome of this thesis. Furthermore, the process of creating the library using DTW method can be found in Algorithm 5.2.

Algorithm 5.2 Appliance Library Creation: Using the extracted features **E** create the library of appliance signatures.

function LIBRARYCREATION(**E**) ▷ Where **E** found in Algorithm 3.2
  DTWLIB = []
  centroid = []
  for $k = 1 : R$ **do**
    $i = 0$
    for $l = 0 : |\mathbf{DTWLIB}|$ **do**
      if $DTW(E(k,:), DTWLIB(i,:)) \geq DTW_{Thr}$ **then**
        $i++$
      end if
    end for
    if $i == |\mathbf{DTWLIB}|$ **then**
      DTWLIB = **DTWLIB** $\cup E(k,:)$
      centroid(i)=average(DTWLIB(i))
    end if
  end for
  return **DTWLIB**, $numcluster = numrows($**DTWLIB**$)$, **centroid**
end function

k-means is performing clustering by using the average power consumption $average(i)$ per event i, and is initialised using the centroids **centroid** obtained through DTW method for creating the **DTWLIB**, as seen in Algorithm 5.2. The number of clusters is defined by the number of unique signatures available in each library $numcluster$.

Algorithm 5.3 shows the inputs and outputs during the k-means clustering. $KMEANS$ function performs clustering and returns **clusters** which are the cluster indices of each event, the new centroids **newcentroid** and the distance matrix **Distance** which shows the distance between each event and the relevant centroid.

Algorithm 5.3 k-means clustering.

function KMEANS(**average**,$numclusters$,**centroid**)
  return **clusters**,**newcentroid**, **Distance**
end function

Algorithm 5.4 shows the classification of the disaggregated aggregate load via clustering using the DTW+kM method. As already discussed, after event detection and feature extraction, k-means is initialised using the centroids from **DTWLIB** and the number of signatures, and performs clustering for the average value of each identified features. After the clustering is performed, the algorithm compares the distance between of each clustered observation and if the distance $Distance(k) \leq W(i)$, the algorithm returns the classified event as then it outputs the event classified as $cluster(k)$. If this

is not true for any of the clusters the event is labeled not classified. The distance check is performed in order to ensure that even though we use the average consumption of its event, the distance of the clustered events and the relevant centroid is still less than the threshold C.

Algorithm 5.4 Disaggregation: Perform clustering on the new collected aggregate load measurements **p** using the DTW+kM method.

function DISAGGREGATION(p(TestStart:TestEnd),**centroid**,$numclusters$, $\mathcal{M}$))
  Mk = []
  for $i = 1 : M$ **do**
    Mk = [**Mk**$_i$]
    Calculate $W(i)$ with (3.10) using **Mk**
    $Event_{Disaggregation}$=EventDetection($p(TestStart : TestEnd)$,$W(i)$)
    [$E_{Disaggregation}, average_{Disaggregation}$]=FeatureExtraction($Event_{Disaggregation}, p(TestStart : TestEnd)$)
    [**clusters**, **newcentroid**, **Distance**]=kmeans(**average**$_{\textbf{Disaggregation}}$,$numclusters$,**centroid**)
    for $k = 1 : numclusters$ **do**
      if $Distance(k) \leq W(i)$ **then**
        $classified(k) = cluster(k)$
      else
        $notclassified$
      end if
    end for
    Output **classified** and **notclassified**
  end for
end function

DTW+kM method is an *unsupervised* NILM approach, as it performs clustering of unlabelled aggregate load measurements. The library of appliance signatures is obtained using historical unlabelled aggregate load measurements, as in Chapter 3. The number of clusters and the centroids used in order to initialise k-means, are obtained through the library, and not through a priori knowledge of the appliances operating in each house. Post-labelling is used in order to match the disaggregation results to the relevant appliance using the IAMs data.

5.3 Unsupervised Combined DTW and k-means Method using DTW Refinement

The combined scheme using the DTW and k-means with DTW refinement, performs classification following the procedure presented in Algorithm 5.5. After k-means performs the initial clustering and classification, the DTW refinement is performing pattern

matching for all the events that were not classified, due to the distance condition. The pattern matching using the refinement method is performed in a similar manner as the process explained in Chapter 3 for the original DTW method, with the exception that it does not remove the classified event operation from the aggregate load.

Algorithm 5.5 Disaggregation: Perform pattern matching using DTW on the non-classified classified aggregate load measurements **p**.

function DISAGGREGATION(p(TestStart:TestEnd), **DTWLIB**, $\mathcal{M}$))
  Mk = []
  for $i = 1 : M$ **do**
    Mk = [**Mk**$_i$]
    Calculate $W(i)$ with (3.10) using **Mk**
    $Event_{Disaggregation}$=EventDetection($p(TestStart : TestEnd)$,$W(i)$)
    $[E_{Disaggregation}, average_{test}]$=FeatureExtraction($Method$,
    $Event_{Disaggregation}, p(TestStart :$
      $TestEnd))$
    [**clusters**, **newcentroid**, **Distance**]=kmeans(**average**$_{\textbf{Disaggregation}}$,$numclusters$,**centroid**)
    for $k = 1 : numclusters$ **do**
      if $Distance(k) \leq W(i)$ **then**
        $classified(k) = cluster(k)$
      else
        notclassified
        for $j = 1 : NumberOfRowsInE_{Disaggregation}$ **do**
          $D = tag$(DTWClass($E_{Disaggregation}(j, :)$, **DTWLIB**))
          $classified1(j) = D$
        end for
      end if
    end for
    Output **classified** and **classified1**
  end for
end function

It is important here to note that the DTW refinement can be performed for the case of the events that were falsely identified as an appliance, after the initial comparison of the classified events to the ground truth operations of the IAMs.

Similarly to the DTW+kM method, kDTW method is an *unsupervised* NILM approach, as it performs disaggregation by only using unlabelled aggregate measurements and it creates the relevant library of appliance signatures using historical unlabelled aggregate load measurements. Pattern matching is performs via an optimised search of the library of appliance signatures, where it only compares new event the signatures with similar average consumption. Post-labelling is used to match the obtained appliance activity to the corresponding appliance using the IAMs data.

5.4 Performance Evaluation of the Proposed Methods

5.4.1 Experimental Setup

The experiments were run on a Toshiba Satellite Intel(R) Core(TM) i3-3110M CPU @2.40GHz with 4GB RAM. *Matlab 2016b* was used for the implementation and the execution of the proposed algorithm.

Similarly with Chapter 4, houses 1, 2 and 6 from REDD dataset [1] and houses 2 and 17 from REFIT [2, 3] have been used for the evaluation of the methods proposed in the current chapter. The periods used for creating the library of the appliance signatures for each house were the same, as discussed in Section 4.1.

The same performance evaluation metrics have been used with regards to classification and estimation accuracy, in line with those applied in Chapter 4. Furthermore, the obtained performance of the proposed DTW+kM and kDTW methods have been compared to that of the DTW-based method, proposed in Chapter 4.

5.4.2 Performance Evaluation using REDD Dataset

For the implementation of the proposed methods of this chapter, the same downsampled version ($1min$) of the REDD dataset has been used, in order to avoid any bias.

5.4.2.1 Disaggregation Case Study using House 1 from REDD Dataset

The disaggregation of house 1 from REDD dataset [1] includes the same appliances presented in Chapter 4, namely dishwasher, fridge, kitchen outlet, microwave, washer dryer, as the two methods presented in this chapter were not able to obtain signatures for bathroom GFI, electric heat, lighting and oven, which were reported in [5].

Table 5.1 shows the obtained TP, FP and FN values of the confusion matrix during the disaggregation of house 1 from the REDD dataset [1] using the two combined k-means and DTW methods, namely DTW+kM and kDTW.

Table 5.1: Confusion Matrix for REDD House 1 using DTW+kM and kDTW.

	TP		*FP*		*FN*	
Apps	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*
DW	0	6	0	0	32	26
F	156	156	24	24	56	56
KO	65	65	9	15	11	11
MW	61	61	13	13	21	21
Washer Dryer	46	46	2	2	22	22

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave. TP=True Positive, FP=False Positive, FN=False Negative.

Figures 5.1(a)-(b) and 5.2 show the comparison of classification accuracy using TP, FP, FN for the two proposed unsupervised NILM methods, DTW with k-means for classification (DTW+kM) and the DTW combined with k-means and post-processing refinement using DTW (kDTW), with respect to the original DTW-based method during the disaggregation of house 1 from REDD dataset [1], based on the results available in Tables 4.2 and 5.1.

According to the obtained results, the DTW+kM method was not able to classify correctly any of the dishwasher operations, even the high consuming washing cycle, as the average consumption value estimated from the algorithm was similar to the average power of the appliance present in the kitchen outlet monitor. This has resulted in higher FP values for kitchen outlet, as some of the missing dishwasher activity was misclassified as kitchen outlet. The DTW refinement was able to correct the misclassification of dishwasher, but was only able to identify the same operations as in the DTW-based method.

For the rest of the appliances both DTW+kM and the implementation with the DTW refinement were able to obtain the same classification results with the DTW-based method, presented in Chapter 4, which means that the DTW refinement was not able to further increase the classification accuracy.

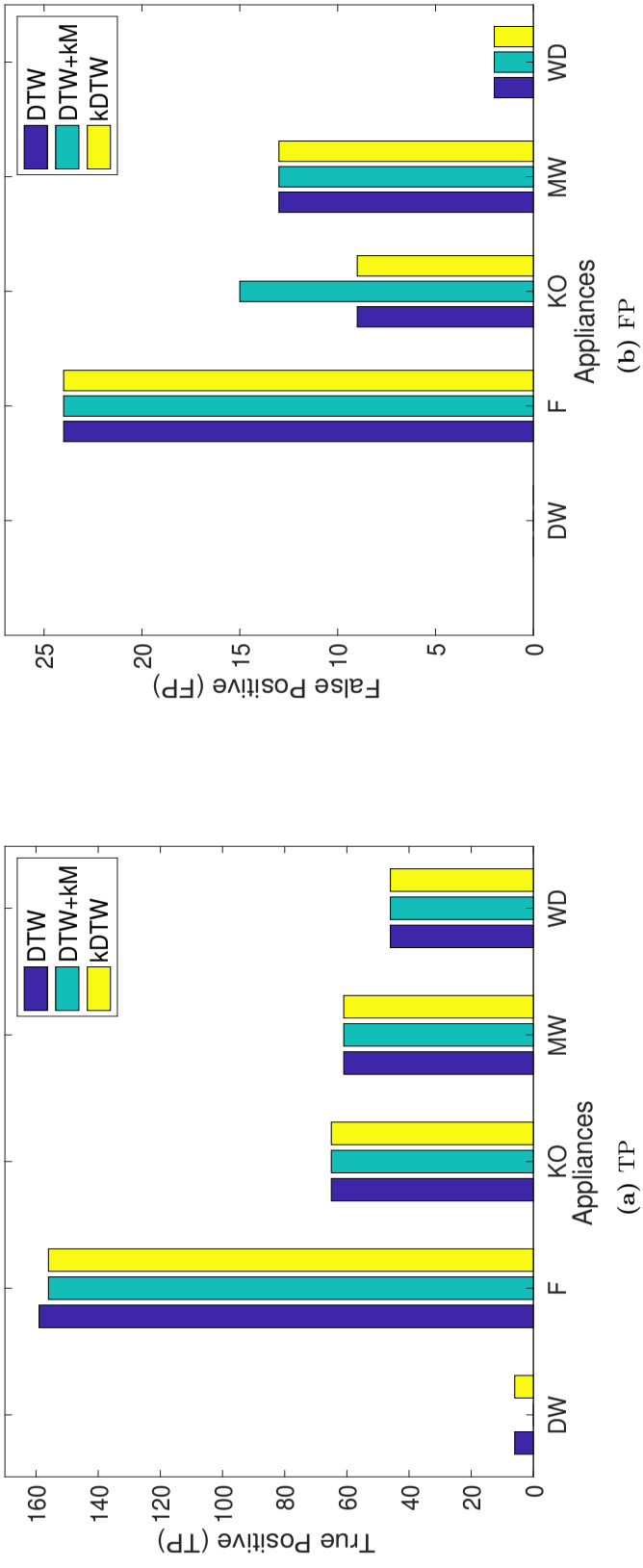


Figure 5.1: Classification accuracy using TP and FP for the proposed DTW+kM and kDTW methods, with respect the DTW-based method from Chapter 4, during the disaggregation for House 1 from REDD dataset [1], as reported in Table 4.2 and 5.1.

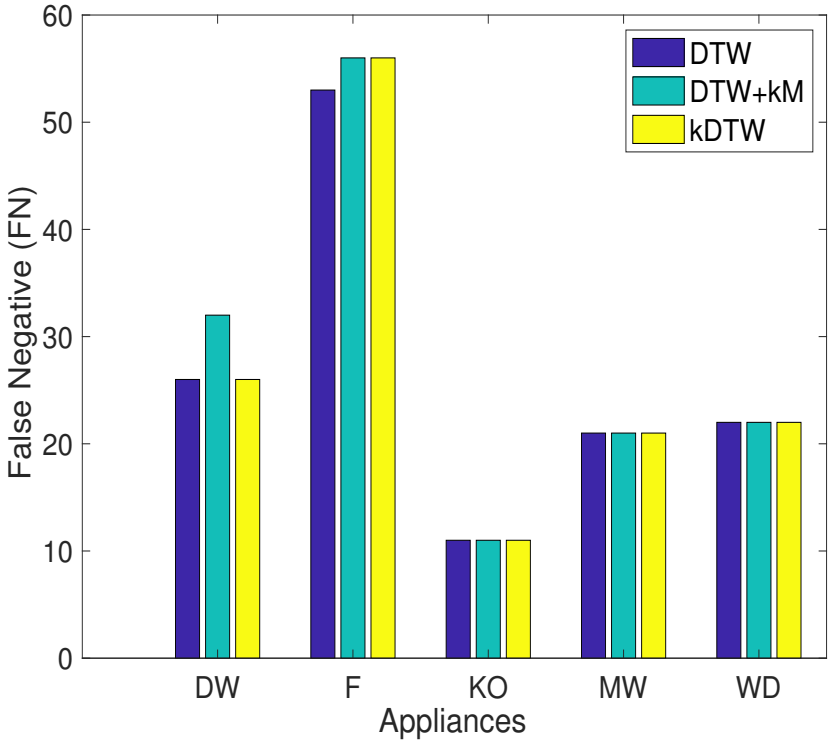


Figure 5.2: Classification accuracy using FN for the proposed DTW+kM and kDTW methods, with respect the DTW-based method from Chapter 4, during the disaggregation for House 1 from REDD dataset [1], as reported in Tables 4.2 and 5.1.

Table 5.2 shows the obtained classification accuracy for the two methods in terms of precision, recall and F_M. As the two methods were able to obtain the same disaggregation for most appliances as with the case of the DTW-based method, the average precision was more than 80% and the relevant recall was more than 70%. Even for the case of the kitchen outlet, where the DTW+kM method misclassified dishwasher operations as kitchen outlet, the obtained precision was 81% compared to the 88% obtained using the DTW method and the kDTW method, which is still more than 80%. Furthermore, the two proposed methods were not able to increase the recall accuracy reported during the DTW-based implementation, thus the obtained recall was more than 70% for most appliances, with the exception of dishwasher, which was either not disaggregated, for DTW+kM, or obtained the same recall as DTW, for the kDTW refinement. Similarly, washer dryer reported the same recall 67% for all three methods.

Table 5.2: Classification Accuracy using Precision, Recall and F_M for REDD House 1 using the proposed DTW+kM and kDTW methods.

	PR		*RE*		F_M	
Apps	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*
DW	–	1	–	0.19	–	0.32
F	0.87	0.87	0.74	0.74	0.80	0.80
KO	0.81	0.88	0.85	0.85	0.83	0.86
MW	0.82	0.82	0.74	0.74	0.78	0.78
WD	0.96	0.96	0.67	0.67	0.79	0.79

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, WD=Washer Dryer. PR=Precision, RE=Recall, F_M=F-Measure.

Table 5.3 shows the obtained classification accuracy using F_M and estimation accuracy using Acc_i of the DTW+kM and kDTW methods, with benchmarks the DTW-based method proposed in Chapter 4. Figures 5.3(a)-(b) show the same results graphically, for visual comparison.

For F_M, kitchen outlet reported 0.83 for DTW+kM method, which is only 3% less compared to the 0.86 obtained using both original DTW-based and the kDTW method, while for the rest of the appliances the corresponding F_M was the same, besides dishwasher for DTW+kM, where no classification was obtained.

In terms of estimation accuracy, Figure 5.3(b) shows that the performance of the three proposed methods is the same for most appliances, while dishwasher, as already discussed, was not disaggregated using DTW+kM, therefore no estimation accuracy was obtained. Although in terms of classification, the proposed methods were overall successful for most appliances, by observing the estimation accuracy, it is apparent that the proposed methods were less accurate for estimating the consumed power per appliance. In most cases this is the result of either overestimation or underestimation of the power and the fact that the proposed methods can not estimate standby load or the low consuming cycles for dishwasher and washer dryer.

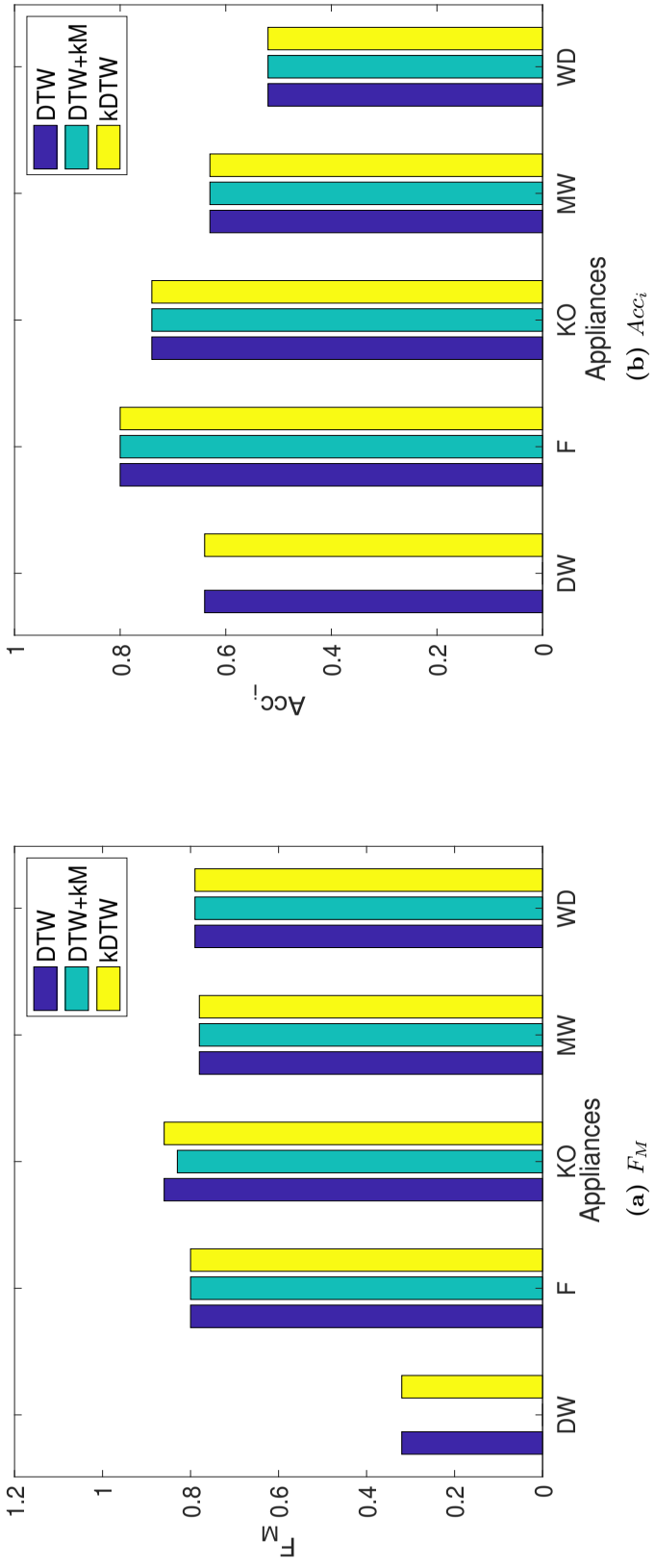
Table 5.3: Classification Accuracy using F_M and Estimation Accuracy using Acc_i of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, for REDD House 1.

	F_M			Acc_i		
Apps	*DTW*	*DTW+kM*	*DTW*	*DTW*	*DTW+kM*	*kDTW*
DW	0.32	–	0.32	0.64	–	0.64
F	0.80	0.80	0.80	0.80	0.80	0.80
KO	0.86	0.83	0.86	0.74	0.74	0.74
MW	0.78	0.78	0.78	0.63	0.63	0.63
WD	0.79	0.79	0.79	0.52	0.52	0.52

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, WD=Washer Dryer.

In Chapter 4, the normalised total power consumption estimation error (TER), which has been proposed in [5], has been included, but, as it is apparent from the obtained estimation accuracy presented in Figure 5.3(b), the TER error for most appliances would be the exact same compared to the one obtain using the DTW-based method, as reported in Table 4.5, with the exception of the DTW+kM method that was not able to classify dishwasher, and therefore estimate its power consumption.

Figures 5.4(a)-(b) show the disaggregation results using the three methods proposed in this thesis, as a percentage of consumption contribution with respect to the actual consumption contribution of the operating appliances (Figure 5.5), using the individual monitors available for house 1 in the REDD dataset [1]. Note that both DTW-based method and the proposed method using the combined k-means and DTW scheme, with the inclusion of the DTW refinement process, achieved almost identical disaggregation results, which results to the same disaggregation contribution as a percentage of the actual load. The DTW+kM, as already discussed, was not able to identify any dishwasher operations, as it was misclassified as kitchen outlet, due to the similarity of the average washing cycle power consumption and the average consumption of the kitchen outlet. Therefore the reported unknown consumption for the DTW+kM method was higher by 2% and 25% compared to the unknown load obtained by the other two methods, and the ground truth unknown consumption contribution, respectively.



(a) F_M (b) Acc_i

Figure 5.3: Performance evaluation using F_M and Acc_i for the proposed DTW+kM and kDTW (Table 5.3) with benchmarks the DTW-based method proposed in Chapter 4 for REDD House 1.

(a) Disaggregation load contribution using DTW+kM method

(b) Disaggregation load contribution using DTW-based and kDTW method

Figure 5.4: Disaggregation results, as percentage of load contribution per appliance relative to the aggregate load, for House 1 from REDD dataset using the DTW based method, proposed in Chapter 4, and the DTW+kM and kDTW methods.

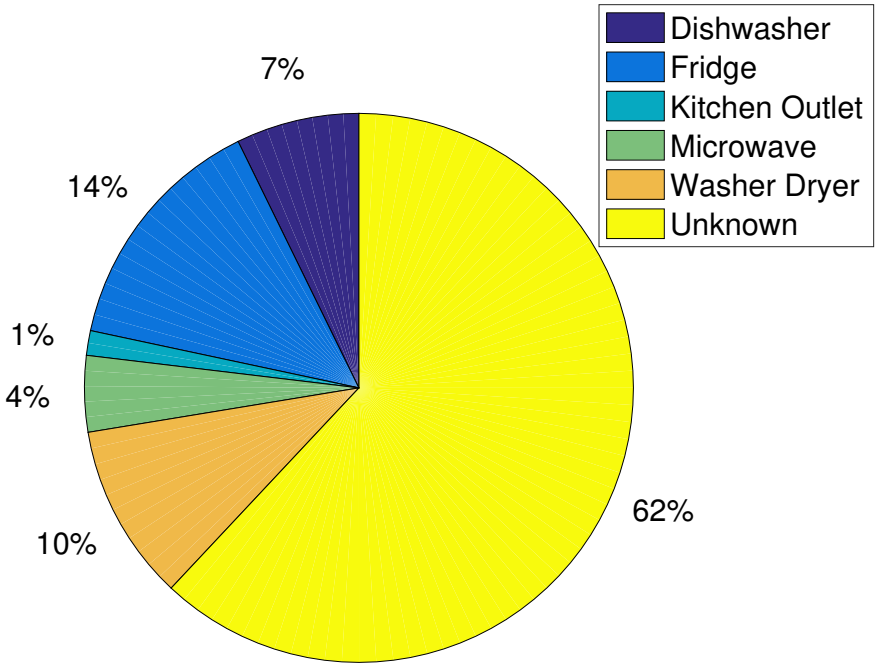


Figure 5.5: Ground truth load contribution, as percentage of load contribution per appliance relative to the aggregate load, for House 1 from REDD dataset.

To summarise, the methods proposed in this chapter were able to show similar performance to the original DTW-based method in Chapter 4 for most appliances, apart from dishwasher, where the DTW+kM method was not able to classify any dishwasher operation, which was successfully addressed when implementing the method using the DTW refinement. In general, as reported in both this chapter and Chapter 4, the proposed methods were able to report good performance in terms of classification for most of the appliance minus the dishwasher.

5.4.2.2 Disaggregation Case Study using House 2 from REDD Dataset

For the case of house 2 from REDD dataset [1], similarly with Chapter 4, the appliances that the proposed methods were able to disaggregate were the following: dishwasher, fridge, two kitchen outlets and microwave. The proposed methods were not able to disaggregate washer dryer, as the original proposed method in Chapter 4, therefore the presented results include only the appliances disaggregated using the proposed DTW-

based method for both classification and estimation accuracy.

Table 5.4 shows the obtained TP, FP and FN values of the confusion matrix during the disaggregation of house 2 from the REDD dataset [1] using the two combined k-means and DTW methods, namely DTW+kM and kDTW.

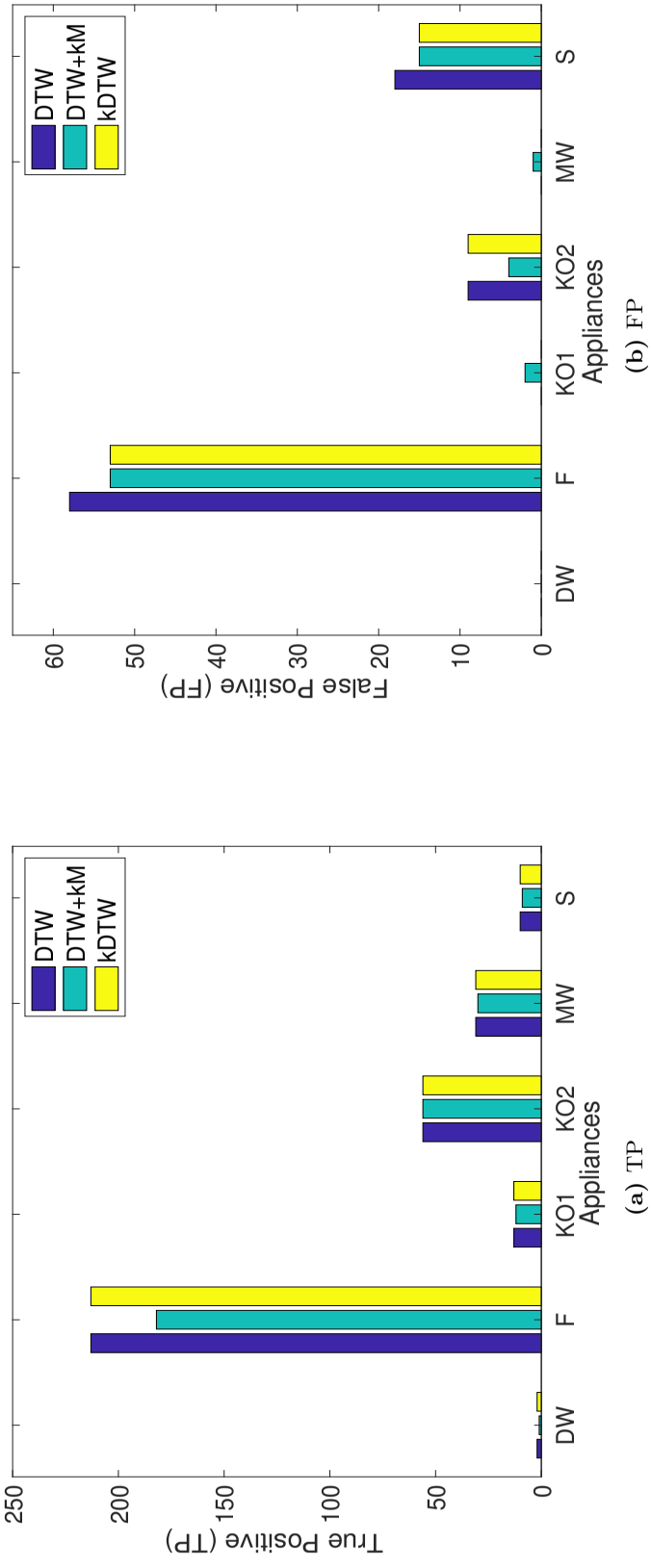
Table 5.4: Confusion Matrix for REDD House 2 using DTW+kM and kDTW.

	TP		*FP*		*FN*	
Apps	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*
DW	1	2	0	0	8	7
F	182	213	53	53	67	36
KO1	12	13	2	0	6	5
KO2	56	56	4	4	20	20
MW	30	31	1	0	8	7
S	9	10	15	15	1	0

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, S=Stove. TP=True Positive, FP=False Positive, FN=False Negative.

Figures 5.6(a)-(b) and 5.7 present the classification results obtained using TP, FP, FN for the two proposed unsupervised NILM methods, DTW with k-means for classification (DTW+kM) and the DTW combined with k-means post-processing refinement using DTW (kDTW) during the disaggregation of house 2 from REDD dataset [1], based on the results available in Tables 4.6 and 5.4.

In terms of correctly identified appliances, the best performance was obtained using kDTW, which includes the DTW refinement, and the original DTW-based method. During the classification of the fridge, the DTW+kM method was able to correctly identify $\sim 15\%$ less true operations compared to those identified using the DTW-based method. kDTW was able to improve the performance in terms of correctly classified events, and was equivalent to the performance of the DTW-based method presented in Chapter 4. Both DTW+kM and kDTW were able to report lower FP values for fridge, kitchen outlet 2 and stove.



(a) TP (b) FP

Figure 5.6: Classification accuracy using TP and FP for the proposed DTW+kM and kDTW methods, with respect the DTW-based method from Chapter 4, during the disaggregation for House 2 from REDD dataset [1], as reported in Table 4.6 and 5.4.

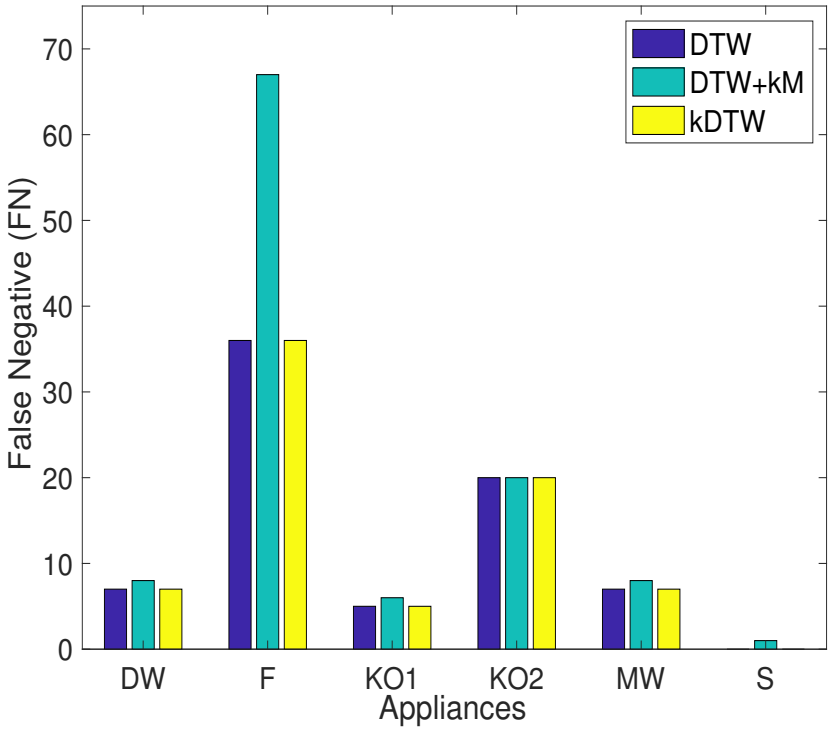


Figure 5.7: Classification accuracy using FN for the proposed DTW+kM and kDTW methods, with respect the DTW-based method from Chapter 4, during the disaggregation for House 2 from REDD dataset [1], as reported in Table 4.6 and 5.4.

Table 5.5 shows the obtained classification accuracy for the DTW+kM and kDTW methods, in terms of precision, recall and F_M. In terms of precision, the two proposed methods were able to report on average more than 80%, with the exception of the stove, were DTW+kM obtained 37% and kDTW 40%, which is slightly better than 36%, which was reported for the proposed DTW-based method in Chapter 4, as shown in Table 4.6. The proposed methods of this chapter have shown higher precision for kitchen outlet 2 at 93% compared to the 86% for the original proposed method in Chapter 4.

Table 5.6 shows the obtained classification accuracy using F_M and estimation accuracy using Acc_i of the DTW+kM and kDTW methods, with benchmarks the DTW-based method proposed in Chapter 4. Figures 5.8(a)-(b) present the obtained results graphically, for visual comparison.

Table 5.5: Classification Accuracy using Precision, Recall and F_M for REDD House 2 using the proposed DTW+kM and kDTW methods

	PR		*RE*		F_M	
Apps	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*
DW	1	1	0.11	0.22	0.20	0.36
F	0.77	0.80	0.73	0.85	0.75	0.82
KO1	0.86	1	0.67	0.72	0.75	0.84
KO2	0.93	0.93	0.74	0.74	0.82	0.82
MW	0.97	1	0.79	0.82	0.87	0.90
S	0.37	0.40	0.90	1	0.52	0.57

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, S=Stove. PR=Precision, RE=Recall, $F_M = F - Measure$.

The DTW+kM method reported the lowest classification accuracy compared to the other two methods, with the exception of the kitchen outlet 2, where it was able to outperform the original DTW-based method, obtaining F_M of 0.82. In general, the kDTW method, which uses the DTW refinement, was able to improve the F_M obtained using the DTW+kM method for most appliances, with the exception of the kitchen outlet 2, where it reported the same accuracy.

The refinement method has shown similar or slightly better performance compared to the original method. More specifically, the refinement method was able to increase F-measure, for both kitchen outlet 2 and stove, by 3% and 4% respectively, compared to the method proposed in Chapter 4.

In terms of estimation accuracy Acc_i, as seen in Figure 5.8(b), the DTW+kM method was able to obtain the same accuracy reported from the DTW-based method in Chapter 4, for most appliances, with the exception of the dishwasher and the stove. For these appliances, this approach reported 0.56 and 0.73 respective, 5% and 2% less compared to kDTW and the original DTW-based method. The DTW refinement was able to improve this performance and obtained the same estimation accuracy as the DTW-based method. In general, it is apparent, that the kDTW method, was not able to further improve the performance of the DTW+kM method, and outperform the estimation accuracy reported by the method presented in Chapter 4.

Table 5.6: Classification Accuracy using F_M and Estimation Accuracy using Acc_i of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, for REDD House 2.

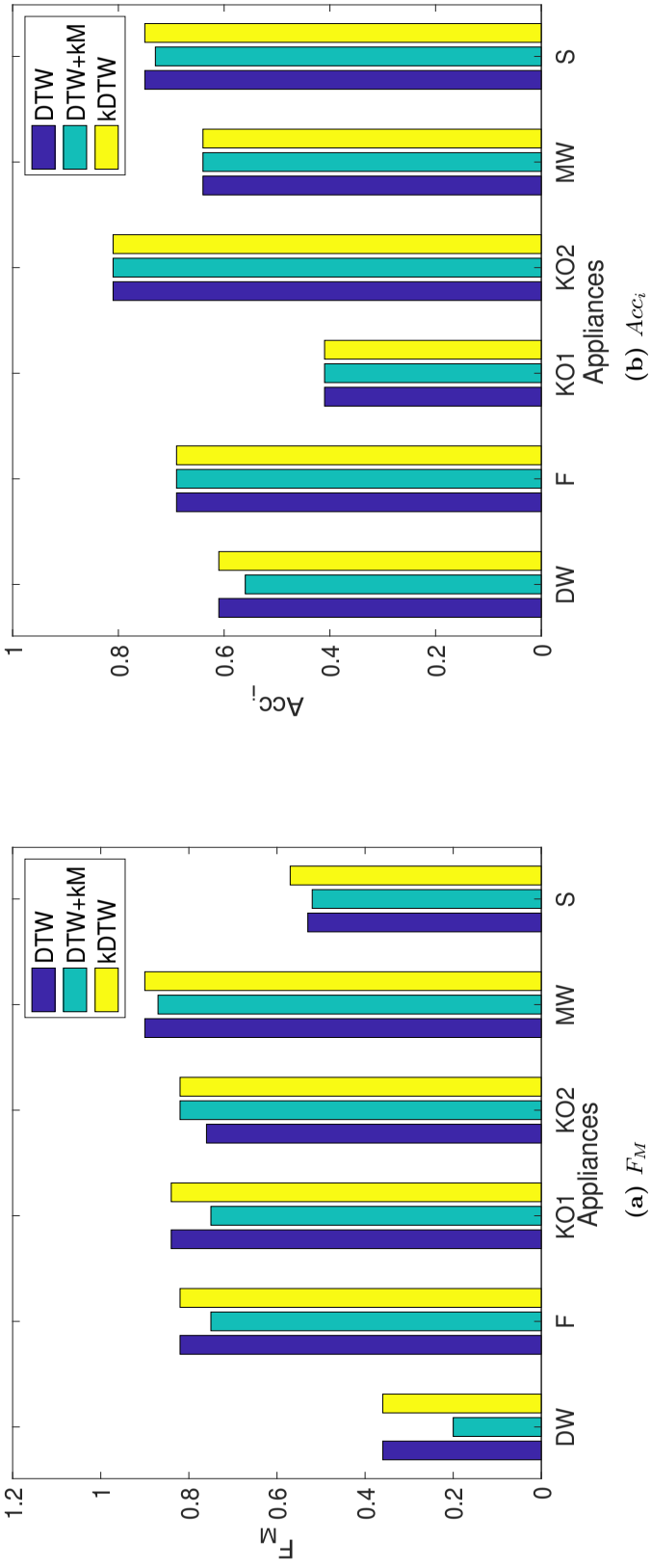
	F_M			Acc_i		
Apps	*DTW*	*DTW+kM*	*DTW*	*DTW*	*DTW+kM*	*kDTW*
DW	0.36	0.20	0.36	0.61	0.56	0.61
F	0.82	0.75	0.82	0.69	0.69	0.69
KO1	0.84	0.75	0.84	0.41	0.41	0.41
KO2	0.79	0.82	0.82	0.81	0.81	0.81
MW	0.90	0.87	0.90	0.64	0.64	0.64
S	0.53	0.52	0.57	0.75	0.73	0.75

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, S=Stove.

Figures 5.9(a)-(b) show the disaggregation results using the three proposed methods as a percentage of consumption contribution in comparison to the ground truth consumption contribution of the disaggregated appliance, shown in Figure 5.10.

Note that both DTW-based method and kDTW method, which combines k-means and DTW scheme, and performs a DTW refinement process, achieved the same disaggregation results. The performance of the proposed DTW+kM is similar to the proposed DTW-based method and the kDTW approach, in terms of the estimation of the consumption contribution per appliance. Although, the DTW+kM method reported higher number of missing fridge operations, the estimated contribution was only $\sim 1\%$ less compared to the corresponding contribution of the other two methods. The contribution of the unknown load was higher at 59%, only 2% more than the other methods, whereas the ground truth consumption of the unknown was 42%.

In summary, the DTW refinement was able to improve the classification performance of the DTW+kM method for most appliances, and reported equivalent accuracy to the proposed DTW-based method in Chapter 4. In terms of estimation accuracy, all methods reported similar performance for most appliances. The kDTW method have proven effective only for the case of the dishwasher and the stove, where it was able to improve the performance of the DTW+kM approach.



(a) F_M (b) Acc_i

Figure 5.8: Performance evaluation using F_M and Acc_i for the proposed DTW+kM and kDTW (Table 5.6) with benchmarks the DTW-based method proposed in Chapter 4 for REDD House 2.

(a) Disaggregation load contribution using DTW+kM method

(b) Disaggregation load contribution using DTW-based and kDTW method

Figure 5.9: Disaggregation results, as percentage of load contribution per appliance relative to the aggregate load, for House 2 from REDD dataset using the DTW based method, proposed in Chapter 4, and the DTW+kM and kDTW methods.

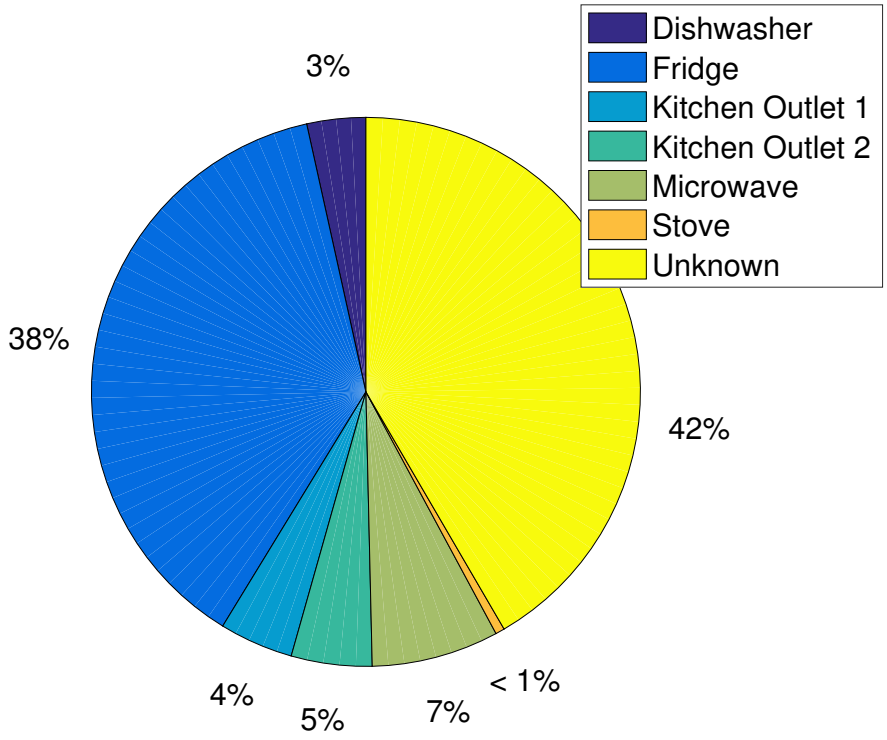


Figure 5.10: Ground truth load contribution, as percentage of load contribution per appliance relative to the aggregate load, for House 2 from REDD dataset.

5.4.2.3 Disaggregation Case Study using House 6 from REDD Dataset

For the case of house 6 from REDD dataset [1], similarly with Chapter 4, the appliances that the proposed methods were able to disaggregate were the following: air conditioning, bathroom GFI, fridge, kitchen outlet and stove. The methods proposed in this chapter, similarly to the original DTW-based method in Chapter 4, were not able to obtain signatures and therefore classify electronics, electric heater and the reported kitchen outlet in [4, 33].

Table 5.7 shows the obtained TP, FP and FN values of the confusion matrix during the disaggregation of house 6 from the REDD dataset [1] using the two combined k-means and DTW based, namely DTW+kM and kDTW.

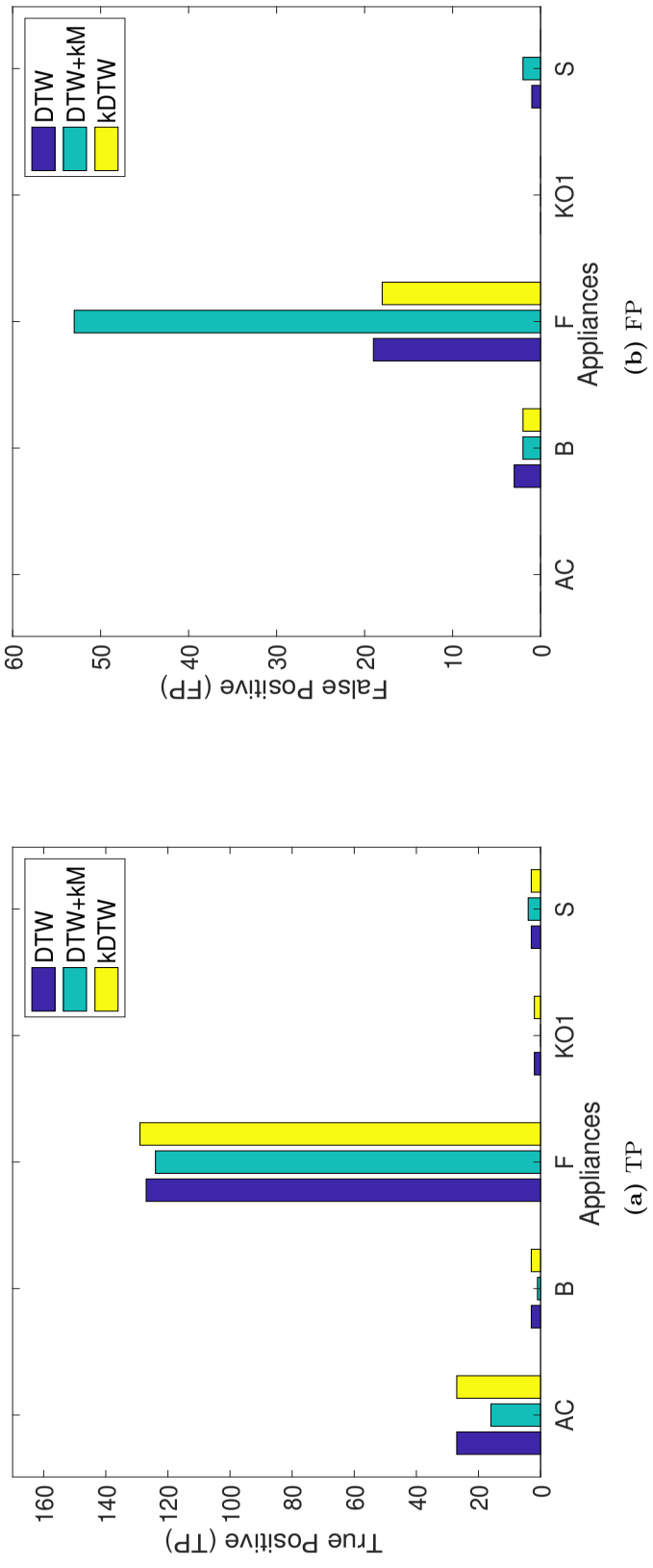
Table 5.7: Confusion Matrix for REDD House 6 using DTW+kM and kDTW.

	TP		*FP*		*FN*	
Apps	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*
AC	16	27	0	0	15	4
B	1	3	2	2	2	0
F	124	129	53	18	51	46
KO1	0	2	0	0	2	0
S	4	5	2	0	1	0

Note: Apps=Appliances, AC= Air Conditioning, B=Bathroom GFI, F=Fridge, KO=Kitchen Outlet, S=Stove. TP=True Positive, FP=False Positive, FN=False Negative.

Figures 5.11(a)-(b) and 5.12 show the classification results obtained using TP, FP, FN for the two proposed unsupervised NILM methods, DTW with k-means for classification (DTW+kM) and the DTW combined with k-means post-processing refinement using DTW (kDTW) during the disaggregation of house 6 from REDD dataset [1], based on the results available in Tables 4.9 and 5.7.

In terms of correctly identified appliances, the best performance was obtained using kDTW, which includes the DTW refinement, and the original DTW-based method. During the classification of the AC, the DTW+kM method was able to correctly identify $\sim 40\%$ less true operations compared to those identified using the DTW-based method. kDTW was able to improve the performance of the DTW+kM method for all disaggregated appliances in terms of true positive values, and generally performed equivalently to the original DTW-proposed method, with the exception of fridge, where it was able to obtain higher TP values from the original DTW method. The DTW+kM method reported higher falsely classified events for the case of the fridge, compared to the other methods and it was not able to correctly identify any operation for the kitchen outlet 1.



(a) TP (b) FP

Figure 5.11: Classification accuracy using TP and FP for the proposed DTW+kM and kDTW methods, with respect the DTW-based method from Chapter 4, during the disaggregation for House 6 from REDD dataset [1], as reported in Table 4.9 and 5.7.

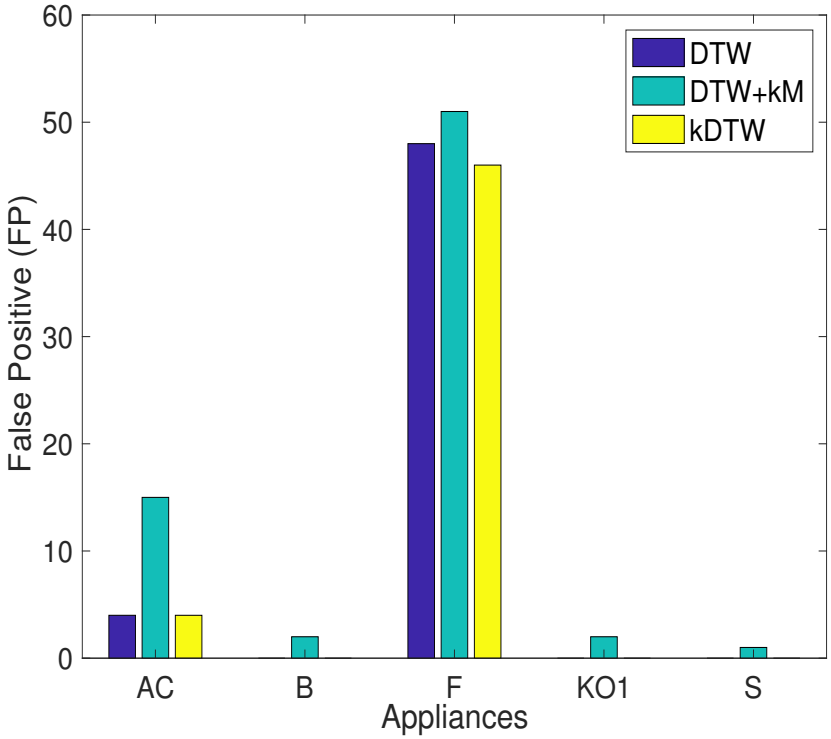


Figure 5.12: Classification accuracy using FN for the proposed DTW+kM and kDTW methods, with respect the DTW-based method from Chapter 4, during the disaggregation for House 6 from REDD dataset [1], as reported in Table 4.9 and 5.7.

Table 5.8 shows the obtained classification accuracy for the DTW+kM and kDTW methods, in terms of precision, recall and F_M.

In terms of precision, the kDTW method was able to obtain on average more than 90%, with the exception of the bathroom GFI, where the reported precision accuracy was 60%, while the DTW+kM method was only able to obtain 33% for the same appliance. For the rest of the appliances, DTW+kM obtained more than $\sim$ 70% for most appliances. Similarly, the kDTW has shown an improved performance with regards to recall accuracy compared to the implementation without the refinement, and was able to acquire more than 80% for most appliances, with the lowest recall being at 0.74 for the fridge.

Table 5.8: Classification Accuracy using Precision, Recall and F_M for REDD House 6 using the proposed DTW+kM and kDTW methods.

	PR		*RE*		F_M	
Apps	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*
AC	1	1	0.51	0.87	0.67	0.93
B	0.33	0.60	0.33	1	0.33	0.75
F	0.70	0.88	0.71	0.74	0.70	0.80
KO1	–	1	–	1	–	1
S	0.67	1	0.80	1	0.73	1

Note: Apps=Appliances, AC= Air Conditioning, B=Bathroom GFI, F=Fridge, KO=Kitchen Outlet, S=Stove. PR=Precision, RE=Recall, F_M=F-Measure.

Table 5.9 shows the obtained classification accuracy using F_M and estimation accuracy using Acc_i of the DTW+kM and kDTW methods, with benchmarks the DTW-based method proposed in Chapter 4. Figures 5.13(a)-(b) present the obtained results graphically, for visual comparison.

In general, the refinement method was able to improve the F_M for all the disaggregated appliances, compared to the combined DTW and k-means method. Both the original DTW-based and the kDTW methods were able to obtain on average classification accuracy of more than 0.8, with the exception of the bathroom GFI, where the kDTW obtained 0.75, outperforming both the DTW-based and the DTW+kM method, which reported 0.67 and 0.33 respectively. Furthermore, the refinement method was able to obtain better classification accuracy, compared to the original DTW-based approach, for both fridge and stove, where it reported F_M of 0.8 and 1 respectively. For the same appliances, the method, proposed in Chapter 4, was able to obtain 0.79 and 0.91 respectively, which is lower by 1% and 9%.

Figure 5.13(b) shows the obtained estimation accuracy per appliance of the proposed DTW+kM and kDTW methods in comparison to those results reported for the DTW-based method in Chapter 4. It is apparent that the original DTW-based method and the kDTW method were able to outperform the implementation of the combined DTW and k-means scheme for all appliances, and they both reported the same estimation accuracy.

Table 5.9: Classification Accuracy using F_M and Estimation Accuracy using Acc_i of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, for REDD House 6.

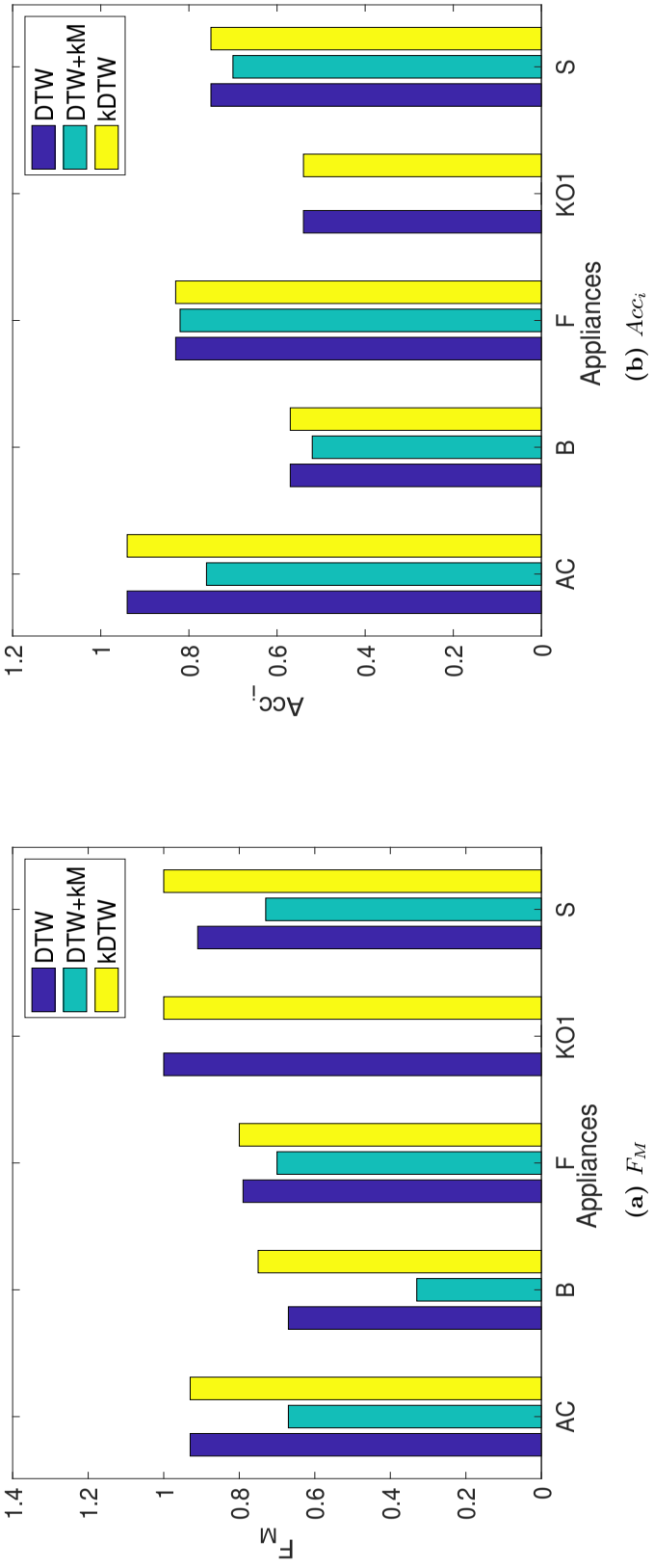
	F_M			Acc_i		
Apps	*DTW*	*DTW+kM*	*DTW*	*DTW*	*DTW+kM*	*kDTW*
AC	0.93	0.67	0.93	0.94	0.76	0.94
B	0.67	0.33	0.75	0.57	0.52	0.57
F	0.79	0.70	0.80	0.83	0.82	0.83
KO1	1	–	1	0.54	–	0.54
S	0.91	0.73	1	0.75	0.70	0.75

Note: Apps=Appliances, AC= Air Conditioning, B=Bathroom GFI, F=Fridge, KO=Kitchen Outlet, S=Stove.

The DTW refinement method was able to improve the performance of the DTW+kM method by almost 20% for the air conditioning and by less than 5% for the other appliances. For kitchen outlet 1, both kDTW and the original DTW-based method were able to estimate the consumption of the appliance.

Figures 5.14(a)-(b) present the disaggregation results of the proposed DTW+kM and kDTW methods, as percentage of load contribution per appliance with regards to the aggregate load, the corresponding result obtained using the DTW-based method from Chapter 4 and the relevant ground truth consumption per appliances (Figure 5.15), using the data from the IAMs as presented in the REDD dataset [1]. Both kDTW and the original DTW-based method obtained the same results in terms of percentage load contribution and thus Figure 5.14(b) is used for both cases.

According to Figure 5.14(a), the DTW+kM method shows only 21% consumption contribution for air conditioning, which is more than 10% less than the actual contribution, which is reported at 33%. The other methods overestimated the consumption of air conditioning by by 3%. Furthermore, DTW+kM estimated the contribution of the unknown load at 69%, which is more than 15% from the consumption estimation obtained using the other methods, and 20% higher than the ground truth contribution of the unknown appliances.



(a) F_M

(b) Acc_i

Figure 5.13: Performance evaluation using F_M and Acc_i for the proposed DTW+kM and kDTW (Table 5.9) with benchmarks the DTW-based method proposed in Chapter 4 for REDD House 6.

(a) Disaggregation load contribution using DTW+kM method

(b) Disaggregation load contribution using DTW-based and kDTW method

Figure 5.14: Disaggregation results, as percentage of load contribution per appliance relative to the aggregate load, for House 6 from REDD dataset using the DTW based method, proposed in Chapter 4, and the DTW+kM and kDTW methods.

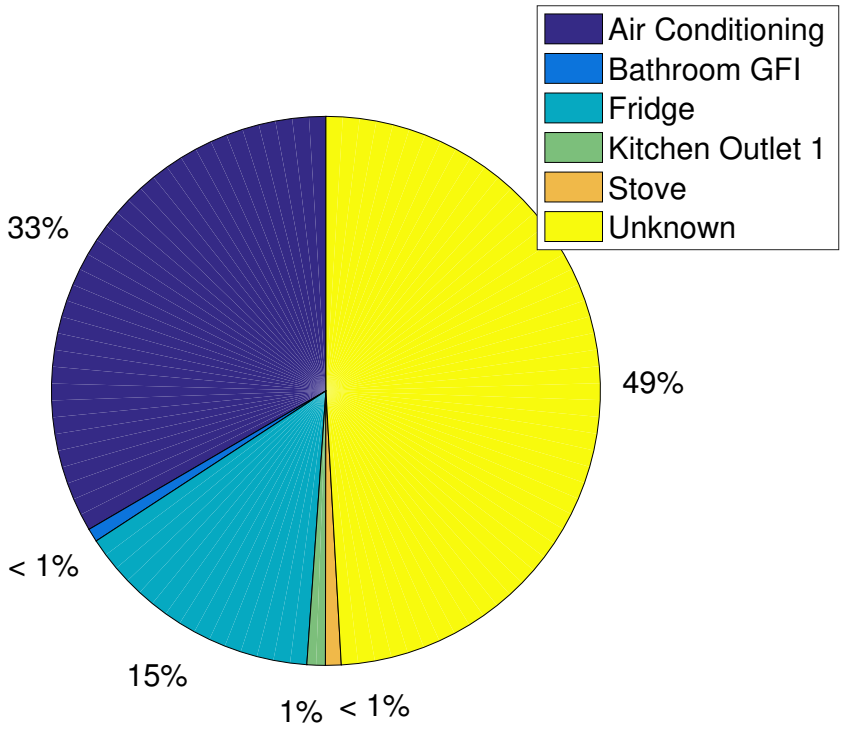


Figure 5.15: Ground truth load contribution, as percentage of load contribution per appliance relative to the aggregate load, for House 6 from REDD dataset.

In general, during the disaggregation of house 6 from REDD dataset [1], it was observed that the DTW refinement method was able to improve for most appliances both classification and estimation accuracy, performing on average better or at least equivalent to the proposed method presented in Chapter 4.

5.4.3 Performance Evaluation using REFIT Dataset

The proposed DTW+kM and kDTW methods have been used for disaggregation of houses 2 and 17 from the same clean version of the REFIT dataset [2, 3], without removal of base load or noise, similarly with Chapter 4.

5.4.3.1 Disaggregation Case Study using House 2 from REFIT Dataset

Similarly with Chapter 4, the same appliances were disaggregated using the proposed DTW+kM and kDTW methods, namely dishwasher, fridge-freezer, kettle, microwave, toaster and washing machine.

Table 5.10 shows the obtained TP, FP and FN values of the confusion matrix during the disaggregation of house 2 from the REFIT dataset [2, 3] using the two k-means and DTW methods, namely DTW+kM and kDTW.

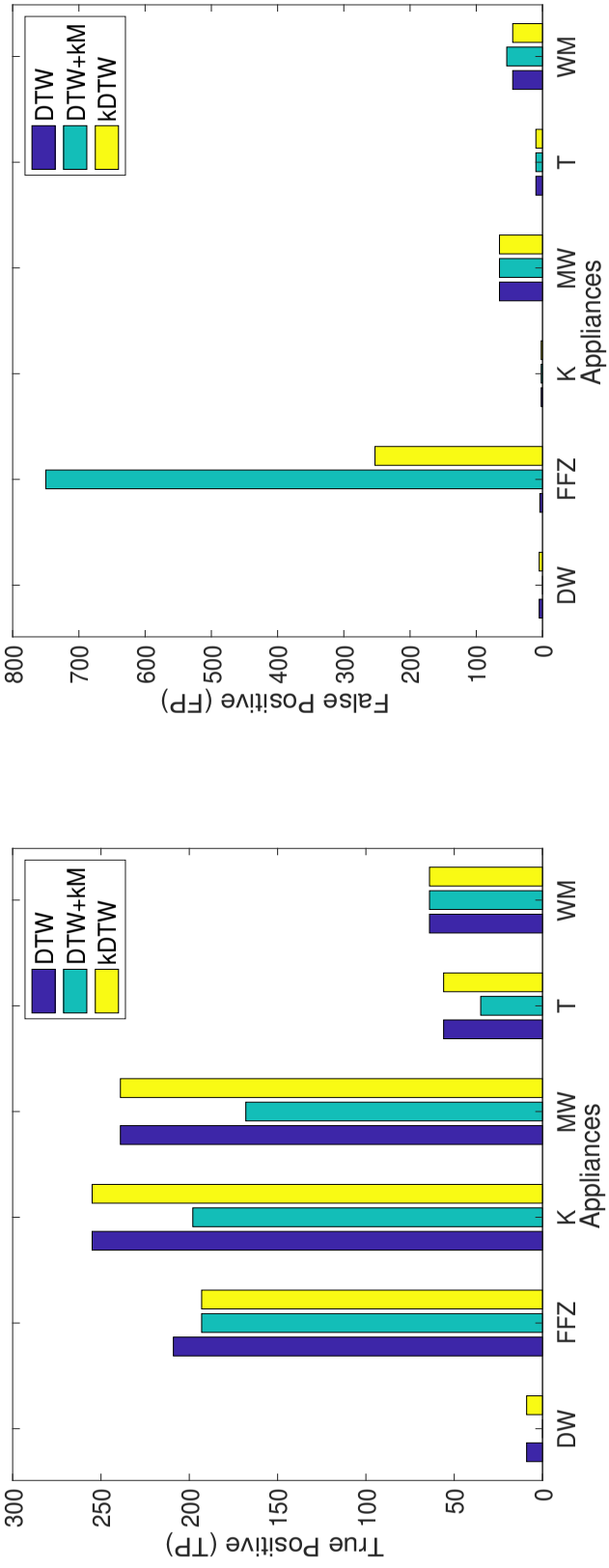
Table 5.10: Confusion Matrix for REFIT House 2 using DTW+kM and kDTW.

	TP		FP		FN	
Apps	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*
DW	0	9	0	5	44	35
FFZ	193	193	750	253	286	286
K	198	255	2	2	74	17
MW	168	239	65	65	80	19
T	35	56	10	10	33	12
WM	64	64	54	45	12	12

Note: Apps=Appliances, DW=Dishwasher, FFZ=Fridge-Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine. TP=True Positive, FP=False Positive, FN=False Negative.

Figures 5.16(a)-(b) and 5.17 show the obtained classification accuracy with regards to TP, FP, FN from the confusion matrix, based on the results presented in Tables 4.11 and 5.10. First and foremost, it can be observed that the DTW+kM method was not able to classify correctly any dishwasher operation, while the kDTW method was able to obtain equivalent results with the DTW-based method, as found in 4.11.

As already discussed in Chapter 4, dishwasher and washing machine have a similar average power consumption for their washing cycle and as the k-means clustering was implemented for clustering the average power of each detected event, for acquiring low computational cost, it is no surprise that the method failed to distinguish the two appliances. With the incorporation of the DTW refinement, the method was able to obtain the same true events as in the case of the DTW-based method.



(a) TP (b) FP

Figure 5.16: Classification accuracy using TP and FP for the proposed DTW+kM and kDTW methods, with respect the DTW-based method from Chapter 4, during the disaggregation House 2 from REFIT dataset [2, 3], as reported in Table 4.11 and 5.10.

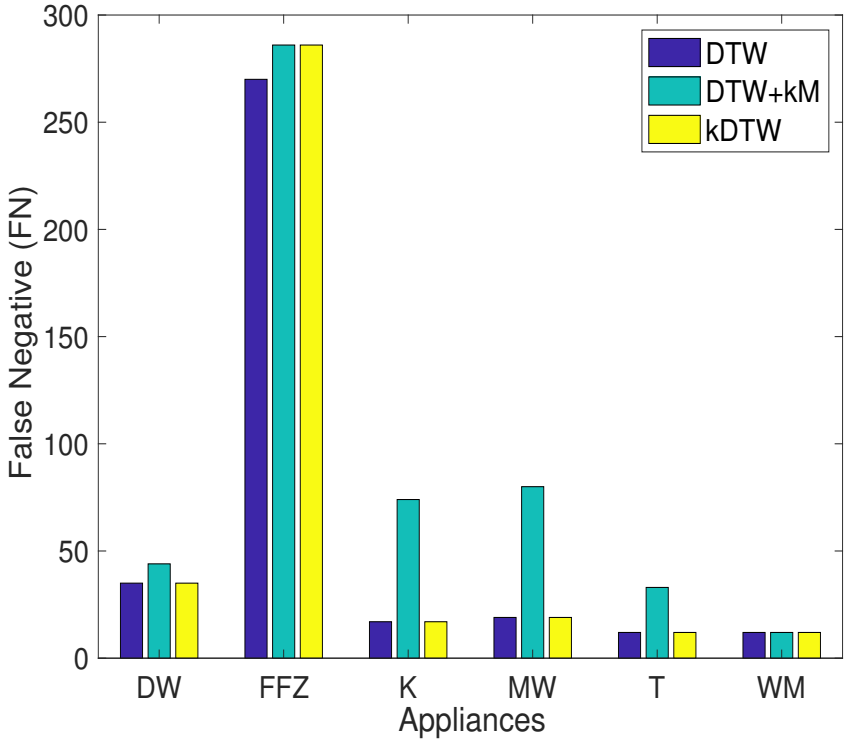


Figure 5.17: Classification accuracy using FN for the proposed DTW+kM and kDTW methods, with respect the DTW-based method from Chapter 4, during the disaggregation for House 2 from REFIT dataset [2, 3], as reported in Table 4.11 and 5.10.

In general, the refinement method was able to improve the correctly identified operations for most appliance, and performed comparable to the original DTW method. False positive values were similar for most appliances across the three methods, with the exception of fridge-freezer, where both DTW+kM and kDTW reported high values. As the two proposed methods did not perform iterative removal of the already detected appliance events during event detection, it has been observed that there was a plethora of rising and falling edges that were occurring during the operation of higher consuming appliances, that were falsely identified as fridge-freezer. During the original implementation those instances were absorbed as a load variation of those appliances, and removed from the load. The refinement method was able to reduce significantly the number of FP values, but the reported number was still much higher compared to the DTW-based method, found in Table 4.11.

Table 5.11: Classification Accuracy using Precision, Recall and F_M for REFIT House 2 using the proposed DTW+kM and kDTW methods.

	PR		*RE*		F_M	
Apps	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*
DW	–	0.41	–	0.2	–	0.27
FFZ	0.2	0.43	0.4	0.4	0.27	0.41
K	0.99	0.99	0.73	0.94	0.84	0.96
MW	0.72	0.79	0.71	0.93	0.71	0.85
T	0.78	0.85	0.51	0.82	0.62	0.83
WM	0.54	0.59	0.84	0.84	0.66	0.69

Note: Apps=Appliances, DW=Dishwasher, FFZ=Fridge-Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine. PR=Precision, RE=Recall, F_M=F-Measure.

As a result, both methods obtained low precision accuracy for fridge-freezer, as reported in Table 5.11, which presents the classification results for the two methods with respect to precision, recall and F_M. The refinement method was able to increase precision by 20%, but still poor compared to the 98% obtained using the DTW-based method. The overall precision for the rest of the appliances was comparable between the methods, and dishwasher did not report any classification accuracy results for the DTW+kM method, as it was not classified.

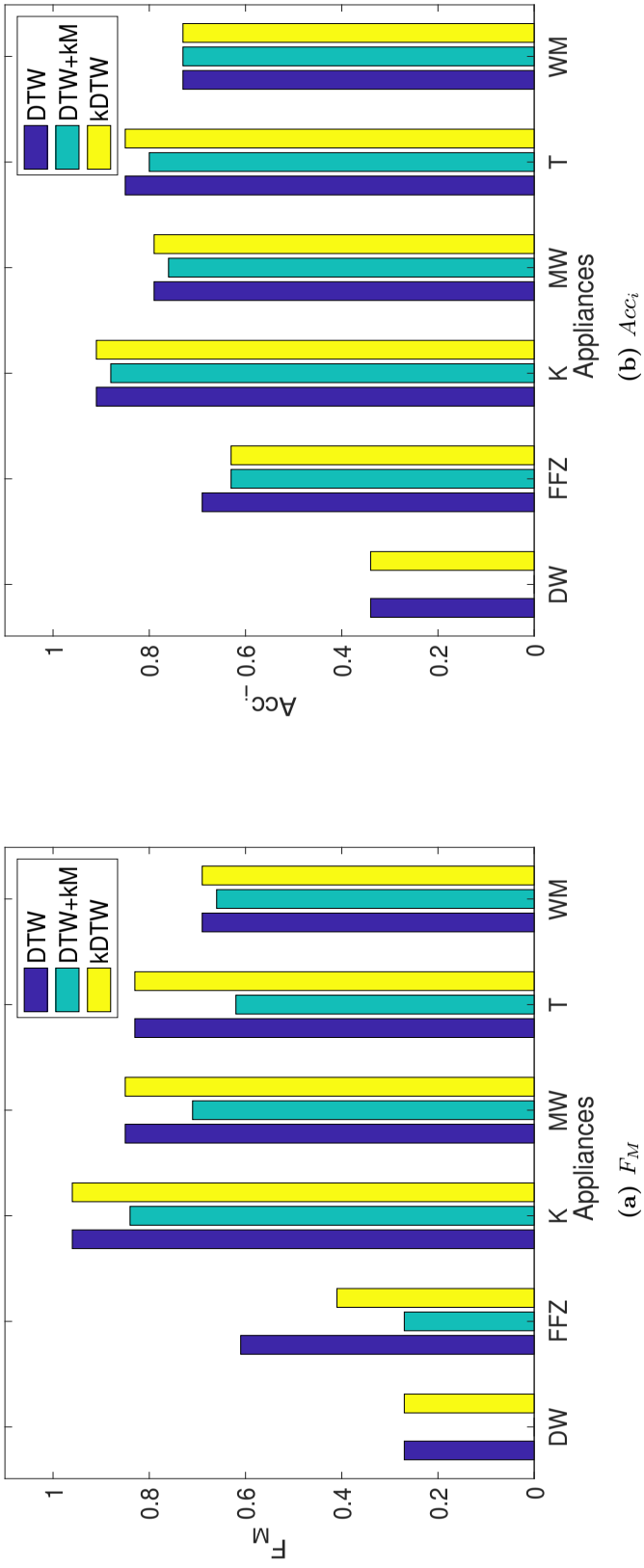
In terms of recall accuracy, both kDTW and DTW-based methods were able to report more than 80% for most appliances, but as expected much lower values for the dishwasher and fridge-freezer, where a substantial number of operations were not classified correctly.

Table 5.12 shows the classification accuracy using F_M, estimation accuracy using Acc_i and TER (normalised total power consumption estimation error) of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4. Figures 5.18(a)-(b) show graphically the classification and estimation accuracy, using F_M and Acc_i respectively.

Table 5.12: Classification Accuracy using F_M, Estimation Accuracy using Acc_i and TER of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, for REFIT House 2.

	F_M			Acc_i			TER		
Apps	*DTW*	*DTW+kM*	*DTW*	*DTW*	*DTW+kM*	*kDTW*	*DTW*	*DTW+kM*	*kDTW*
DW	0.27	–	0.27	0.34	–	0.34	0.43	–	0.43
FFZ	0.61	0.27	0.41	0.69	0.63	0.63	0.70	0.73	0.73
K	0.96	0.84	0.96	0.91	0.88	0.91	0.019	0.031	0.019
MW	0.85	0.71	0.85	0.79	0.76	0.79	0.09	0.09	0.09
T	0.83	0.62	0.83	0.85	0.80	0.85	0.21	0.33	0.21
WM	0.69	0.66	0.69	0.73	0.73	0.73	0.42	0.42	0.42

Note: Apps=Appliances, DW=Dishwasher, FFZ=Fridge-Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine.



(a) F_M **(b)** Acc_i

Figure 5.18: Performance evaluation using F_M and Acc_i for the proposed DTW+kM and kDTW (Table 5.12) with benchmarks the DTW-based method proposed in Chapter 4 for REFIT House 2.

With respect to the classification accuracy F_M, the DTW+kM approach reported overall the lowest accuracy for all appliances and was not able to disaggregate the dishwasher, as already discussed. The DTW refinement method was able improve the performance, and reported the same classification accuracy as the original DTW-based method for most appliances. The kDTW approach was outperformed by the DTW-based method for the case of the fridge-freezer, where the former obtained 0.41 and the latter 0.61.

Similarly with the classification accuracy, the DTW+kM method reported the lowest estimation accuracy for most appliances, as seen in Figure 5.18(b), with the exception of the washing machine, where it reported Acc_i of 0.73, the same as the other two methods. The kDTW method was able to improve the performance reported by the DTW+kM approach for most appliances, apart from the fridge-freezer and the washing machine, where they reported the same accuracy. While the DTW refinement was able to report higher estimation accuracy compared to the DTW+kM method, it was not able to outperform the original DTW-based method. For the majority of the appliances, both approaches obtained the same accuracy, while for the fridge-freezer the DTW-based method outperformed the refinement approach by 6%.

Furthermore, Figure 5.19 reports the evaluation of the proposed method of this chapter with benchmarks the DTW-based method proposed in Chapter 4, using normalised total power consumption estimation error (TER), proposed in [5]. The proposed kDTW and DTW methods from Chapter 4, reported the lowest TER estimation error for the kettle, microwave and toaster, where the methods have shown high estimation accuracy per appliance.

The kDTW approach was able to reduce the TER error for most appliances compared to the DTW+kM method, with the exception of the fridge-freezer and the washing machine, where both methods reported error of 0.73 and 0.42 respectively.

In general, the DTW refinement reported either the same or higher estimation error compared to the DTW-based method, therefore it was not able to increase the overall performance, which has been apparent through out this case study.

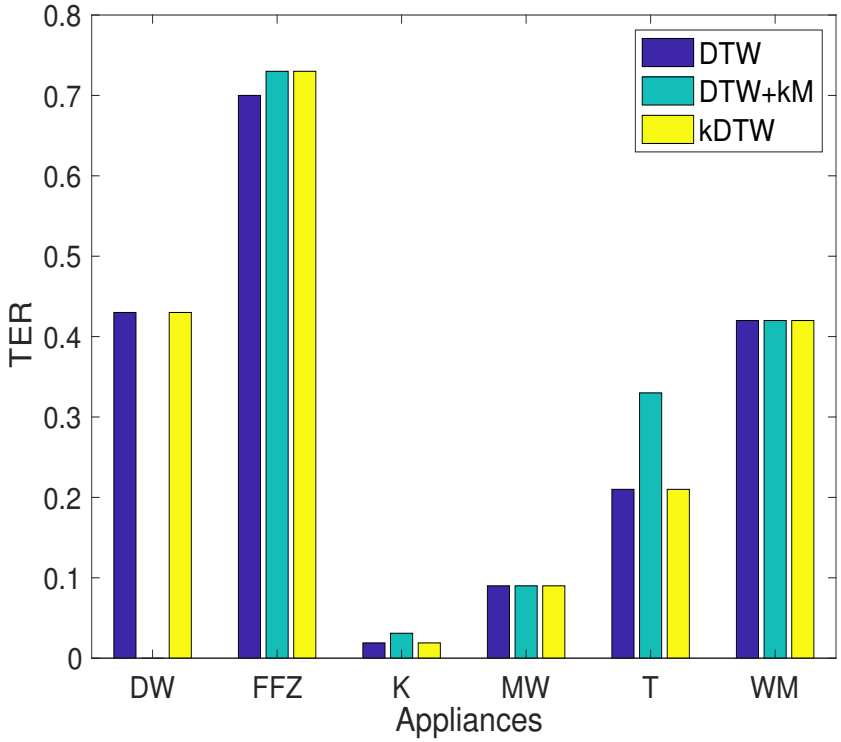


Figure 5.19: Performance evaluation using TER for the proposed DTW+kM and kDTW, with benchmarks the DTW-based method proposed in Chapter 4 for REFIT House 2.

Figures 5.20(a)-(b) show the disaggregation results of the proposed DTW+kM and kDTW NILM methods as percentage of load contribution per appliance with regards to the aggregate load and the relevant ground truth consumption per appliances (Figure 5.21), available from the IAMs of the REFIT dataset [2, 3].

The kDTW method obtained equivalent disaggregation results with the original DTW-based method, thus both are presented using Figure 5.20(b). As DTW+kM was not able to classify and therefore estimate the consumption contribution of dishwasher, it has reported unknown load contribution of 89%, which is 8% and 20% higher than the estimated contribution for the other methods and the ground truth contribution.

(a) Disaggregation load contribution using DTW+kM method

(b) Disaggregation load contribution using DTW-based and kDTW method

Figure 5.20: Disaggregation results, as percentage of load contribution per appliance relative to the aggregate load, for House 2 from REFIT dataset using the DTW based method, proposed in Chapter 4, and the DTW+kM and kDTW methods.

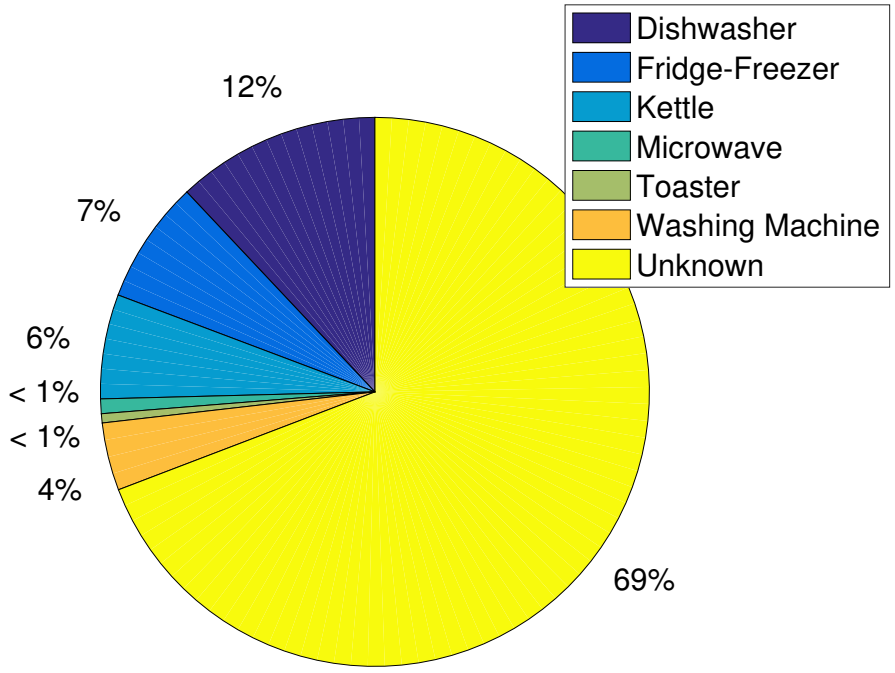


Figure 5.21: Ground truth load contribution, as percentage of load contribution per appliance relative to the aggregate load, for House 2 from REFIT dataset.

In conclusion, the proposed DTW refinement was able to improve the performance of the DTW+kM method, in terms of classification and estimation accuracy, and reduce the normalised total estimation error for most appliances, and reported equivalent results with the proposed method in Chapter 4. In general, lower performance was reported for dishwasher and washing machine, where, due to the similarity of the appliance signatures and their average consumption power, there were misclassifications between the appliances and for the case of DTW+kM no classification of dishwasher. Similarly, fridge-freezer disaggregation reported high falsely identified operations while using the k-means based DTW methods, which can be explained by the acquisition of similar events in terms of average consumption, which during the original implementation were treated as part of the higher consuming appliances operating at the same time.

5.4.3.2 Disaggregation Case Study using House 17 from REFIT Dataset

The final case study, as in Chapter 4, is focused on the disaggregation of house 17 from the REFIT dataset [2, 3]. The combined DTW and k-means methods will be evaluated using the same appliances, as in the previous chapter, for reference, freezer, kettle, microwave and toaster, as none of the methods was able to obtain signatures for both washing machine and fridge-freezer, as discussed in Chapter 4.

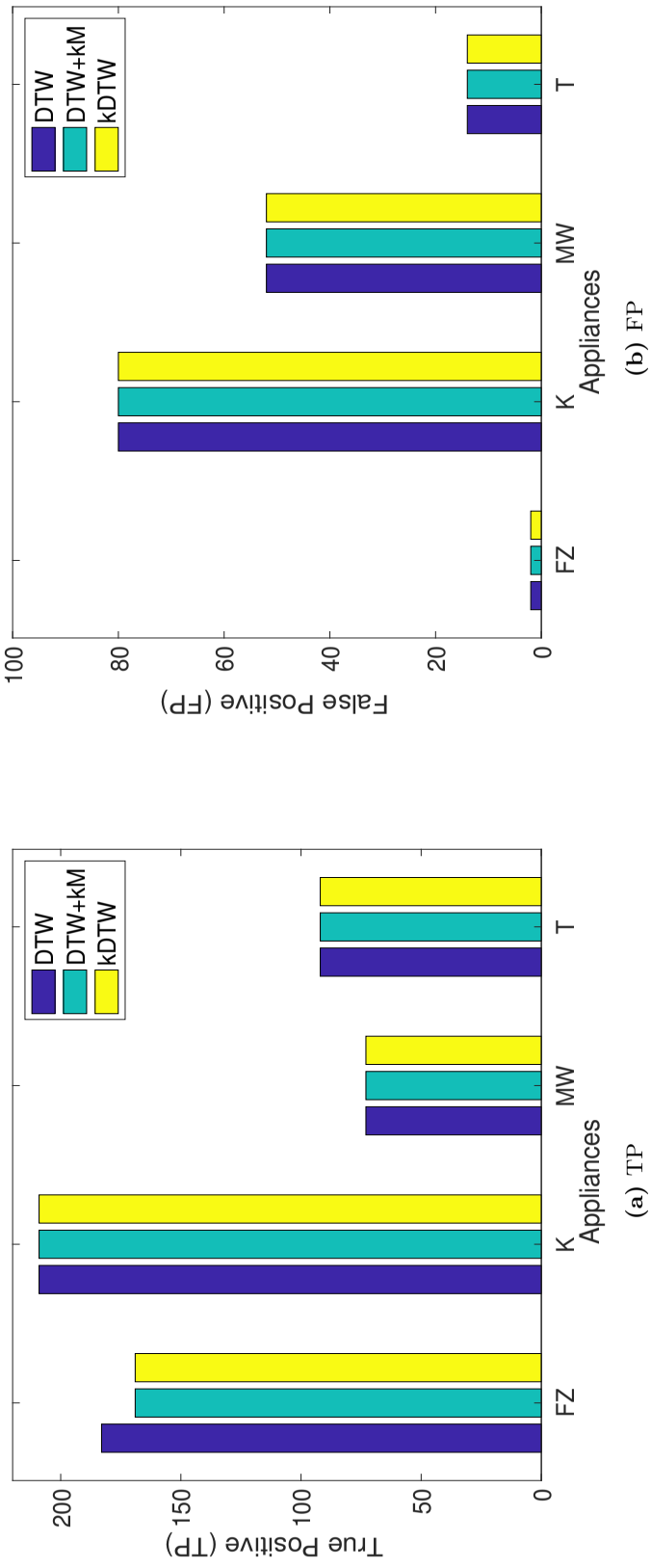
Table 5.13 shows the obtained TP, FP and FN values of the confusion matrix during the disaggregation of house 17 from the REFIT dataset [2, 3] using the two k-means and DTW methods, namely DTW+kM and kDTW.

Table 5.13: Confusion Matrix for REFIT House 17 using DTW+kM and kDTW.

	TP		FP		FN	
Apps	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*
FZ	169	169	2	2	353	353
K	209	209	80	80	12	12
MW	73	73	52	52	15	15
T	92	92	14	14	9	9

Note: Apps=Appliances, FZ=Freezer, K=Kettle, MW=Microwave, T=Toaster. TP=True Positive, FP=False Positive, FN=False Negative.

Figures 5.22(a)-(b) and 5.23 show the obtained classification accuracy with regards to TP, FP, FN from the confusion matrix, based on the results presented in Tables 4.15 and 5.13. It can be easily observed that all the DTW based methods show a very similar performance in terms of TP, FP and FN. DTW+kM and kDTW were able to classify $\sim$ 10% less freezer operations compared to the original DTW-based method. Those freezer events were not detected during edge detection as the corresponding rising edge was *"covered"* by higher events that seemed to last only for one sample, that during the original implementation were iteratively removed by the aggregate load, *"uncovering"* the hidden start event of the freezer.



(a) TP (b) FP

Figure 5.22: Classification accuracy using TP and FP for the proposed DTW+kM and kDTW methods, with respect the DTW-based method from Chapter 4, during the disaggregation House 17 from REFIT dataset [2, 3], as reported in Table 4.15 and 5.13.

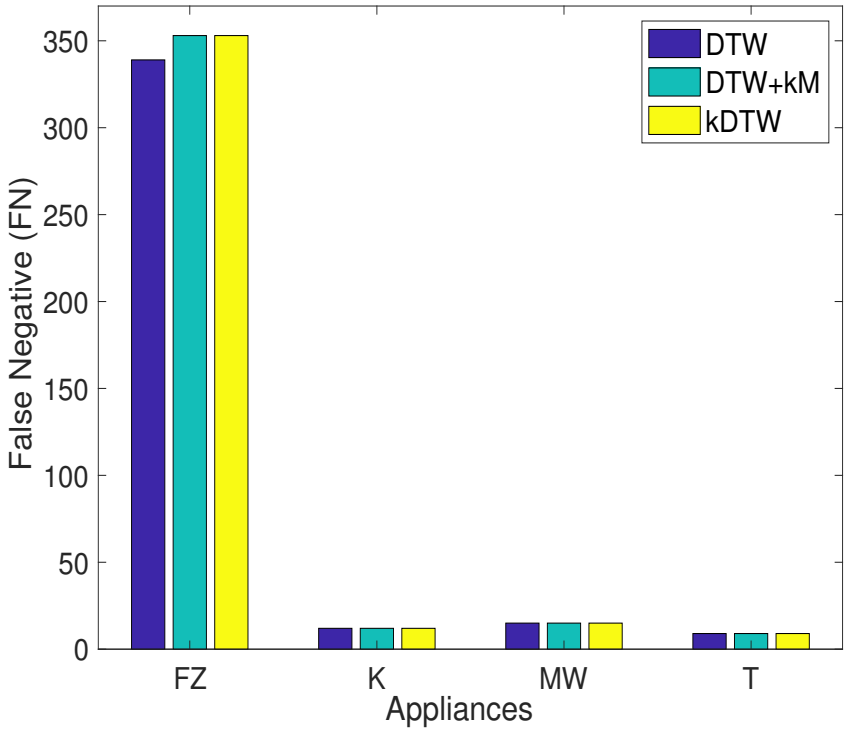


Figure 5.23: Classification accuracy using FN for the proposed DTW+kM and kDTW methods, with respect the DTW-based method from Chapter 4, during the disaggregation for House 17 from REFIT dataset [2, 3], as reported in Table 4.15 and 5.13.

Table 5.14 presents the classification accuracy of the proposed DTW+kM and kDTW methods with regards to the obtained precision, recall and F_M for house 17 from REFIT dataset [2, 3].

At a glance, it can be observed, that the DTW refinement did not provide any improvement for the specific case study, therefore the classification accuracy using all the metrics is the same for both DTW+kM and kDTW methods. Both methods reported the same precision and recall with the original DTW-based method, proposed in Chapter 4, for most of the appliances. The only exception was the freezer, where the combined k-means and DTW methods reported recall equal to 0.32, while the DTW-based method reported 0.35. This was the result of the higher number of false negatives reported during the disaggregation of house 17 using the DTW+kM and kDTW methods.

Table 5.14: Classification Accuracy using Precision, Recall and F_M for REFIT House 17 using the proposed DTW+kM and kDTW methods.

	PR		*RE*		F_M	
Apps	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*	*DTW+kM*	*kDTW*
FZ	0.99	0.99	0.32	0.32	0.48	0.48
K	0.72	0.72	0.95	0.95	0.82	0.82
MW	0.58	0.58	0.83	0.83	0.68	0.68
T	0.87	0.87	0.91	0.91	0.88	0.88

Note: Apps=Appliances, FZ=Freezer, K=Kettle, MW=Microwave, T=Toaster. PR=Precision, RE=Recall, F_M=F-Measure.

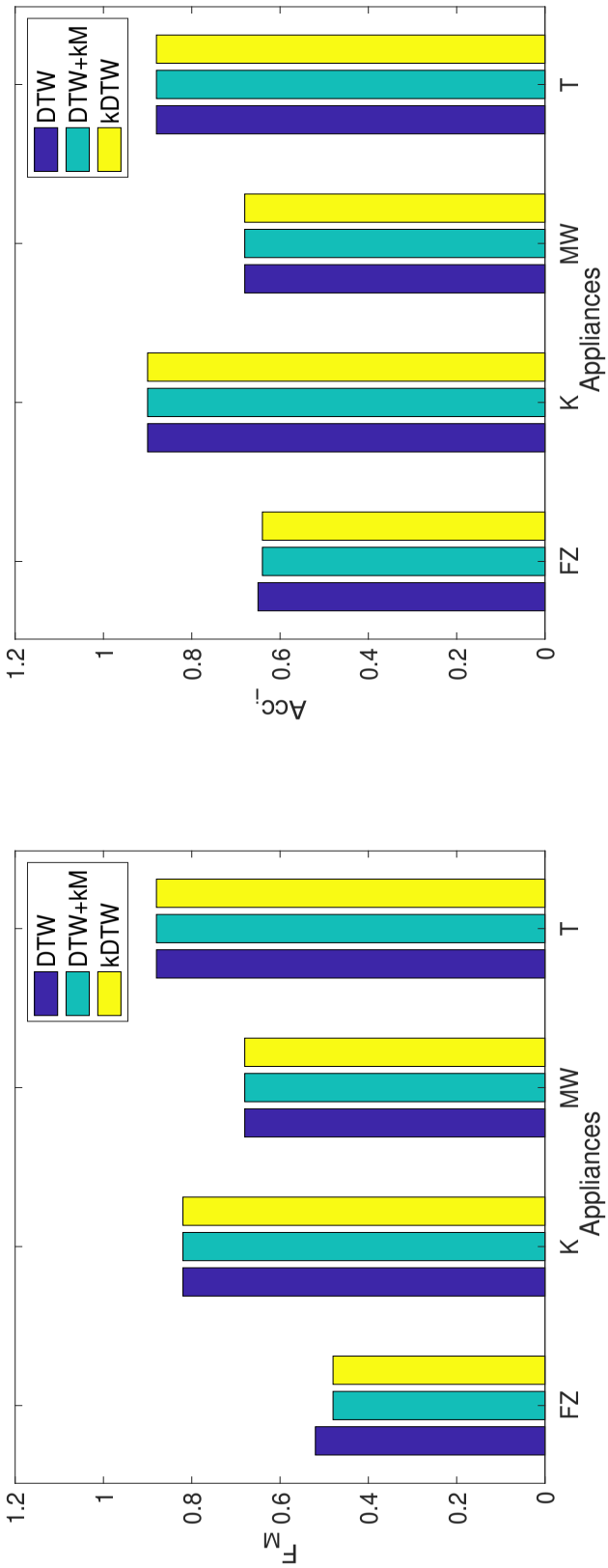
Table 5.15 shows the classification and estimation accuracy, using F_M and Acc_i respectively, between the proposed methods of this chapter, the DTW-based method from Chapter 4, while Figures 5.24(a)-(b) present those results graphically, for visual comparison.

Table 5.15: Classification Accuracy using F_M and Estimation Accuracy using Acc_i of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, for REFIT House 17.

	F_M			Acc_i		
Apps	*DTW*	*DTW+kM*	*DTW*	*DTW*	*DTW+kM*	*kDTW*
FZ	0.52	0.48	0.48	0.65	0.64	0.64
K	0.82	0.82	0.82	0.90	0.90	0.90
MW	0.68	0.68	0.68	0.68	0.68	0.68
T	0.88	0.88	0.88	0.88	0.88	0.88

Note: Apps=Appliances, FZ=Freezer, K=Kettle, MW=Microwave, T=Toaster.

Similarly with what has been already discussed earlier in this case study, the combined k-means and DTW methods reported the same classification accuracy with the DTW-based method for most appliance. For the case of the freezer, where the DTW+kM and kDTW methods reported higher number of false negatives, F_M was 0.48, which is 4% less compared to the accuracy reported by the original DTW-based method.



(a) F_M (b) Acc_i

Figure 5.24: Performance evaluation using F_M and Acc_i for the proposed DTW+kM and kDTW (Table 5.15) with benchmarks the DTW-based method proposed in Chapter 4 for REFIT House 17.

(a) Disaggregation load contribution using DTW+kM and the kDTW method

(b) Disaggregation load contribution using DTW-based method

Figure 5.25: Disaggregation results, as percentage of load contribution per appliance relative to the aggregate load, for House 17 from REFIT dataset using the DTW based method, proposed in Chapter 4, and the DTW+kM and kDTW methods.

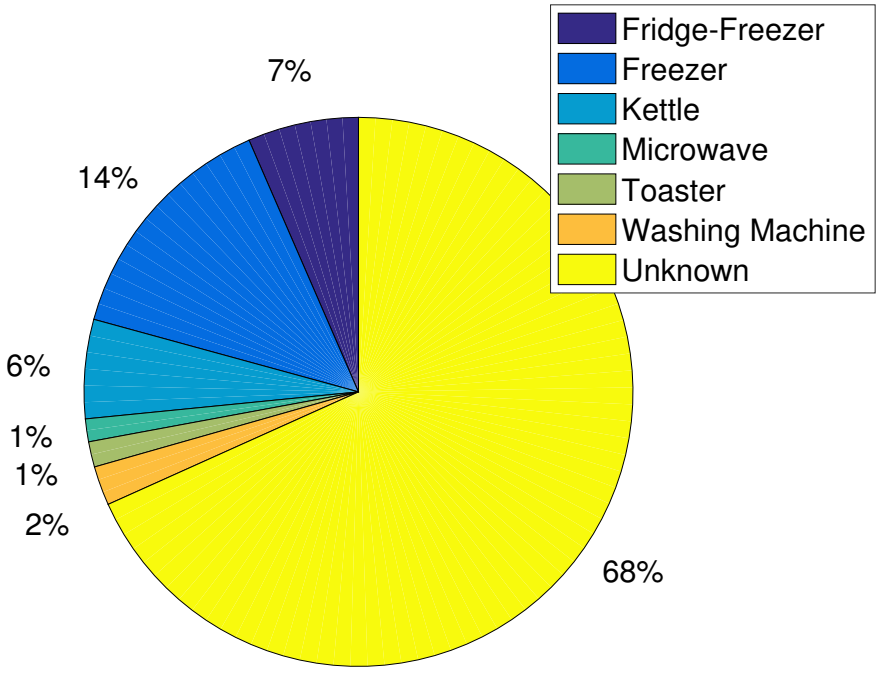


Figure 5.26: Ground truth load contribution, as percentage of load contribution per appliance relative to the aggregate load, for House 17 from REFIT dataset [2, 3].

Furthermore, in terms of estimation accuracy per appliance, the DTW+kM and kDTW methods reported the same accuracy for most appliance, which the exception of the freezer. They obtained comparable accuracy at 0.64, only 1% lower than the estimation accuracy reported by the DTW-based method.

Figures 5.25(a)-(b) show the disaggregation results of the proposed DTW+kM, kDTW NILM and original DTW-based methods as percentage of load contribution per appliance with regards to the aggregate load and the relevant ground truth consumption per appliances (Figure 5.26), available from the IAMs of the REFIT dataset [2, 3]. Both DTW+kM and kDTW obtained equivalent disaggregation results, therefore 5.25(a) represents the consumption contribution of both. The only difference between those methods and the DTW-based method from Chapter 4 is the estimated consumption contribution of the freezer, which is 1% more for the latter. Overall the other appliances, both known and unknown reported similar contribution.

In summary, during the disaggregation of house 17 from the REFIT dataset, it has been identified that the DTW refinement method was not able to improve the performance of the DTW+kM method and both methods reported equivalent classification and estimation accuracy, comparable to the DTW-based NILM method proposed in Chapter 4.

5.5 Overall Performance Evaluation

Similarly with Chapter 4, this section will evaluate the overall performance of the unsupervised methods presented in this chapter, with benchmarks the unsupervised DTW-based method proposed in Chapter 4 and the methods presented in [5], together with any observations regarding the advantages and drawbacks of the proposed methods. Table 5.16 presents the overall classification accuracy using $F1_{micro}$, as defined in equation 2.6, which calculates the *micro-average* using the total values of TP, FP and FN for all classified appliances.

All the proposed unsupervised methods were able to perform disaggregation for house 2 in both REDD and REFIT for the same appliances as presented in [5]. Similarly with Chapter 4, the DTW+kM and kDTW methods were not able to train and classify bathroom GFI, lighting and oven for house 1 in REDD and fridge-freezer and washing machine for house 17 in REFIT. For this purpose, the m- results for each proposed method are introduced, in order to account for the total number of relevant appliance operations and include them as false negative values, in order to perform a better comparison with the methods in [5]. Figure 5.27 shows the graphical representation of the overall performance reported in Table 5.16.

For house 2 from REDD dataset [1], the proposed combined methods of this chapter, together with the original DTW-based method, were able to obtain the highest overall classification accuracy, with best obtained when using the DTW+kM together with the DTW refinement, which reported 0.81, slightly better than the 0.8 obtained using the original method. Even when obtaining disaggregation solely by using the DTW+kM method, the reported results have shown an accuracy of 0.76, better but comparable to that obtained by the P refined methods in [5].

Table 5.16: Overall Classification Accuracy using $F1_{micro}$, as defined in equation 2.6, with benchmarks P-SGSP [5, 33], P-UGSP [4, 5] and P-DT [5, 32], as presented in [5].DTW, DTW+kM and kDTW present the overall accuracy for only the disaggregated appliances, whereas *m-* is accounting for the missing operations of the appliances that where not disaggregated using DTW, but were present in the disaggregation using the methods in [5].

Datasets	*REDD*		*REFIT*	
NILM Methods	*House 1*	*House 2*	House 2	House 17
DTW	0.79	0.80	**0.77**	0.68
m-DTW	0.49	0.80	**0.77**	0.51
DTW+kM	0.77	0.76	0.48	0.67
m-DTW+kM	0.48	0.76	0.48	0.50
kDTW	0.78	**0.81**	0.68	0.67
m-kDTW	0.49	**0.81**	0.68	0.50
P-SGSP	0.69	0.73	0.61	**0.62**
P-UGSP	0.67	0.72	0.59	0.59
P-DT	**0.70**	0.68	0.59	0.60

Furthermore, when reporting only the classified appliances, the DTW-based method from Chapter 4 was able to outperform both the proposed method and the methods presented in [5] for house 1 from REDD dataset, while the remaining proposed methods obtained the comparable classification accuracy of 0.77 and 0.78, which is higher than the methods used for benchmarking. When taking into account the appliances, that were not classified using the proposed methods of this thesis, the best performance is reported with 0.7 for P-DT [5, 32], with the other P refined methods claiming a comparable overall classification accuracy. The reported accuracy for the proposed methods was for most cases ~ 0.49, which is $\sim 20\%$ less than the P refined methods. As already explained in Chapter 4, the DTW algorithm had difficulty in separating the bathroom GFI events from microwave operations, which was not resolved with the implementation of the combined DTW and k-means schemes.

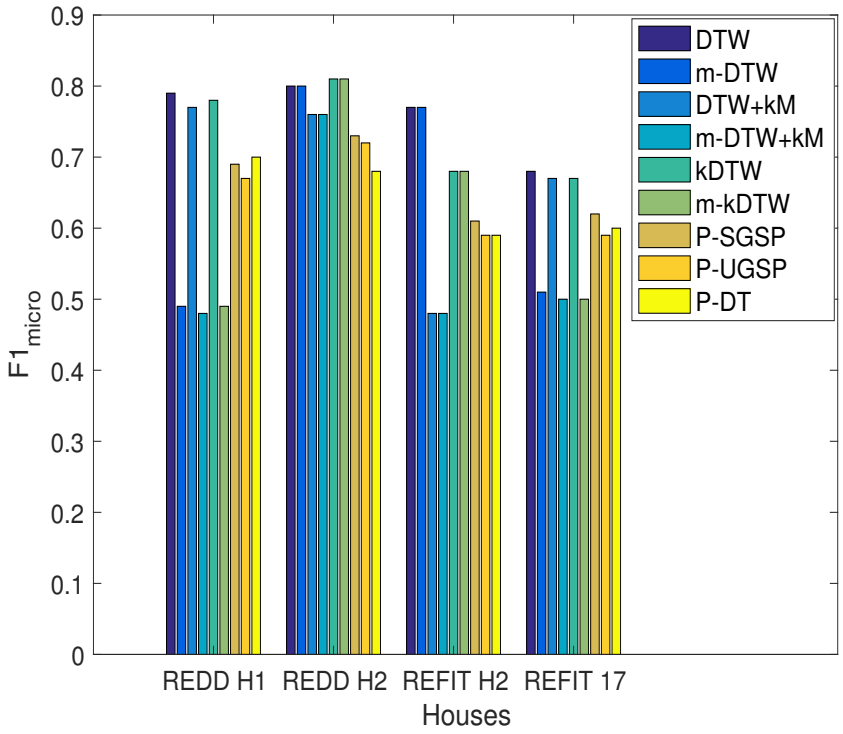


Figure 5.27: Overall classification performance evaluation using $F1_{micro}$ for the proposed DTW-based NILM method, as presented in Table 5.16, with benchmarks P-SGSP [5, 33], P-UGSP [4, 5], P-DT [5, 32], HMM [35], for Houses 1 and 2 from REDD dataset [1] and Houses 2 and 17 from REFIT [2, 3], as presented in [4, 33]. m-DTW presents the overall performance by taking into account the FN of the appliances that were disaggregated in [5] but not using the proposed DTW methods.

For house 2 from the REFIT dataset [2, 3], the DTW-based method from Chapter 4 reported the highest overall classification by more than $\sim$ 10% from all the other methods. The DTW+kM method obtained the lowest results at 0.48 and the DTW refinement was able to improve it by 20%. This was the result of the increased number of events wrongly identified as fridge-freezer, due to the selection of not performing iterative disaggregation and removal of the detected load. For house 17 from the REFIT dataset, the proposed unsupervised methods, obtained accuracy of $\sim$ 0.67 and $\sim$ 0.5, when taking into account the missing fridge-freezer and washing machine, which is $\sim$ 10% less than the methods presented in [5].

Furthermore, the DTW refinement was able to improve the performance of the DTW+kM method for most houses with the exception of house 17 from REFIT, where all methods obtained the same classification accuracy. Similarly, all the proposed

methods seemed to perform with high classification and estimation accuracy, when disaggregating kitchen appliances for most houses in REDD and REFIT datasets. The reported normalised total power consumption estimation error (TER) for the proposed methods as a whole was the lowest for the kitchen appliances of house 2 from REFIT dataset compared to the methods presented in [5]. The difficulty of the original method to distinguish the washing cycle of the dishwasher and washing machine during the disaggregation of house 2 from REFIT dataset was not resolved, and the DTW+kM method was not able to classify any dishwasher operations, similarly with house 1 from the REDD dataset, where the dishwasher was misclassified as kitchen outlet. k-means seemed to *"merge"* the clusters with similar average power, even though they started from different centroids. The refinement method in both cases was able to obtain the equivalent events as with the original method, without being able to further increase the accuracy.

Another interesting observation was the increased number of misclassified events as fridge-freezer operation when using the DTW+kM method with and without the DTW refinement. As it has already been explained, this is the result of events with similar average power consumption, that during the disaggregation were *"absorbed"* by the higher consuming appliances depending on the parameter C.

5.6 Execution Time Evaluation

This section is focused on evaluating the performance of the three proposed methods in terms of their execution time for both creating the library of appliance signatures and disaggregation of a given aggregate load. The performance will be presented using house 2 from the REFIT dataset, but the performance of the methods was equivalent for the rest of the houses in both REDD and REFIT dataset.

Figure 5.28(a) shows the relevant execution time versus the number of aggregate load samples used for the case of house 2 in REFIT dataset, when creating the library of appliance signatures. These aggregate data correspond to the dates used for creating the library, as found in Table 4.1. Note that for each method, the algorithms were executed for a variety of number of samples, eg. 1000, 10000, 20000, etc., starting from the beginning of the dataset that corresponds to the dates in Table 4.1.

(a) Processing Time vs Number of aggregate load samples during library creation

(b) Execution time vs Number of aggregate load samples during disaggregation

Figure 5.28: Execution time during library creation and disaggregation for REFIT House 2.

According to Figure 5.28(a), it is apparent that both DTW+kM and kDTW approaches perform the same during this step, as they create the library in exactly the same way, as already discussed earlier in this chapter. Furthermore, the methods proposed in this chapter were able to obtain significantly the execution time when creating the library, compared to the DTW-based method.

Figure 5.28(b) shows the processing time while performing disaggregation versus the number of aggregate load samples for house 2 in REFIT dataset. These aggregate data correspond to the dates used for disaggregation, as found in Table 4.1. Note that for each method, the algorithms were executed for a variety of number of samples, eg. 1000, 10000, 20000, etc., starting from the beginning of the dataset that corresponds to the dates in Table 4.1.

The original DTW-based method is the most computational complex, while the DTW+kM method reported the lower execution time, as the classification is performed using k-means clustering on the average power consumed per appliance operation. Furthermore, the kDTW method, was more efficient in terms of execution time compared to the DTW-based method, but was outperformed by the less complex DTW+kM approach, as seen in Figure 5.28(b).

5.7 Summary

The aim of this chapter, was to propose an optimised method using the concept of the DTW-based method, proposed in Chapter 4, in order to improve the overall performance and reduce the computational complexity of the method, but without sacrificing the unsupervised learning, which is the main focus of this thesis. k-means is a commonly used clustering method with low computational cost and the capability to support both semi-supervised and unsupervised learning.

Furthermore, k-means has been used in various NILM works, especially for clustering features, as a standalone method or combined with other methods, such as SVM and more combined with DTW, by using DTW as a distance measure for k-means [87, 117, 118, 170]. The proposed combined DTW and k-means method, uses DTW for the creation of a library of appliance signatures and the initialisation of the number of clusters and the centroids, using unlabelled historical aggregate load data. Classification

through clustering of the average power consumption of each detected event.

In addition, a refinement method has been proposed, that allows the use of DTW for further refining the events that were either misclassified or not classified at all, as it has been observed by using as a feature the average power of each event it can lead to misclassification between appliances with similar average consumption. Any differences in terms of pre-processing, event detection, feature extraction and classification, similarly with the method proposed in Chapter 4, have been discussed and the relevant algorithmic representation was included.

For evaluation purposes, the same datasets have been used as in Chapter 4, for reference REDD [1], downsampled at $1min$, and REFIT [2, 3] datasets, with benchmarks to the DTW-based method of Chapter 4. Appendix B provides tables and graphs with the obtained performance with benchmarks the same state-of-the-art NILM algorithms, as in Chapter 4, for any interested party.

Furthermore, both methods were able to reduce the complexity of the original DTW-based method, in terms of execution time, with the DTW+kM method showing the lowest processing time, compared to the other two methods. The DTW+kM method was able in most cases to report equivalent performance for most appliance across houses, in terms of classification and estimation accuracy, but failed to distinguish appliances with similar average power consumption, such as dishwasher and washing machine. Moreover the kDTW method was able to improve the performance of the DTW+kM in most houses, while the method was able to obtain comparable performance to the original DTW-base method without the overhead of iterative removal of disaggregated loads.

Chapter 6

Data Analysis for other IoT Applications

6.1 Introduction

Advancements in monitoring and sensor technology together with software development for interfacing and managing sensors and databases make the concept of IoT a realistic but challenging scenario. Proposing methods for handling and analysing data from various disciplines is of great importance, thus this chapter attempts to investigate such methods for two scenarios, one for earthworks monitoring, and one for user's attention assessment using eye-tracking devices. These scenarios are presented in the format of two different case studies, with related background, data analysis techniques and outcome discussion.

The first case study, published in [177], proposes a prototype low-cost platform with a variety of sensors located on one sensor node for gathering real-time data for resistivity, ground movement and pressure, that can assist to predictive maintenance by monitoring and analysing moisture (via resistivity sensors) and pressure in the embankment. This application falls in the smart environment realm of IoT (see Appendix C).

Eye-tracking devices can be used for monitoring both pupil dilation and gaze fixation, features that can be used in order to extract useful information regarding user's attention and interaction with visual content. In the context of IoT, this information could be applied for advertisement and personalised services, when using devices with enabled eye-tracking properties through cameras and in the healthcare domain, for psychological assessment and patient assistance, similarly with the system proposed in [17], which has been discussed in Section C.2.2.2. Furthermore, this could be applied in driving assistance systems by identifying patterns in the driver's behaviour, through a

correlation of outdoor camera data and vehicle implemented eye tracking devices. Thus, the purpose of the second case study, published in [178], is to evaluate signal processing techniques, that could be incorporated for assessing user's attention when interacting with various contents.

Those case studies show the variability in terms of measurements and features extracted, in addition with various sampling rates, depending on the sensors/equipment used, which represents the most challenging problem in IoT.

6.2 Embankment Monitoring

Asset monitoring of canal structures, embankments, bridges and a number of other critical assets will deliver sustainability through the reduction of human intervention for assessing the health of the asset and promote enhanced data quality and accessibility for best practice in environmental management, as required by environmental regulators and other government departments.

6.2.1 Background

A key component of management of water resources lies in asset monitoring of the structures that contain water, e.g., dams and embankments. Earthworks failures can lead to disastrous consequences, including flooding, and can be very expensive to remediate. Early intervention and prevention requires identification of the incremental development of internal conditions that ultimately trigger failure. Spatially continuous data can achieve a level of sub-surface resolution significantly closer to the scale of true heterogeneity than currently achieved using conventional intrusive point sensing approaches alone.

While current automated procedures, sensors and SCADA systems provide information regarding the health of the assets, they have a number of limitations: (a) the cost of deploying/maintaining these solutions; (b) the level of intrusiveness; (c) the need for engineers validating measurements by visual inspection; (d) low temporal resolution with limited scope for predictive approaches to asset failure; (e) limited support for strategic decision-making. For this purpose, there are currently available different solutions in order to monitor the health of an earthwork. According to Sellers

et al. [179], one of the most commonly used methods is soil resistivity surveys and a 2D or 3D mapping of the ground using these results, known as *Electrical Resistivity Tomography* (ERT), as it can provide information about the moisture content of the earthwork in a non intrusive way. An example of this method is the ALERTme system, which maps 3D the resistivity of a railway embankment using a kit designed for this purpose (specifications of 500V/up to 500mA), as it can be seen in Gunn *et al.* [180]. Similarly, there are available many ready to use solutions, but similarly with the ALERTme project, they use high voltage and current, which requires expensive voltage transformers, and cabling in order to mitigate health and safety risks during the experiments.

As the purpose of the work presented in this thesis was not to perform ERT analysis, a detailed overview of ERT related research can be found in [181], where the authors reviewed various 2D and 3D methods. According to the authors, the systems can be categorised as *static* and *dynamic*. Static systems, are the systems that require a high number of electrodes interconnected through cabling, that remain in the ground during the survey, whereas dynamic systems use a much smaller number of sensors, that have to be moved during the survey, in order to map the asset in question [181]. The prototype presented in this case study can be classified as a static system, due to its specifications.

Together with the resistivity and moisture content, which can be also measured using dedicated pore pressure sensors, the field experts suggested that the movement of the ground (vibration, acceleration) can be a very informative measurand, as this could provide early notification about the possibility of earthworks failure. With the available solutions, a survey like that will require a high budget and additionally the usage of many different sensors, that generally do not allow the measurements using one and only interface.

This case study presents a solution to assess the physical integrity of vulnerable earth structures (dams, embankments and cuttings) - thereby facilitating the shift from more costly responsive remediation of earthwork failures to early intervention. A unique, customised and cost-effective platform for automated monitoring of earthworks through prototyping a novel hardware/firmware solution in consultation with various stakeholders: (i) integration of analogue and digital sensors for measuring pressure and

motion, (ii) resistivity sensor (board) that is controlled by main hardware (board) and requires low voltage compared to the off-the-shelf resistivity solutions, (iii) variable and on-demand sampling rates that can be dynamically controlled, (iv) a prototype mechanical waterproof design for housing main hardware, resistivity sensor-board and relevant sensors.

6.2.2 System Set-up

The proposed monitoring platform could integrate a range of sensors for monitoring the condition of earthworks assets (embankments, canal infrastructures etc) with minimised cost and high accuracy. This project has delivered proof-of-concept by deployment of a 12V network of five integrated-sensor nodes, appropriately cased and connected for power and communications under the ground surface. Measurements of resistivity, ground movement and pressure from the sensor network are communicated periodically and autonomously to a gateway and then onto a data collection hub. The outcome of this research underpins deployment of a network of these sensor nodes, application of monitoring techniques, intelligent data mining and data analytics to derive models concerning the condition of the assets and will assist in assessing infrastructures and informing management decisions.

6.2.3 Sensor Network

The deployment of the sensor network has been mainly affected by the measurement of soil resistivity, where a grid network of electrical resistivity electrode arrays is formed over the area of interest. The distance between the sensors is a direct function of the accuracy of the 3D mapping of the earthwork. For the above purposes, the sensors were deployed in a line of subsurface sensors, which could be extended in the future to multiple lines, in order to create the array that is usually used for resistivity measurements (Figure 6.1). In relation to Figure 6.1, the top row of sensor nodes has been deployed for proof-of-concept.

The deployed sensor network comprises the following:

- *Sensor Nodes* deployed in arrays: each node is a customised reprogrammable board that was designed and prototyped and is connected with three sensors and

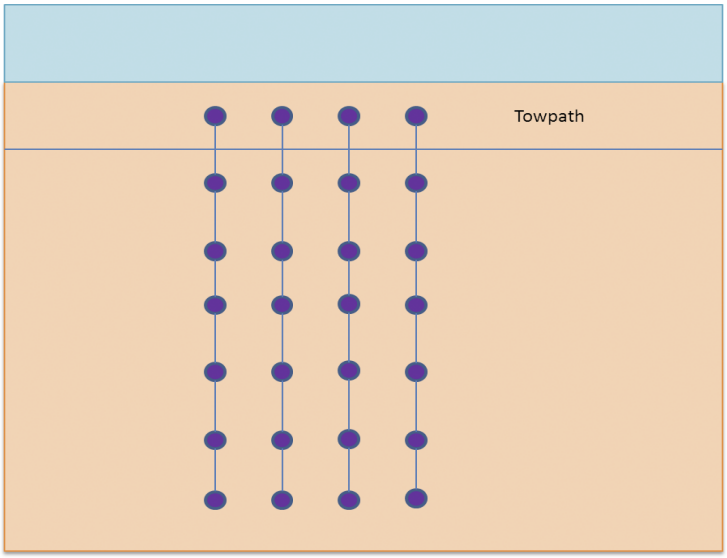


Figure 6.1: Embankment Deployment (Lateral View).

resistivity circuitry, and can be enabled for the usage of more sensors.

- *Sensor Communication Module*: Interfacing the sensor nodes to the gateway node using a *Controlled Area Network* (CAN).
- *CAN network*: The CAN cable consists of 6 pins. One pin is used to power the sensor nodes from the power supply, another is ground and the other two are used for CAN high and CAN low. The fifth pin is used as a ground sense, in order to have a reference for the resistivity measurement.
- *Hub for data collection*: For the deployment, the data were collected through a Raspberry Pi, but an interfacing of the Gateway with the Cloud could be performed through near white space communication ($\sim$ 433 MHz).

6.2.4 Prototype Sensor

The prototype sensor node consists of the following on and off-board sensors:

1. Resistivity board for injecting current, sensing voltage, sinking current, sensing voltage.
2. A Digital Accelerometer. This on-board sensor will be able to sense acceleration or vibrations ($\pm 2g/\pm 4g/\pm 8g$ dynamically selectable full-scale) in the soil.

3. An Analogue Pressure Sensor that can measure absolute pressure (0- 200kPa) using a single port, which was in contact with the soil (off board). Through a future calibration, the absolute pressure measurements can be potentially translated in terms of pore pressure. Currently, for measuring pore pressure, specialised and expensive sensors are required.

All the hardware has been integrated in PVC (*Polyvinyl Chloride*) tube, that provides both endurance and waterproof protection. The casing allows the removal of the hardware for for reprogramming or debugging should the need arise. At the bottom of the node, there is a short copper probe, that is used for the resistivity measurements and is safely connected to the resistivity board. While the voltage supply used for both for the sensor network was an enclosed 12 VDC, 1.3A Switch Mode Power Supply (SMPS).

A summary of the sensors and their sampling rates can be found in Table 6.1. These sampling rates can be adapted easily depending the circumstances and the defined requirements.

Table 6.1: Sensors Summary and Sampling Rates.

Sensor Types	*Sampling Rate*
Accelerometer	$12.5Hz$
Analogue Pressure	$1sec$
Resistivity Injection for measurement (Resistivity Board)	$5sec$

6.2.5 Resistivity Measurements Specifications

According to Wenner [182] and the IEEE standards in [183, 184], soil resistivity is measured using the 4-pin Wenner method. The first probe, as seen in Figure 6.2, injects current according to the system's specifications, the two intermediate probes sense the voltage in relation to a common ground and the last probe sinks the current.

The voltage measured at each probe provides the voltage drop required to calculate the Wenner resistance and consequently the apparent resistivity using the formulas found in equations 6.1 or 6.2. These calculations can be executed at the data collection hub, which will receive all the relevant measurements.

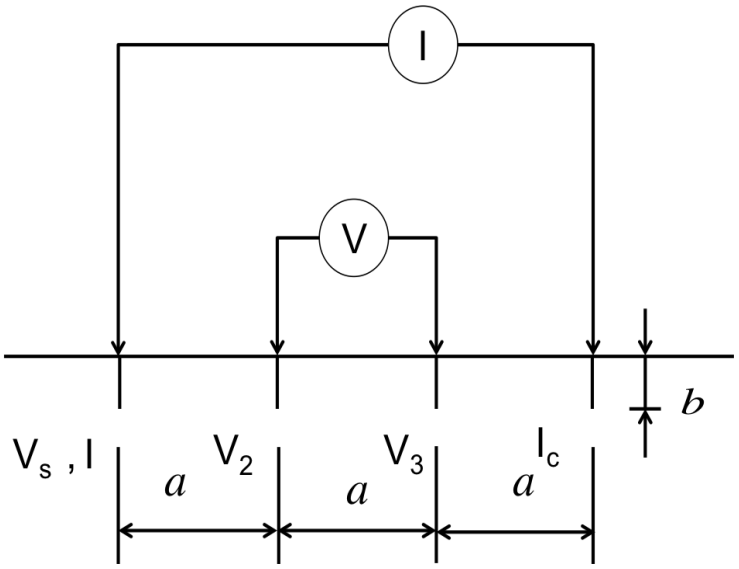


Figure 6.2: 4-pin Wenner Method.

The spacing between the probes for the deployment was selected to be 1m, and the depth of the probes is 44.5cm, the height of the casing is 32.5 cm and the length of the rod that is placed at the bottom of the tube is 12 cm.

$$\rho_w = \frac{4\pi a R_w}{1 + \frac{2a}{\sqrt{a^2+4b^2}} - \frac{a}{\sqrt{a^2+b^2}}} \tag{6.1}$$

where ρ_w is the *apparent resistivity* (Ωm), a is the spacing between the probes (m) b is the depth of the probes (m) and $R_w = \frac{V}{I}$ is the *Wenner Resistance* (Ω). If b is small compared to a, as is the case of probes penetrating the ground only for a short distance (as normally happens), the previous equation can be reduced to:

$$\rho_w = 2\pi a R_w \tag{6.2}$$

The resistivity sensor node can measure the voltage that is sensed at the two intermediate nodes, namely V_2 and V_3, the current injected I, the Voltage supply V_s and the sink current at the last node I_c. These measurements are critical for both calculating Wenner Resistance R_w and respectively soil resistivity ρ_w, but also in order to validate the measurements. For this purpose an additional measurement called ground compensation is acquired by using one of the extra leads of the CAN cable, which can compensate for power losses that occur due to the length of the CAN cable.

Table 6.2: Typical Soil Resistivity based on [183–185].

Soil Type	*Average Resistivity (Ωm)*
Well - graded gravel	$600 - 1000$
Poorly graded gravel	$1000 - 2500$
Clayey gravel	$200 - 400$
Silty sand	$100 - 800$
Clayey sands	$50 - 200$
Silty or clayey sand with slight plasticity	$30 - 80$
Fine sandy soil	$80 - 300$
Gravelly clays	$20 - 60$
Inorganic clays of high plasticity	$10 - 55$
Surface Soils	$1 - 50$
Clay	$2 - 100$
Sandy clay	$100 - 150$
Moist gravel	$50 - 700$
Dry gravel	$700 - 1200$
Limestone	$5 - 10000$
Porous limestone	$30 - 100$
Quartz, crystalline limestone	$100 - 1000$
Sandstone	$20 - 2000$
Granites	$900 - 1100$
Concrete	$300 - 500$

In order to determine soil resistivity from the above measurements, the voltage drop between the two nodes ($V = V_2 - V_1$) is calculated and followed by $R_w = V/I$, where I is the injected current. Finally, using equation 6.1, the soil resistivity measured values can be determined. Table 6.2 provides average soil resistivity values for different soil types.

Due to the voltage limitations, the global resistivity measurements will have to be redefined, as the maximum distance between the nodes will not exceed 1 m. It is important to mention here that commercial resistivity kits can use voltages up to 800V and inject currents up to 2.5 A. The proposed resistivity circuit is able to inject much lower currents, which was set for the deployed nodes up to 119 mA, and was selected by taking into consideration the common values of resistivity (1-10000 Ωm) and an average spacing of 1m. Every sensor board is connected to a solid copper probe, similar to commercial resistivity kits and can work using both injection, sensing, or sinking mode. The most common material for these rods is stainless steel, but solid copper rods are also widely used and also due to the voltage limitations of the specific project a solid copper rod would offer higher conductivity compared to stainless steel.

6.2.5.1 Deployment

Before the deployment, an in-house and residential field testing was performed in order to ensure the system's performance in terms of communication, readings especially for the case of a field-testing and validate resistivity measurements using resistors (in-house testing). During the residential test, it became apparent, that the resistivity readings were affected by the mains ground, as both laptop and lab power supply were powered by house mains. The voltages sensed when the power supply was injected and when it did not, had a small deviation, which shows that there was increased noise due to the residential appliances and ground.

For the above purpose, it was decided for the deployment to use the SMPS, which provides a ground for the deployed system that is isolated from the mains ground, as in Falkirk Wheel, the site is close to high voltage facilities, that will also provide supply to the SMPS. Additionally, the power cable used to power the SMPS, does not use the ground lead.

Deployment was carried out in Falkirk Wheel at Falkirk, Scotland at an embankment

that is maintained by Scottish Canals during late February-beginning of March 2014. During the test period the weather at area was close to the average temperatures of the area with no extreme below zero temperatures.

6.2.6 Results and Discussion

All the collected data from each resistivity sensor were further processed in order to get the resistivity, ground movement and pressure. Due to the current setup, the sensors are sending directly their measurements to the gateway, where they receive a timestamp. Due to the amount of the data and each sensor's sampling rates, there have been cases of dropped or wrong order packets. This synchronisation is more important for the resistivity analysis, as in order to define one resistivity value, measurements from all sensors are required.

During the pre-processing of the data, all resistivity data has been partitioned into windows, with each window starting from the message that the source is open (current will be injected), and finishing when the closing message was received. All remaining windows were discarded as not useful. For the case of accelerometer and pressure data, any missing or incomplete data have been replaced using linear interpolation between the preceding and subsequent measurements.

Two different approaches have been used in order to calculate resistivity: (1) Assuming that the data arrive at the gateway with the same order that each sensor receives its measurements, and (2) assuming that the data can arrive with a different order but still can be grouped per sensor.

Figure 6.3(a) shows the results from the first scenario and Figure 6.3(b) from the second scenario. For both cases the average resistivity varies between 40-60 Ωm. In both cases, there are some higher resistivity values that reach around 90 Ωm, which can be either the result of missed data at the specific measuring window, though it could still represent a normal resistivity value, which will be discussed further.

Figure 6.3(b) is clearly more settled, as all the data that do not follow the initial assumption are discarded, thus the different number of measurements. This does not affect marginally the average resistivity, due to the high sampling rate.

(a) Resistivity using sensor network order to arrive at gateway.

(b) Resistivity when different arrival occurs

Figure 6.3: Resistivity.

According to Nwankwo *et al.* [186] and [183–185], the resistivity measurements that are presented in Figures 6.3(a) and 6.3(b) can be categorised as clay, which is one of the most common soil types in Scotland. During the set up period, there were no extreme temperatures that would affect severely the measurements (extreme cold/dryness). Pangonilo in [185] claims that clay resistivity can be between 2-100 Ωm.

The accelerometer sensor is able to capture 3-axis vibrations/movement. Figure 6.4 shows the accelerometer sensor readings. While readings do not exceed a sum of squares of 200 to 250, there is a clear peak at 300. This occurred during an artificially induced disturbance (jump) in the vicinity of the sensor node, which resulted in noticeable ground movement. This indicates the need for destructive testing, further data analytics and the potential to detect clear patterns of embankment failure.

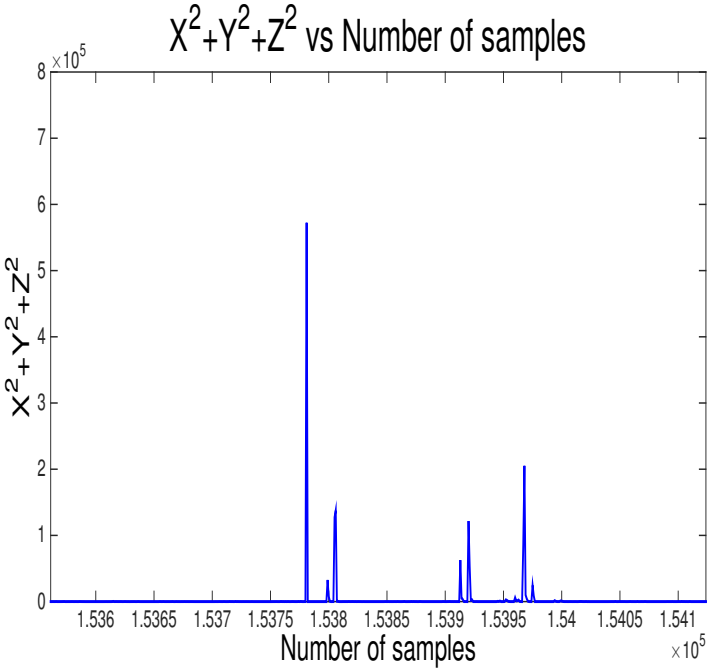


Figure 6.4: $X^2 + Y^2 + Z^2$ vs number of samples.

Figures 6.6(a)-(b) show the pressure readings and the conversion to kPa units, respectively.

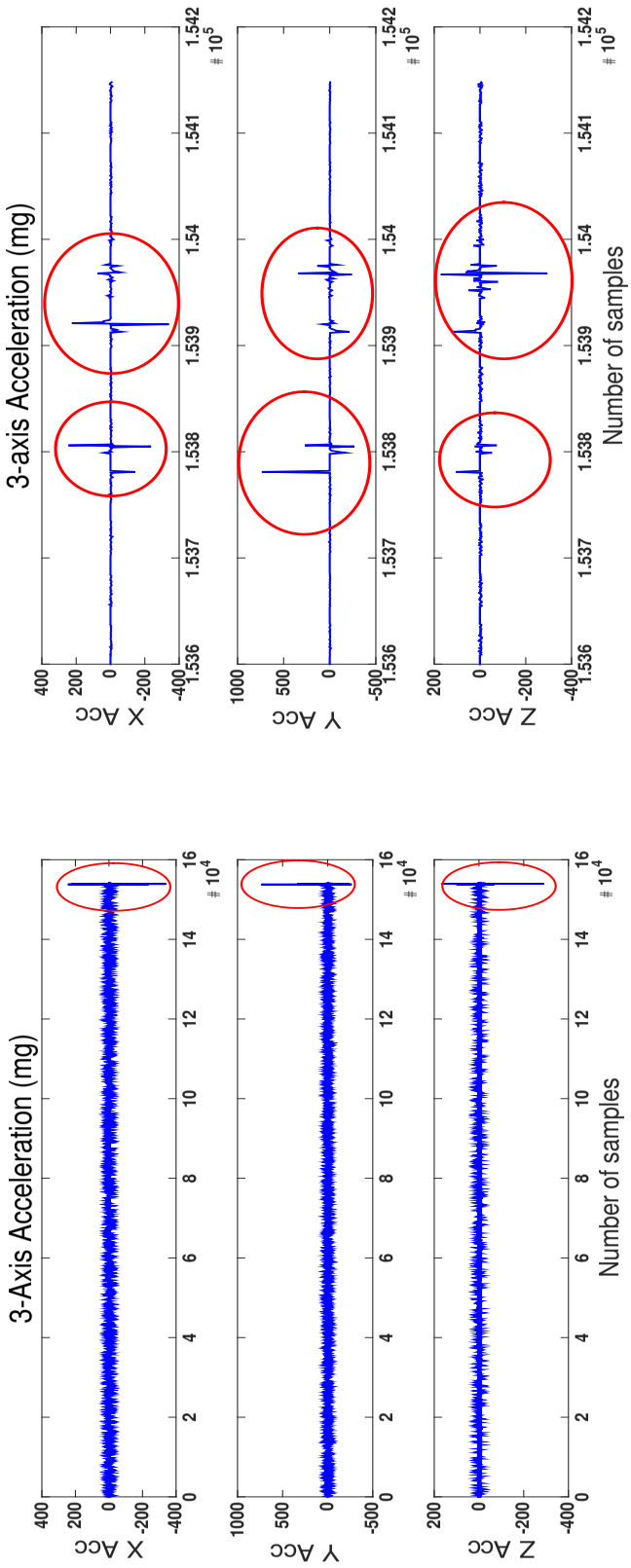
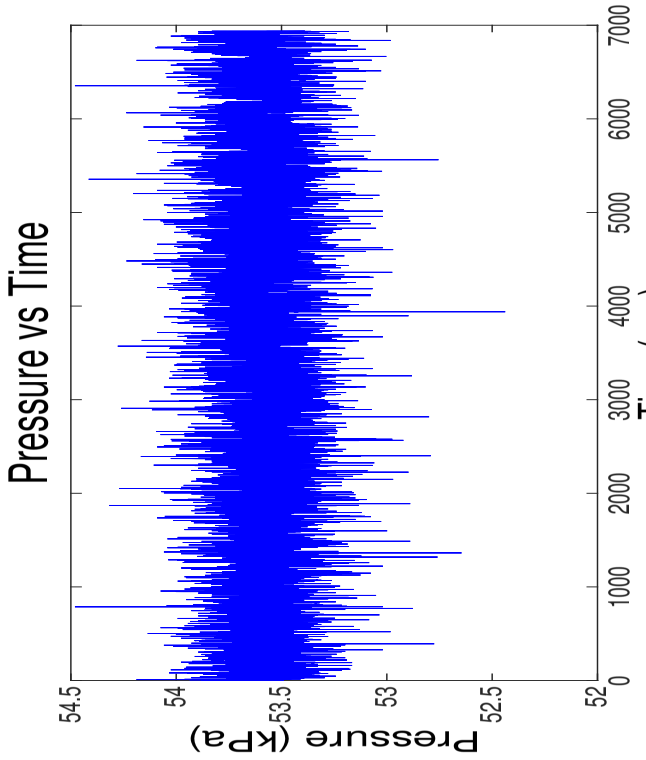
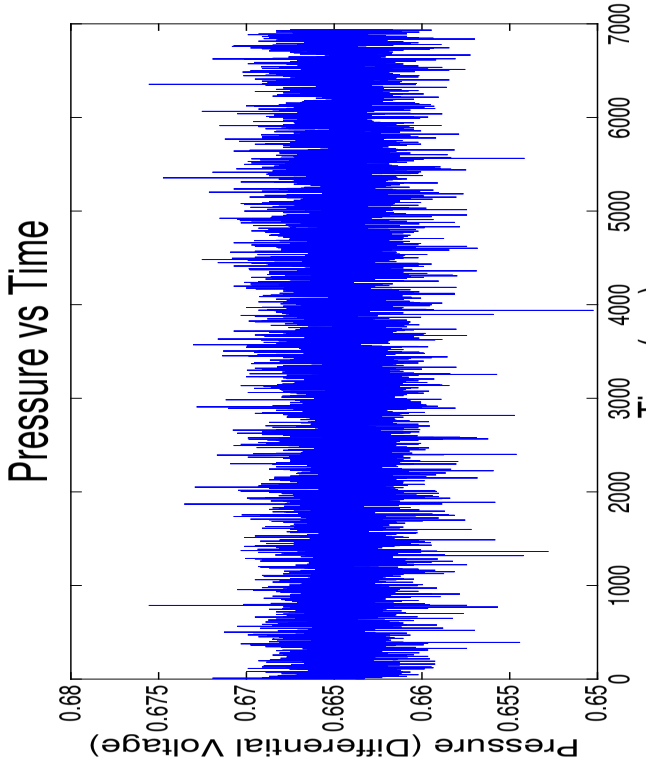


Figure 6.5: 3-Axis Acceleration vs number of samples.



(b) Pressure variations (kPa) versus time.



(a) Pressure variations versus time.

Figure 6.6: Pressure variations versus time.

The deployed pressure sensor is a differential pressure sensor and it provides as an output the differential voltage, which is proportional to the differential pressure applied. This voltage output can be found in Figure 6.6(a) and it has been amplified by a gain of 62. Figure 6.6(b) shows the converted voltage output to pressure units by using the sensitivity of the sensor (according to the specific application) $S = 0.2mV/kPa$ and by attenuating for the output gain. According to the figure, a variation of around 2 kPa is noticeable during the 2 hours sample timespan and the above results translate to almost 0.5 atm.

In conclusion, the proposed prototype system incorporates heterogeneous sensors, which represent one of the most important IoT-enabling technologies. These sensors allow the monitoring of various parameters that could affect the integrity of the embankment infrastructure. Data pre-processing included the partitioning of resistivity data into windows, using the injection and sinking of current process and the linear interpolation of adjacent measurements when data were missing from the accelerometer and pressure datasets. The obtained data have been used in order to identify the usefulness of features such as resistivity, 3-axis movement and pressure, which could allow future implementation for full scale ERT and correlation of all features in order to create predictive models using suitable data mining techniques.

6.3 Methods for Attention Assessment in Visual Content Interaction

Eye movements and changes in pupil dilation are known to provide information about viewer's attention and interaction with visual content. The purpose of this case study was to evaluate different statistical and signal processing methods for autonomously analysing pupil dilation signals and extracting information about viewer's attention when perceiving visual information. In particular, using a commercial video-based eye tracker to estimate pupil dilation and gaze fixation, this work demonstrates that wavelet-based signal processing provides an effective tool for pupil dilation analysis and discusses the effect that different image content has on pupil dilation and viewer's attention.

6.3.1 Background

Objectively assessing users' experience when interacting with visual content is gaining increased research interest due to its relevance for numerous applications ranging from video compression, Human Computer Interaction (HCI)-based decision support tools design, web-site design, and Internet visual searches, to those related to marketing, science and medicine. Eye trackers that use high-resolution, high-speed video cameras to record eye movements and corneal reflections are cost-effective tools for high-precision measurements of the size of the pupil, gaze locations and the time length of fixation. They consist of a video camera and infrared illuminators, positioned in front of an eye, used to track the movement of the eyes which are then mapped into real-world coordinates by camera calibration.

The gaze location and the time length of fixation obviously show which features of the image the user is looking at and can reveal, for example, which features attract "the eye", which features are missed, and identify the point of visual fixation. However, since purely looking at an image feature does not necessarily mean that the feature attracted attention and caused the desired cognitive reaction, relating pupil dilation to cognition is a promising research direction.

Numerous studies in psychology have demonstrated that changes in pupil dilation consistently occur during the cognition process, including reading, visual search, and problem solving, as well as valence, arousal, pain, etc. (see [187–196]).

Extracting useful information from raw pupil dilation signals is not a trivial task due to high measurement noise of the commercial camera-based eye trackers, distortion due to gaze angle [187], frequent eye blinking, effects of illumination changes, irregular time delays in pupil response to stimuli, as well as the fact that pupil reaction is caused by different factors, which are hard to separate.

A set of simple experiments was designed based on a commercially available eye tracker, and the generated data were used in order to evaluate several statistical and signal processing tools for extracting useful information from pupil dilation signal when a user is presented with a sequence of coloured images. The obtained results show that mean, peak and variance are insufficient to capture the dynamics of pupil dilation change and therefore frequency-based and wavelet-based analysis is proposed,

for defining and extracting features that can be further used for clustering or pattern matching.

Pupils respond to different stimuli, including pain, emotional reaction, mental workload, and arousal with a very uneven reaction time delay and intensity that depends on the intensity and type of stimuli. The reaction delay can range from 0.1sec (mainly in the case of pain) to 2-7 seconds for emotional stimuli [187]. For example, [188] investigated reaction to sound stimuli, and observed that only after 400ms the pupil starts to sharply dilate, reaching a peak 2-3sec after the stimuli.

Beatty [192] concluded that pupil dilation is a good representation of the difficulty and amount of mental workload across tested subjects and cognitive tasks. For example, calculating 16×23 causes 10% larger pupil dilation than 7×8, or memorising a 3-digit number and 7-digit number caused 0.1mm and 0.55mm of pupil dilation, respectively (see [187]).

Different measures, such as mean dilation, peak dilation, variance, as in [189], and response time, have been proposed to quantify pupil dilation as a response to cognitive tasks (see [187] and references therein). However, these measures are not very robust and often not informative enough.

The key challenge in extracting meaningful information from pupil dilation lies in distinguishing the exact cause of pupil reaction. The pre-processing task needs to remove distortion and camera noise as well as natural blinking and pupillary light reflex. Indeed, as a reaction to brightness change, pupils naturally dilate, which can significantly affect measurements. To mitigate this problem, in [190] and [191], *principal component analysis* (PCA) is used on the pupil dilation data. Another approach can be found in [193], where a Hilbert transform method was used in order to study cognitive overload and cognitive dissonance. Though the initial results show potential, they are not conclusive, according to the authors, requiring further studies.

[195] proposed a measure called *Index of cognitive activity* (ICA), that represents the average number of "abrupt" changes in pupil size per second. This was estimated using wavelet decomposition, a technique that proved capable of filtering out the change of brightness effect.

Building on this work, Marshall in [196] compared pupil dilation with other measures such as blink rates, fixation time, saccade distance and speed, during different tasks,

such as driving a car and visual search, in order to identify the best combination of measures for assessing the cognitive state of a subject. This research has proposed a combination of seven "eye metrics", left and right index, left and right blink, left and right movement, and divergence.

All these methods are either limited in conclusions they can make due to a restricted extracted feature space, or require additional measurands. Despite the fact that the pupil dilation signal has been studied for long (see [187–196]), there are still many unknowns w.r.t exploiting pupil dilation analysis for assessing multimedia experience, and appropriate signal processing and machine learning tools are needed to make the information extraction process fast and automated, to open the door for real-time visual feedback design mechanisms.

6.3.2 Methodology

Experiments were designed to examine the gaze position and pupil responses to different, "neutral-content" images (indoors and outdoors) with and without searching for a specific target. The "neutral-content" images were selected as the research's aim was to examine pupil activity that is not a product of emotional triggering (pleasant/unpleasant), which is expected to create a more intense response. Through the experiment, it would be also possible to assess the effect of "busy" indoor images versus less busy outdoor images.

The eye tracker used during all experiments is the Tobii X2-60 Eye Tracker [197], which provides pupil dilation and gaze fixation data with a 60Hz sampling rate. Data collection and stimuli presentation were obtained using Ogama Version 4.5 [198].

6.3.2.1 Experimental Setup

Experiments were performed in a laboratory environment using moderate artificial light conditions, which were kept constant for the duration of all trials. Ten subjects who participated in the experiment ranged between 25 and 50 years old, both male and female, either with normal or corrected-to normal vision. The subjects did not have previous knowledge of the stimuli presentation and did not receive any training, before the initial trial.

The subjects were sitting in front of a screen with a resolution of 1920x1080 pixels,

at a $\sim$ 70 cm distance (35° angle from the eye tracker). Calibration was performed using Ogama's calibration process, where a coloured dot was moving in the corners and centre of the screen and the subjects were asked to gaze at the dot. This process was performed before each trial.

6.3.2.2 Stimuli

The stimuli for this experiment were four high quality images that were acquired from Flickr under Creativity Commons Licences and are shown in Figure 6.7. Two of the images represent the outside of residential properties and the other two are indoor bedroom images [199–202].

Outdoor images did not resemble the architecture of the area where the test subjects reside. Between each image, and at the beginning and end of the presentation, a whole grey-coloured blank image of the same size was used to separate each different stimuli. Each image (stimuli or grey) was shown for 10 sec.

For each subject, the experiment comprised three trials, which took place at the same time and place, with a gap of less than a minute between the two trials. During the first trial (Trial 1), the subjects were shown the stimuli presentation, and were asked to watch the presentation with no further instructions. After the end of the trial, the same stimuli was shown for the second time (Trial 2), with no further instructions. The motivation for having two identical trials, one after another, is to see how the subjects react when they are shown a stimuli that they have already familiarised themselves with.

During the final trial (Trial 3), all subjects were requested verbally to locate the flower(s) in the images at the beginning of the trial, without doing any task when this occurred or without verbally suggesting that they had identified the target.

6.3.2.3 Pre-processing

During pre-processing stage, both pupil dilation and gaze data were cleaned by removing all eye blink artifacts, as in [189, 191]. Eye blink artifacts are usually represented in the data either with missing or zero data, or with abnormal values, such as negative pupil dilation values. All missing data were replaced using data interpolation from both right and left pupil data using linear interpolation.

(a) Content Image 1 [199]

(b) Content Image 2 [200]

(c) Content Image 3 [201]

(d) Content Image 4 [202]

Figure 6.7: Content images used as a stimuli during the experiments.

Three different types of signal processing analysis were performed on the processed data: (1) statistical analysis using dilation mean, variance, and peak; (2) frequency analysis; and (3) wavelet-based analysis.

6.3.2.4 Harmonic Analysis

Harmonic Analysis is performed by using the Welch Method [203], which is a commonly used method for estimating the *Power Spectral Density* (PSD) of a signal in the presence of noise. It splits the data into overlapping segments, computes modified periodograms of the overlapping segments and then averages them in order to estimate the PSD and mitigate effects of random noise. In the applied procedure, the signal is segmented into eight sections of equal length, each with 50% overlap. All remaining signal parts that cannot be included into these eight segments are discarded. Each segment is windowed with a Hamming window of the same length as the segment. Hamming windows are commonly used in digital signal processing and they provide good cancellation of the

first side lobe [204].

6.3.2.5 Wavelet-based Analysis

After interpolation and prior to wavelet-based analysis, to mitigate the effect of random measurement noise, the pupil dilation signal was filtered by a 5th order low-pass Butterworth filter with a cutoff frequency of f_c =4Hz, which was selected as in [189] and [191], since the pupil servomechanism's break frequency is roughly 2Hz (see [191] and references therein).

After filtering, *Discrete Wavelet Transform* (DWT) decomposition of the signal and wavelet de-noising using soft thresholding [205] is performed. This is done by passing the signal through a low-pass and a high-pass filter, and then down-sampling the filtered signals in order to remove the over-completeness of the transform coefficients.

Due to the properties of DWT, the energy of the transformed signal is concentrated in only few DWT coefficients that have high magnitudes, and the energy of the noise is spread across a large number of DWT coefficients that have low magnitudes. Wavelet De-noising by Soft Thresholding [205] can be applied to remove the remaining noise in the 0-4Hz band, by minimising *mean square error* (MSE) of the reconstructed signal compared to the original signal under the constraint that with high probability the reconstruction is at least as smooth as the original. This allows for the removal of undesirable noise ripples or oscillations that would not be removed with a simple MSE minimisation. The idea of wavelet de-noising by soft thresholding is to first decompose the noisy signal into N levels using a pyramidal wavelet filter, and then apply thresholding on the wavelet coefficients coordinate-wise with a specially selected threshold.

All DWT coefficients whose absolute value is less than the predefined threshold are set to zeros and all remaining coefficients will have magnitude reduced by the applied threshold. In the proposed method, *Minimax thresholding* is used and the level of noise is estimated based on the first level coefficients. Finally, the inverse transform is applied to recover the original signal. A similar wavelet de-noising procedure was used in [195] to evaluate the level of cognitive activity based on pupil dilation.

6.4 Results and Discussion

The filtered pupil dilation signal was initially separated into image segments and then all content image segments and all grey image segments were concentrated, forming in this way two signals: a signal carrying four content images and a signal carrying grey images. Figure 6.8 shows the filtered pupil dilation signal for Subject 5 and 10 during all three trials for the content images. Each vertical line represents the temporal transition between images. These subjects were selected as a representative example of all subjects.

The pupil dilation range was around 1mm for both subjects and it dynamically changed during the experiments. By reviewing the Figures 6.8(a)-(b), it is apparent that the dilation on average is higher during Trial 3, compared to the other two trials.

Figure 6.9(a)-(b) show the 3D gaze fixation (X,Y) versus time, i.e, the sample number, for Subjects 5 and 10 during Trial 3, where the subjects were asked to locate "flower(s)" in the images. Figures 6.10(a)-(b) and 6.11(a)-(b) illustrate the gaze position for both subjects for each content image. It is apparent by looking at the original images in Figures 6.7 that clusters of gaze points are located in the image areas with flowers. Indeed, for Image 2, in Figure 6.10(b), it is apparent that Subject 5 was able to target the living flowers. On the other hand, Subject 10 noticed many artificial flowers in the image. Note that there was no specification about the type of flower(s) the subjects should locate. Similarly, Images 1 and 3 have multiple flowers, hence the multiple clusters of gaze fixation points. In Image 4, both subjects were able to locate the vase on the table. The 3D graphs provide extra information of when the different targets were identified, spread out across the period the image(s) were shown.

Tables 6.3 and 6.4 show the Mean values and Variance of pupil dilation when viewing the Content Images, after filtering. From Table 6.3, we can see that predominately the highest values (bold) of mean pupil dilation were during Trial 3, which is expected.

In general, though mean, variance and number of peaks rather irregularly change across the images and subjects, failing to capture signal transients. Thus, time averaging over the images does not, in this case, provide information that can be used to assess user's attention and experience, since the signal transition information is lost.

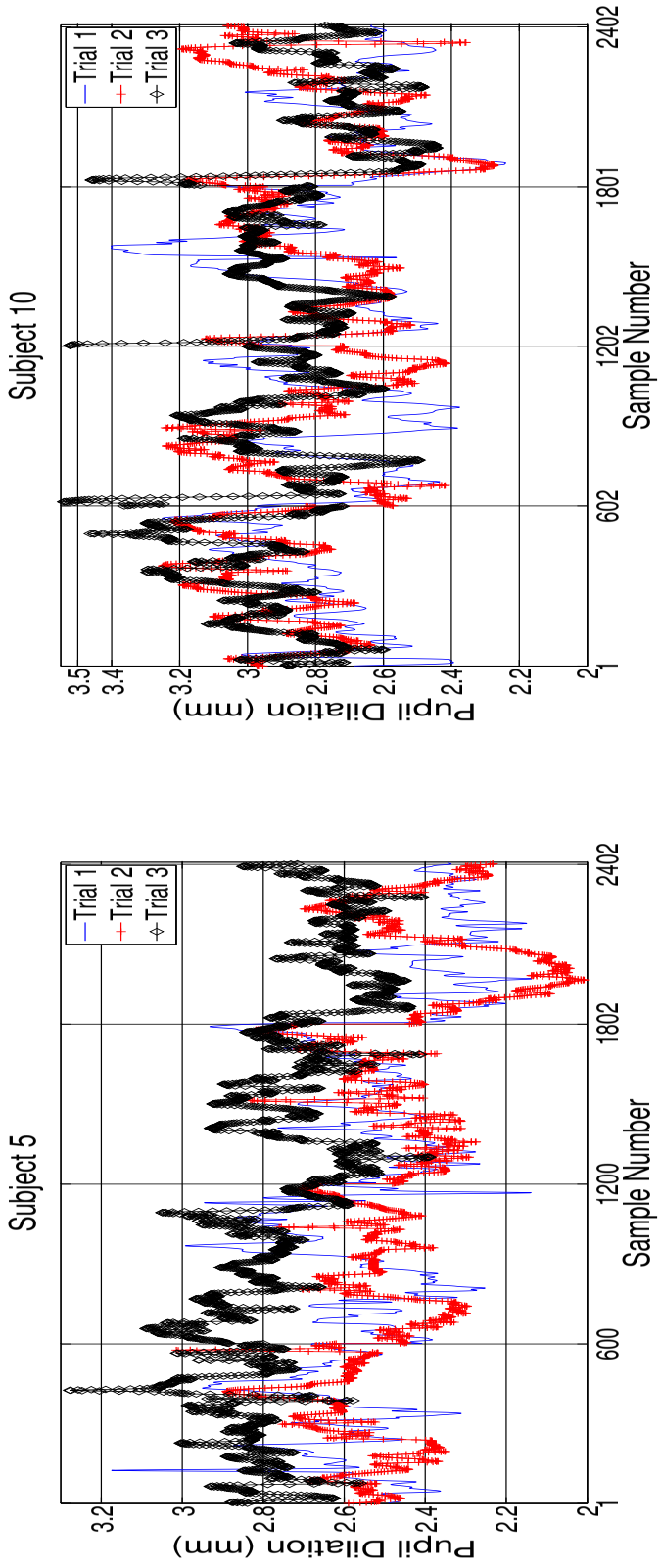
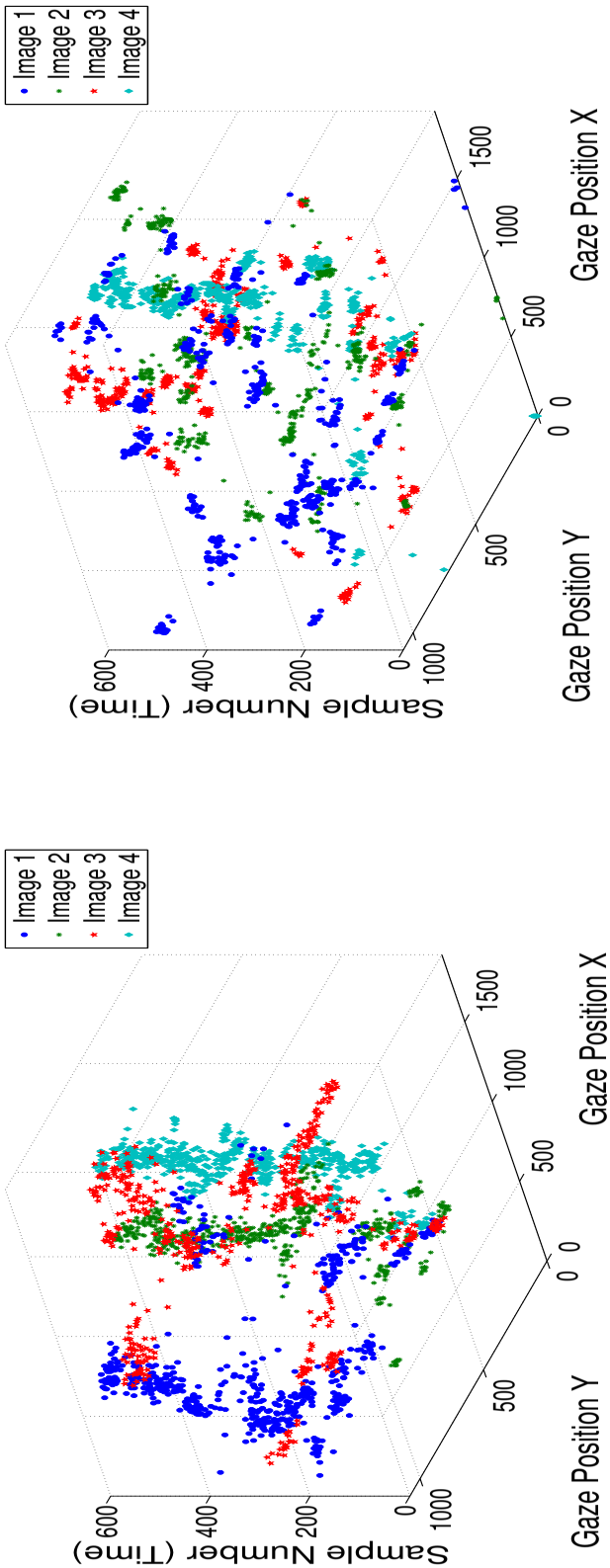


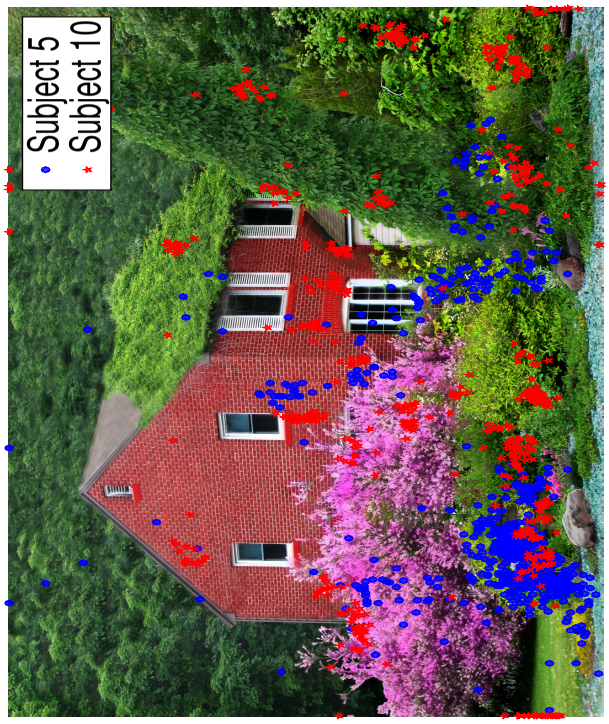
Figure 6.8: Filtered Pupil Dilation for Subjects 5 and 10 respectively during all three trials.



(a) 3D Gaze Position for Subject 5

(b) 3D Gaze Position for Subject 10

Figure 6.9: 3D Gaze Position for Subject 5 and 10 for all images during Trial 3.

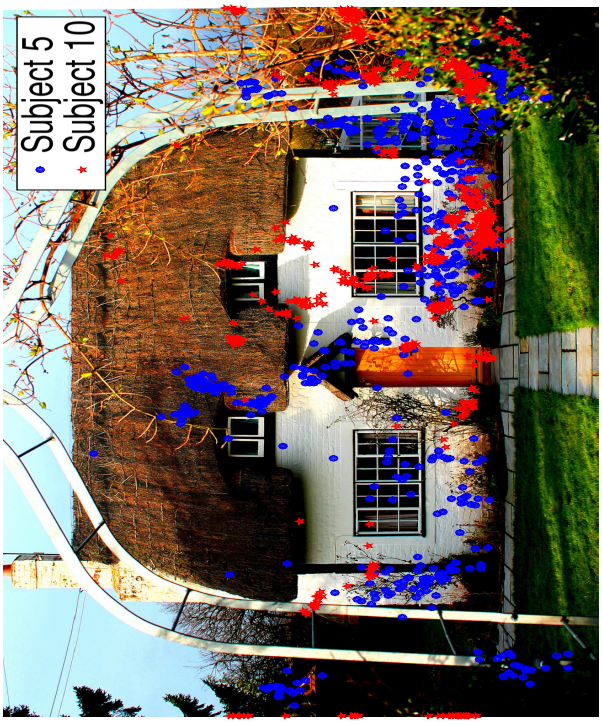


(a) Gaze Position for Content Image 1

(b) Gaze Position for Content Image 2

Figure 6.10: Gaze Position for Subject 5 and 10 for content images 1 and 2 during Trial 3.

(b) Gaze Position for Content Image 4



(a) Gaze Position for Content Image 3

Figure 6.11: Gaze Position for Subject 5 and 10 for content images 3 and 4 during Trial 3.

Table 6.3: Mean of the Filtered Pupil Dilation Signal of the Content Images [in mm]. S stands for Subject and **Bold** represents the highest mean value per image.

	Content Image 1			*Content Image 2*			*Content Image 3*			*Content Image 4*		
S	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*
1	2.180	**2.193**	2.123	2.215	**2.236**	2.213	2.135	2.227	**2.315**	2.111	**2.258**	2.246
2	**2.576**	2.500	2.546	2.606	2.751	**2.795**	2.479	2.583	**2.745**	2.458	**2.481**	2.417
3	3.104	**3.241**	3.215	3.312	3.103	**3.429**	3.091	2.991	**3.097**	**3.038**	2.849	2.958
4	**3.109**	2.631	2.749	**3.157**	2.823	2.843	2.779	2.699	**2.808**	2.642	2.644	**2.703**
5	2.550	2.520	**2.771**	2.528	2.442	**2.765**	2.477	2.428	**2.640**	2.275	2.270	**2.530**
6	2.578	**2.689**	2.673	2.645	2.665	**2.833**	2.609	2.683	**2.781**	2.499	**2.608**	2.586
7	2.614	2.544	**2.657**	**3.023**	2.653	2.872	**2.680**	2.627	2.668	**2.830**	2.612	2.542
8	**3.402**	3.204	3.374	3.481	3.185	**3.500**	3.209	3.050	**3.350**	**3.265**	3.139	3.221
9	3.131	3.059	**3.134**	**3.186**	3.159	3.032	2.929	3.029	**3.163**	2.804	**2.829**	2.707
10	2.757	2.873	**2.925**	2.639	2.739	**2.828**	2.770	2.743	**2.828**	2.546	**2.690**	2.653

Table 6.4: Variance of the Filtered Pupil Dilation Signal of the Content Images [in mm^2]. S stands for Subject and **Bold** represents the highest variance per image.

	Content Image 1			*Content Image 2*			*Content Image 3*			*Content Image 4*		
S	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*
1	0.119	**0.131**	0.109	**0.129**	0.118	0.121	0.118	0.130	**0.131**	0.111	0.131	**0.133**
2	**0.183**	0.169	0.177	0.182	**0.200**	0.179	0.148	0.164	**0.165**	**0.162**	0.147	0.141
3	0.237	**0.284**	0.266	0.285	0.237	**0.292**	0.229	**0.260**	0.244	**0.238**	0.199	0.207
4	**0.255**	0.168	0.182	**0.238**	0.191	0.193	**0.194**	0.179	0.192	0.168	0.170	**0.173**
5	0.169	0.160	**0.185**	0.176	0.147	**0.186**	0.155	0.150	**0.175**	0.124	0.151	**0.152**
6	0.156	0.173	**0.182**	0.165	0.164	**0.207**	0.165	0.173	**0.180**	0.160	**0.165**	**0.165**
7	0.173	0.165	**0.182**	**0.241**	0.179	0.200	0.177	0.177	**0.254**	**0.426**	0.176	0.155
8	0.298	0.243	**0.331**	0.318	0.250	**0.331**	0.246	0.225	**0.295**	**0.261**	0.243	0.259
9	0.286	**0.305**	0.252	0.267	**0.314**	0.260	0.237	0.268	**0.292**	0.190	**0.206**	0.178
10	0.212	0.214	**0.228**	0.193	**0.225**	0.214	**0.229**	0.198	0.187	0.170	**0.216**	0.190

6.4.1 Harmonic Analysis

Next, harmonic analysis is performed by estimating the PSD of the signal in the range of interest $(0 - 2Hz)$ using the Welch method described above. PSD was separately estimated for each grey image and each content image. The results for all content images are shown in Table 6.5. Frequency analysis shows the distribution of the power indicating in which frequency sub-band most of the signal's energy is concentrated. It could potentially indicate increased mental activity, if small enough time windows are applied.

This can be seen from Table 6.5, as on average Trial 3 shows increased power values per image, when compared with other trials. For example, Subject 5 in Image 1 shows 564.4W, which is higher compared to 477.6W and 468.5W for Trials 1 and 2, respectively. This pattern is similar to the case of mean values in Table 6.3 and generally power values are higher for Trial 3 when the subjects were asked to perform a target search task. This is mostly pronounced for Image 3 where all subjects show higher energy at Trial 3 (see bold values in Table 6.5). As frequency analysis loses time information and makes it difficult to conclude which time stimuli caused the reaction, therefore a wavelet-based analysis is proposed.

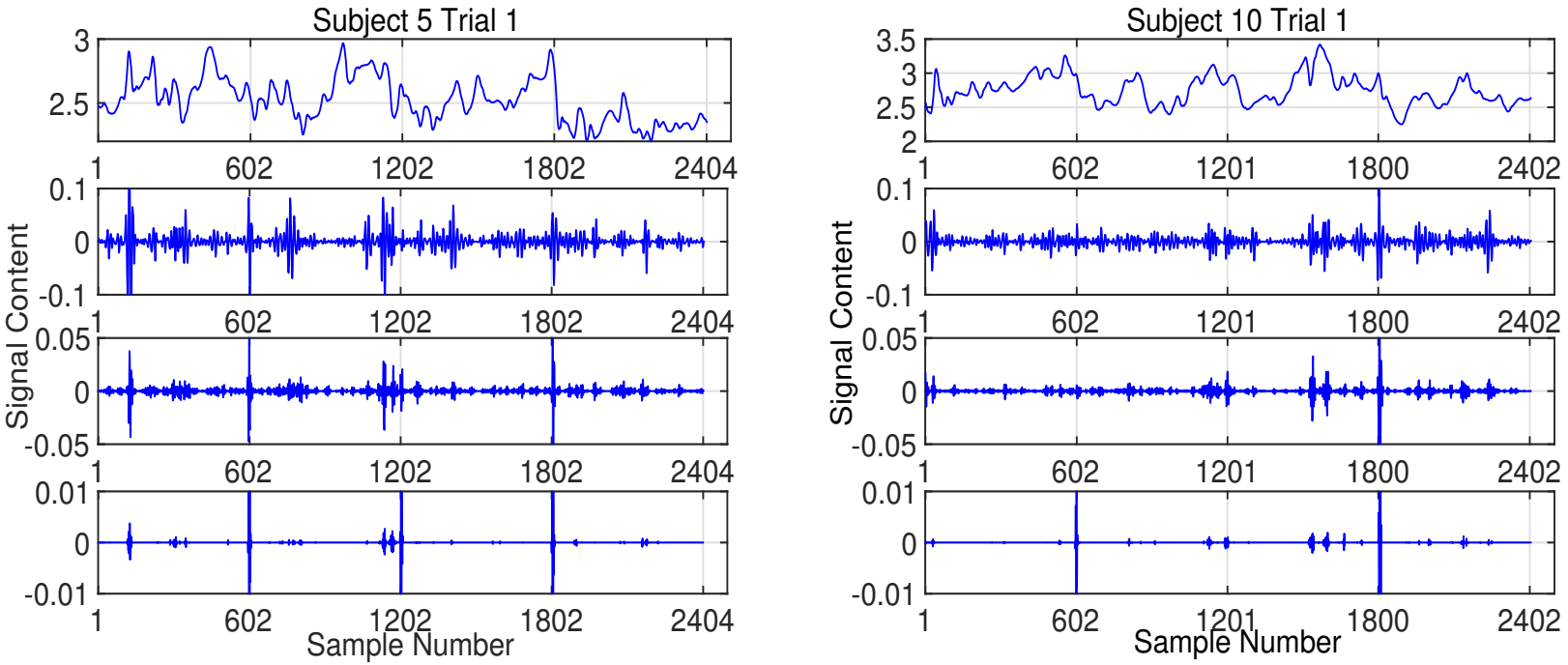
6.4.2 Wavelet-based Signal Processing

Figures 6.12(a)-(b), 6.13(a)-(b) and 6.14(a)-(b)show the wavelet-based analysis for Subjects 5 and 10 for all each trial respectively. Similar results are obtained for other subjects. Horizontal axis again shows the sample number with vertical lines pointing to the image transition moments; vertical axis denotes the right eye pupil dilation in mm.

A Daubechies-4 wavelet filter was used, as it is one of the most popular orthogonal wavelet filter with fast wavelet transform. In contrast to the Fourier analysis, DWT can tradeoff frequency and time resolution allowing for detection of time interval when a specific frequency component occurred. A 4-level wavelet decomposition was used, decomposing the signal into 4 frequency bands, to maintain high frequency resolution.

Table 6.5: Signal Power in the 0-2Hz Band in [W] per content image. **Bold** represents the highest power per image.

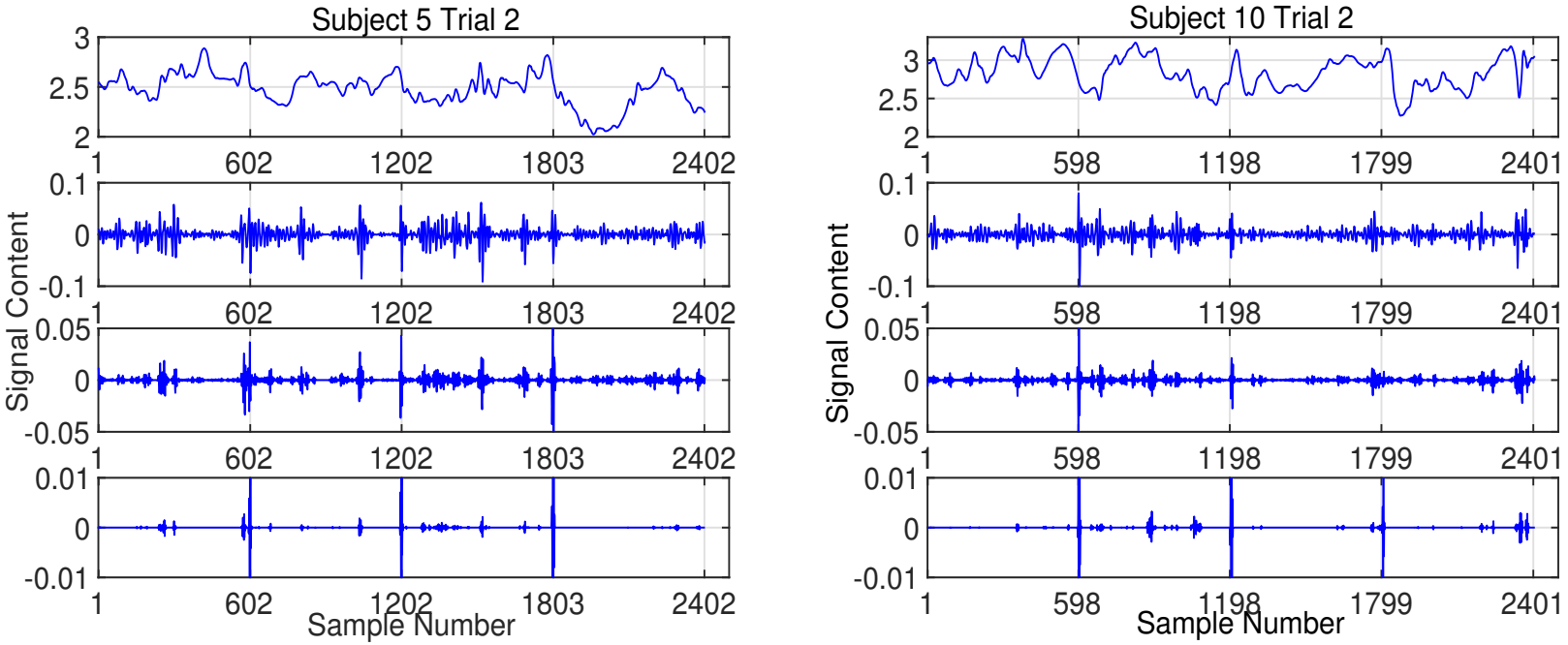
	Content Image 1			*Content Image 2*			*Content Image 3*			*Content Image 4*		
S	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*	*Trial 1*	*Trial 2*	*Trial 3*
1	348.8	**355.4**	335.1	362.2	**367.7**	359.5	336.4	365.5	**392.2**	329.8	**373.2**	373.2
2	490.2	461.8	**506.9**	502.6	**557.4**	520.4	451.1	490.2	**491.1**	447.0	**451.9**	428.9
3	707.9	**772.9**	763.1	802.5	708.6	**863.1**	701.6	663.8	**703.2**	**677.3**	596.9	642.6
4	**710.8**	510.2	554.5	**731.0**	585.7	592.4	567.6	538.0	**580.0**	512.7	516.2	**538.9**
5	477.6	468.5	**564.4**	470.0	436.5	**559.0**	451.8	432.9	**512.6**	379.1	379.8	**470.9**
6	490.2	**530.6**	523.5	513.5	521.3	**585.3**	500.6	529.3	**567.0**	458.0	**498.6**	491.9
7	502.2	476.9	**523.2**	**672.6**	516.4	602.9	528.8	507.2	**531.7**	**620.3**	502.1	474.5
8	**855.8**	756.3	852.0	891.8	749.3	**910.0**	757.0	687.1	**823.2**	**782.6**	721.4	771.4
9	**726.8**	695.3	722.2	**746.2**	742.2	676.3	636.3	678.7	**736.3**	579.8	**590.6**	540.5
10	561.0	605.3	**624.4**	517.0	551.2	**588.4**	567.6	554.4	**587.6**	477.4	**539.6**	519.6



(a) Subject 5 Trial 1

(b) Subject 10 Trial 1

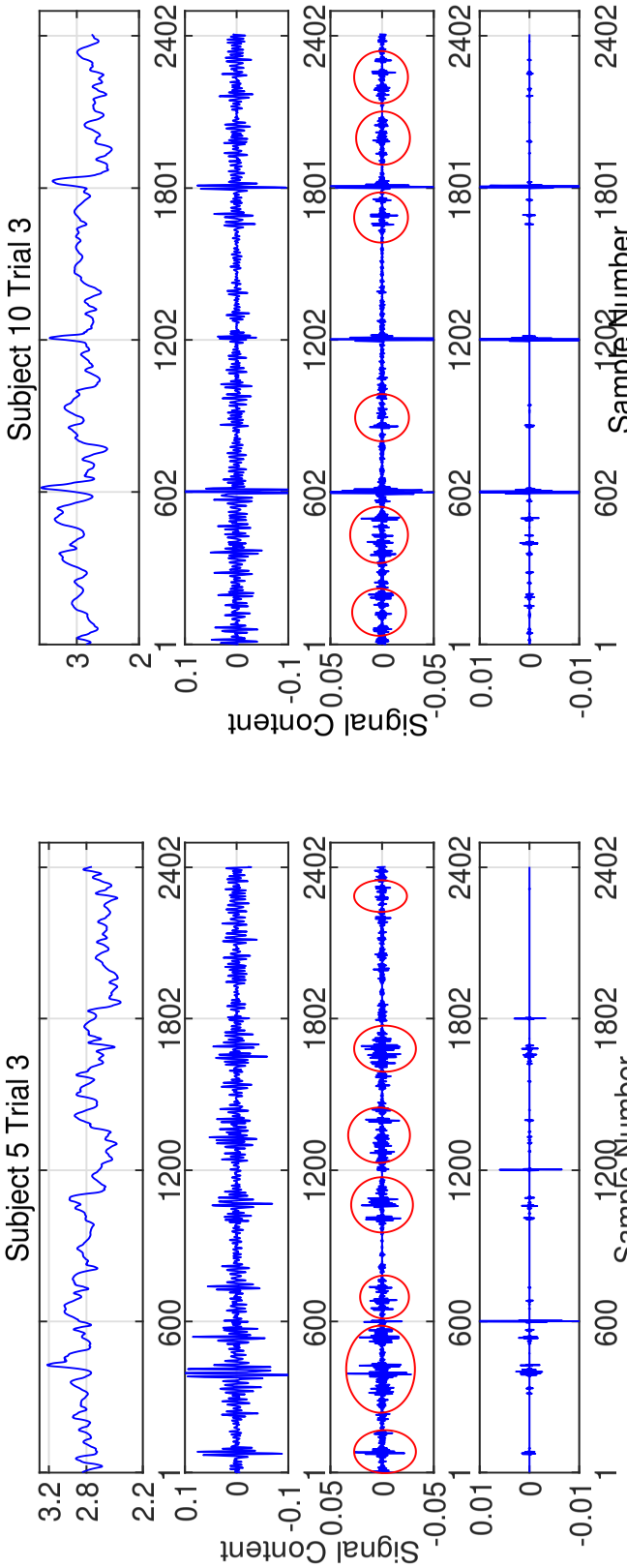
Figure 6.12: Wavelet Analysis of Content Images for Subject 5 and 10 for Trial 1.



(a) Subject 5 Trial 2

(b) Subject 10 Trial 2

Figure 6.13: Wavelet Analysis of Content Images for Subject 5 and 10 for Trial 2.



(a) Subject 5 Trial 3

(b) Subject 10 Trial 3

Figure 6.14: Wavelet Analysis of Content Images for Subject 5 and 10 for Trial 3.

The areas around the image transitions should be ignored as they are caused by signal stitching. An increased activity can be clearly observed in the band of interest. Trial 3 is characterised by significant pupil dilation activity in the beginning (until the task is solved, bearing in mind delayed reaction) and then reduction, while the other two trials have more evenly spread activity. Trial 1 has evidently more activity than Trial 2, since new content was presented in Trial 1. This activity has been marked for both subjects in Figures 6.14(a) and (c) with red circles, where activity is more clear. In Trial 3, the pupil dilation activity indicates that Image 2 was most challenging, which is true since the flower position is not so obvious.

In general, video-based eye trackers can be incorporated in the IoT concept, similarly with RFID tags and sensors, (see Appendix C). Gaze fixation and wavelet analysis of pupil dilation, compared to other signal processing methods presented throughout this case study, have been proven suitable for assessing user's attention. Therefore, they could be used as features for identifying patterns in the user's content interaction behaviour using data mining techniques. Data pre-processing included removal of eye blink artifacts and replacing missing values using linear interpolation of both left and right eye data.

6.5 Summary

The aim of this chapter was to investigate different data monitoring, data analytics and signal processing tools, that can be implemented in different IoT applications with diverse output data. For this purpose, two case studies related to different disciplines have been presented.

The first case study proposed a cost-effective prototype sensor solution for monitoring earthworks. The presented setup can measure soil resistivity, ground movement and pressure, but allows the incorporation of other sensors. The obtained results show expected resistivity values for the weather condition and soil material at the deployment site. Ground movement sensor sensitivity was proven and can be used in a future destructive test that could provide the profile of a healthy and failing earthwork. Finally, a further calibration of the absolute pressure with a pore pressure sensor would provide a cost-effecting alternative of the current methods of measuring pore pressure.

The proof of concept of the proposed monitoring system and the inferred results could allow a future implementation of sensors' arrays in order to produce a full landscape survey and therefore induce ERT applications.

The second case study discussed different signal processing methods for analysing pupil dilation signals with applications to multimedia experience assessment. The main goal was to review and test different methods, in order to evaluate their future use for feature definition and extraction for autonomous pattern matching and event detection. More specifically, mean, variance and number of peaks of the pupil dilation have been proven to be inadequate to provide clear patterns that would help assessing user's attention. On the other hand, harmonic analysis obtained similar results with the case of mean values, with higher power values on average during Trial 3, but it is difficult to identify when this reaction occurred solely using frequency domain data.

As in [195], the presented findings show that clearly wavelets provide a clearer view of activity on pupil dilation and can be essentially used as a helping tool for extracting signatures from the pupil dilation signal in order to relate each segment to image information/task for automated pattern matching. A possible combination of the above methods could provide a more accurate activity recognition process.

Both case studies make apparent that although the main data analysis steps, such as data pre-processing, are similar, the further analysis requires a more in depth knowledge of the features monitored in each application. Additionally, depending on the sampling rate, useful information can be extracted through signal processing and pattern matching in both time and frequency domain.

Finally, in the following chapter, Chapter 7, the main findings of the research work presented in this thesis will be presented, with respect the research questions defined in Chapter 1. Furthermore, a further discussion will be included regarding any possible limitations of the presented research and solutions will be presented in the context of a further research.

Chapter 7

Summary and Implications

7.1 Introduction

The realisation of Internet of Things will provide various benefits in our communities by automating processes and assisting even with the simplest everyday tasks, such as driving a car, visiting a hospital or utilising the capabilities of a smart house. Together with every positive effect, IoT will instigate various challenges for researchers, companies and lawmakers, in the context of data security and privacy, data storing and data utilisation and analysis. An enormous amount of data would be produced simultaneously, originating from smart devices and sensors incorporated in various and heterogeneous fields. One of the most interesting challenges, which was the main motivation behind this research work, will be the extraction of meaningful information from the available data and the application appropriate data analytics and signal processing methods that can be applied for this purpose.

This chapter will provide a summary of the research outcomes with regards to the stated motivation and the research questions, defined in Chapter 1. Any observations and limitations of the proposed methods and solutions will be further analysed, in order to lay the foundations for any future research that could optimise and advance further the research work presented in this thesis.

7.2 Research Findings

The main findings of this research work are associated with the smart grid and smart metering application of IoT from the perspective of Non-Intrusive Appliance Monitoring, where an extended background research has been performed in order to identify both the advancements and the limitations of the NILM research field. Three unsupervised

methods were proposed as a potential solution of the NILM problem and a performance evaluation has been presented with benchmarks state-of-the-art NILM methods using publicly available datasets, namely REDD [1] and REFIT [2, 3].

Furthermore, two other IoT applications have been presented in Chapter 6 .The first application includes the monitoring of large infrastructures, specifically a canal embankment, where a prototype monitoring system was proposed and a proof of concept has been reported. The second case relates to the assessment of human interaction with visual content, where various statistical and signal processing techniques have been investigated for extraction of useful information.

7.2.1 Low Resolution Non-Intrusive Load Monitoring

Through the extended NILM background review, available in Chapter 2, it has become apparent that the vision of an online and real-time load disaggregation, require NILM techniques, that can accurately identify appliance usage using solely aggregate power data at low sampling rates, similar to the standards of the vast majority of the commercial smart meters deployed by utilities and governments around the world. Therefore, the main focus in both Chapters 3 and 5 was to propose unsupervised NILM techniques for low resolution disaggregation, using as a feature only the information available from the smart meter and without having a priori knowledge of the appliance operations through individual monitors. Throughout this thesis, the IAMs data have been only incorporated for naming the already obtained appliance signatures and to evaluate the accuracy of the proposed methods with regards to the ground truth operation of each appliance.

In Chapter 3, an unsupervised method, using solely Dynamic Time Warping (DTW), has been proposed. The proposed method obtains offline a library of appliance signatures for each operating appliance using the aggregate load. These signatures are used during disaggregation in order to classify the obtained disaggregation result. One of the main advantages of this method is that allows the incorporation of new appliances, by only learning for the new appliance, to obtain the relevant signature and compare it to the available appliance signatures library. Thus, it does not require relearning of the appliances already present in the household. In Chapter 4, the proposed DTW-based method was evaluated using five houses from publicly available NILM datasets, three

from the US-based REDD dataset [1], downsampled to 1min resolution, and 2 from the UK-based REFIT dataset [2, 3], with benchmarks the state-of-the-art NILM methods, as presented in the works of Zhao *et al.* [4, 5] and He *et al.* [33], which include a variety of supervised and unsupervised Graphical Signal Processing based methods, Decision Tree [32], HMM [35] and FHMM [34].

The proposed method was able to obtain, in most houses, higher or comparable performance in comparison to most of the methods used for benchmarking. It has been proven more successful for disaggregating on/off appliances, especially with high consumption, e.g. kettle toaster, and found more challenging the classification of multi-state appliances, such as dishwasher, and washing machine, but was still able to identify the high consuming washing cycle with comparable accuracy in most cases. Although, the proposed DTW-method obtained overall good performance, an increased computational cost has been reported.

In Chapter 5, two methods have been proposed for reducing the computational complexity of the original DTW-based method, with both performing unsupervised learning, similar to the method proposed in Chapter 4. The first method, referred as DTW+kM, creates the library of appliance signatures using DTW similarly with the original implementation and initialises the cluster centroids that will be used by k-means in order to perform clustering and classification of the detected aggregate load. k-means increased significantly the classification speed, and was able to report in most cases similar performance with the original DTW-based method, for most houses, with the exception of some appliances, such as dishwasher and washing machine, where it has merged the original clusters into one due to similarity of the average consumption of the appliances. Overall the DTW+kM method has shown sufficiently good performance, and it was the method with the lowest execution time.

The second method, namely kDTW, was introduced in order to provide a further refinement of the proposed DTW+kM method. The refinement method is using the DTW method for the non-identified and the falsely identified operations, by selectively comparing the corresponding events to the appliance signatures with the closest average consumption. The implementation increased the computational complexity of the DTW+kM method, but still was significantly faster than the proposed DTW-based method in Chapter 4. Overall the refinement has been proven successful, and it was

able to obtain the same accuracy in most houses with the original method, therefore comparable to most of the state-of-the-art NILM, presented in [4, 4, 33].

7.2.1.1 Large Infrastructures Monitoring

Earthworks infrastructures require regular monitoring and maintenance, as a potential failure could affect the smooth operation of railways, and canals, to name the least, and could even lead to fatal accidents. This monitoring process usually involves expensive monitoring systems and human involvement for operating the monitoring systems. In Chapter 6, a prototype and cost-effective solution has been proposed for automated monitoring of a canal embankment, and generally earthwork assets, for assisting in proactive maintenance of the asset. Through background research and expert knowledge from the stakeholders, the suitable monitoring parameters have been identified, namely soil resistivity, pressure and ground movement and a customisable sensor node, with prototype casing, has been designed and deployed for monitoring the allocated site in Falkirk Wheel.

Data pre-processing has been performed in order to synchronise the data using the injection and sinking of the current process, and data imputation was performed using linear interpolation. The obtained resistivity was in line with the expected values of the soil type present in the testing site, and due to the normal weather conditions, there were not any extreme variations. An artificially induced disturbance has shown that the ground movement/vibration could provide a potential feature for capturing the failure of an embankment. Pressure has reported only small variation of $\sim 2kPa$, therefore in order to obtain useful information from the pressure, a further calibration using a moisture content sensor is required. In general, we were able to show the potential of monitoring resistivity, 3-axis movement and pressure, all in one prototype and customisable sensor node, compared to the off-the-shelf solutions, that require the use of multiple and expensive monitoring systems, in order to obtain the different parameters.

7.2.1.2 Attention Assessment in Visual Content Interactions

The assessment of the user's attention when interacting with visual content, can be useful in various IoT applications, from healthcare to even personalised advertisement based on the visual interaction of the user when using a computer embedded with an eye tracking device. In Chapter 6, a three-stage experiment has been designed and executed in order to assess the attention of the test subjects with respect to the visual content presented, in the specific case "neutral-content" images (indoors and outdoors), with and without the search of a target. Data pre-processing has been performed for removing eye blink artifacts and linear interpolation was implemented for recovering any missing data. Three different signal processing approaches have been investigated, namely statistical analysis using pupil dilation mean, variance, and peak, frequency analysis and wavelet-based analysis, in order to evaluate their suitability for extracting useful information from the pupil dilation and gaze position data.

The obtained results have shown that mean, variance and number of peaks did not provide useful information and were not able to depict the signal's transient. Harmonic analysis has reported increased power for most of the subjects during the third trial, where the subjects were asked to search for a target, but it does not allow the correlation of the increased power to the time stimuli that triggered this reaction. For this purpose, a wavelet-based method has been proposed using a Daubechies-4 wavelet filter, to decompose the signal into 4 frequency bands, but still maintain high frequency resolution. The reported results have shown an increased activity in terms of pupil dilation during the trial, where the subjects were performing the requested task, for reference trial 3. Therefore, this wavelet approach together with the gaze position information have been proven successful in identifying activity that could be potentially used as an extracted feature, in order to further analyse and classify the human attention behaviour and maybe correlate this behaviour to emotions that the visual content could create towards the user.

7.3 Research Limitations

Even though this research work has achieved the original research aims and objectives, it is important to identify and discuss any limitations that have been observed throughout this thesis. These limitations can be summarised as follows:

- The proposed DTW-based method, as already discussed, has shown high computational cost, thus it is not a suitable candidate for online and real-time disaggregation. This has been addressed with the use of the combined k-means and DTW scheme, which was able to reduce significantly the processing time of the classification.

- The proposed DTW-based method was not able to obtain signatures for all appliances, especially low consuming ones, and various states of multi-state appliances, therefore the reported disaggregation results in some of the test houses included less appliances compared to the benchmark methods.

- All proposed methods have a difficulty in successfully separating the washing machine and dishwasher in House 2 from REFIT dataset, which consume similar average power during their washing cycle, which is the state that the proposed methods can disaggregate.

- The event detection method, using the "rising and falling edges", does not allow the extraction of all the appliance operation events present in the aggregate data, that could be potentially classified using the proposed methods.

- The DTW refinement was not able to further increase the performance of the original DTW-based method.

- The resistivity measurements obtained, using the prototype monitoring system in Chapter 6, have reported synchronisation issues, which was challenging in terms of pre-processing the data.

- The deployed monitoring system included only four nodes, in order to prove the concept of resistivity and the rest measured parameters, which although successful it terms of understanding the usefulness of the monitored measurands, is not

sufficient for understanding in full extent the behaviour of the embankment. This would require the implementation of a full array of sensors and a destructive test, which was not the purpose of this research.

- The use of "neutral-content" visual content, although it was specifically selected to serve the objectives of this thesis, makes it difficult to identify intense pupil dilation variations, which would be more evident when using content that could trigger emotional reactions.

7.4 Further Research

The potential future research, as an extension of this thesis, would include the following:

- Investigation of alternative edge/event detection methods in order to overcome the limitations of the "rising and falling" edge detection method, used throughout this thesis, increase the quality of the detected events, acquire the complete signature of multi-state appliances, such as dishwasher and washing machine, and overcome more successfully the overlapping events. Evaluation of the complexity and performance of the proposed DTW-based, DTW+kM and kDTW methods.
- Implementation of the proposed DTW refinement, as a post-processing tool, using other state-of the-art NILM methods, for evaluating the potential application for improving the disaggregation process for other event based NILM methods.
- Implementation of a firmware that will allow the precise synchronisation of the resistivity measurements, in order to overcome the synchronisation issues observed during the original deployment.
- Performance of a destructive testing in order to acquire the "profile", in terms of measurements, of a failing earthwork asset, and perform data analysis and predictive modelling. Investigation of the potential use of DTW for classification of the obtained features.
- Deployment of the prototype monitoring system in a large-scale array, in order to monitor the resistivity for performing 2D or 3D mapping of the earthwork asset through ERT.

- The research and development of a more compact casing design, that would allow easier and less intrusive deployment of the sensor nodes.

- Performance of a more visually challenging experiment, in order to acquire distinct features using the proposed wavelet-based signal processing approach and evaluate appropriate signal processing and machine techniques for classification of the user's attention behaviour. Investigation of the potential use of DTW for the classification purpose.

Bibliography

[1] J. Z. Kolter and M. J. Johnson, "REDD: A Public Data Set for Energy Disaggregation Research," in *SustKDD Workshop on Data Mining Applications in Sustainability*, 2011.

[2] D. Murray, J. Liao, L. Stankovic, V. Stankovic, R. Hauxwell-Baldwin, C. Wilson, M. Coleman, T. Kane, and S. Firth, "A data management platform for personalised real-time energy feedback," in *Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting (EEDAL)*, pp. 1–15, August 2015.

[3] D. Murray, L. Stankovic, and V. Stankovic, *An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study*, vol. 4, pp. 160122 EP –. The Author(s) SN -, Jan 2017.

[4] B. Zhao, L. Stankovic, and V. Stankovic, "On a training-less solution for non-intrusive appliance load monitoring using graph signal processing," *IEEE Access*, vol. 4, pp. 1784–1799, 2016.

[5] B. Zhao, K. He, L. Stankovic, and V. Stankovic, "Improving event-based non-intrusive load monitoring using graph signal processing," *IEEE Access*, vol. 6, pp. 53944–53959, 2018.

[6] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-intrusive load disaggregation using graph signal processing," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2016.

[7] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787 – 2805, 2010.

[8] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645 – 1660, 2013.

[9] L. D. Xu, W. He, and S. Li, "Internet of things in industries: A survey," *IEEE Transactions on Industrial Informatics*, vol. 10, pp. 2233–2243, Nov 2014.

[10] I. Lee and K. Lee, "The internet of things (iot): Applications, investments, and challenges for enterprises," *Business Horizons*, vol. 58, no. 4, pp. 431 – 440, 2015.

[11] M. Yun and B. Yuxin, "Research on the architecture and key technology of internet of things (iot) applied on smart grid," in *2010 International Conference on Advances in Energy Engineering*, pp. 69–72, June 2010.

[12] T. Rymarczyk, E. Kozłowski, and G. Kłosowski, "Intelligent system for analysis and monitoring of flood embankments based on electrical impedance tomography, machine learning and internet of things," in *Preprints 2018*, Aug. 2018.

[13] N. Gondchawar and R. S. Kawitkar, "Iot based smart agriculture," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, pp. 838–842, June 2016.

[14] M. S. Mekala and P. Viswanathan, "A novel technology for smart agriculture based on iot with cloud computing," in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 75–82, Feb 2017.

[15] S. R. Prathibha, A. Hongal, and M. P. Jyothi, "Iot based monitoring system in smart agriculture," in *2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)*, pp. 81–84, 2017.

[16] A. T. Duchowski, "A breadth-first survey of eye-tracking applications," *Behavior Research Methods, Instruments, & Computers*, vol. 34, pp. 455–470, Nov 2002.

[17] P.-W. Lui, F. M. Lai, K.-C. Su, J. Y. Lin, H. W. Chi, J.-S. Wang, and Y. W. Chen, "Use eye tracker to design an intelligent patient bed," *Energy Procedia*, vol. 143, pp. 553 – 558, 2017. Leveraging Energy Technologies and Policy Options for Low Carbon Cities.

[18] M. Eriksson and N. P. Papanikotopoulos, "Eye-tracking for detection of driver fatigue," in *Proceedings of Conference on Intelligent Transportation Systems*, pp. 314–319, Nov 1997.

[19] X. Liu, F. Xu, and K. Fujimura, "Real-time eye detection and tracking for driver observation under various light conditions," in *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 2, pp. 344–351 vol.2, June 2002.

[20] A. J. Adams, "Acute effects of alcohol and marijuana on vision," in *Frontiers in Visual Science* (S. J. Cool and E. L. Smith, eds.), (Berlin, Heidelberg), pp. 93–105, Springer Berlin Heidelberg, 1978.

[21] C. Hayashi, "What is data science ? fundamental concepts and a heuristic example," in *Data Science, Classification, and Related Methods* (C. Hayashi, K. Yajima, H.-H. Bock, N. Ohsumi, Y. Tanaka, and Y. Baba, eds.), (Tokyo), pp. 40–51, Springer Japan, 1998.

[22] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? the case of electricity," *Energy Policy*, vol. 52, pp. 213 – 234, 2013. Special Section: Transition Pathways to a Low Carbon Economy.

[23] G. Hart, M. I. of Technology. Energy Laboratory, and E. P. R. Institute, *Nonintrusive Appliance Load Data Acquisition Method: Progress Report.* MIT Energy Laboratory, 1984.

[24] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, pp. 1870–1891, Dec 1992.

[25] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. D. Abowd, *At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line (Nominated for the Best Paper Award)*, pp. 271–288. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[26] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, "Power signature analysis," *IEEE Power and Energy Magazine*, vol. 1, pp. 56–63, Mar 2003.

[27] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, "Enhancing electricity audits in residential buildings with nonintrusive load monitoring," *Journal of Industrial Ecology*, vol. 14, no. 5, pp. 844–858, 2010.

[28] H.-H. Chang, "Load identification of non-intrusive load-monitoring system in smart home," *WSEAS Transactions on Systems*, vol. 9, pp. 498–510, 05 2010.

[29] H. H. Chang, C. L. Lin, and J. K. Lee, "Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms," in *The 2010 14th International Conference on Computer Supported Cooperative Work in Design*, pp. 27–32, April 2010.

[30] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Transactions on Consumer Electronics*, vol. 57, pp. 76–84, February 2011.

[31] A. Zoha, A. Gluhak, M. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, pp. 16838–16866, Dec 2012.

[32] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Non-intrusive appliance load monitoring using low-resolution smart meter data," in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 535–540, Nov 2014.

[33] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-intrusive load disaggregation using graph signal processing," *IEEE Transactions on Smart Grid*, vol. 9, pp. 1739–1747, May 2018.

[34] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "Nilmtk: An open source toolkit for non-intrusive load monitoring," in *Proceedings of the 5th International Conference on Future Energy Systems*, e-Energy '14, (New York, NY, USA), pp. 265–276, ACM, 2014.

[35] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, AAAI'12, pp. 356–362, AAAI Press, 2012.

[36] S. Makonin, *Approaches to Non-Intrusive Load Monitoring (NILM) in the Home.* Vancouver, Canada: Technical Report. Simon Fraser University, 10 2012.

[37] J. Froehlich, E. Larson, S., G. Cohn, M. Reynolds, and S. Patel, "Disaggregated end-use energy sensing for the smart grid," *IEEE Pervasive Computing*, vol. 10, pp. 28–39, Jan 2011.

[38] S. Makonin, B. Ellert, I. V. Bajic, and F. Popowich, "Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014," *Scientific Data*, vol. 3, no. 160037, pp. 1–12, 2016.

[39] Nilm Wiki, "Nilm Datasets." http://wiki.nilm.eu/datasets.html, 2017. Accessed 6 June 2018.

[40] L. Pereira and N. Nunes, "Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools-a review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 6, p. e1265, 2018.

[41] A. G. Ruzzelli, C. Nicolas, A. Schoofs, and G. M. P. O'Hare, "Real-time recognition and profiling of appliances through a single electricity sensor," in *2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, pp. 1–9, June 2010.

[42] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng, "Load signature study #x2014;part i: Basic concept, structure, and methodology," *IEEE Transactions on Power Delivery*, vol. 25, pp. 551–560, April 2010.

[43] H. Y. Lam, G. S. K. Fung, and W. K. Lee, "A novel method to construct taxonomy electrical appliances based on load signaturesof," *IEEE Transactions on Consumer Electronics*, vol. 53, pp. 653–660, May 2007.

[44] S. Makonin and F. Popowich, "Nonintrusive load monitoring (nilm) performance evaluation," *Energy Efficiency*, vol. 8, pp. 809–814, Jul 2015.

[45] J. Peppanen, , S. Grijalva, and M. J. Reno, "Handling bad or missing smart meter data through advanced data imputation," in *2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, Sep. 2016.

[46] J. Kelly and W. Knottenbelt, "The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes," *Scientific Data*, vol. 2, pp. 150007 EP –, 03 2015.

[47] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," in *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS-12)* (N. D. Lawrence and M. A. Girolami, eds.), vol. 22, pp. 1472–1482, 2012.

[48] D. Egarter, V. P. Bhuvana, and W. Elmenreich, "Paldi: Online load disaggregation via particle filtering," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, pp. 467–477, Feb 2015.

[49] M. Aiad and P. H. Lee, "Non-intrusive load disaggregation with adaptive estimations of devices main power effects and two-way interactions," *Energy and Buildings*, vol. 130, pp. 131 – 139, 2016.

[50] M. Figueiredo, B. Ribeiro, and A. de Almeida, "Electrical signal source separation via nonnegative tensor factorization using on site measurements in a smart home," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, pp. 364–373, Feb 2014.

[51] C. C. Yang, C. S. Soh, and V. V. Yap, "A systematic approach to on-off event detection and clustering analysis of non-intrusive appliance load monitoring," *Frontiers in Energy*, vol. 9, pp. 231–237, Jun 2015.

[52] S. Pattem, "Unsupervised disaggregation for non-intrusive load monitoring," in *2012 11th International Conference on Machine Learning and Applications*, vol. 2, pp. 515–520, Dec 2012.

[53] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-markov models," *J. Mach. Learn. Res.*, vol. 14, pp. 673–701, Feb. 2013.

[54] K. Basu, A. Hably, V. Debusschere, S. Bacha, , and A. Ovalle, "A comparative study of low sampling non intrusive load dis-aggregation," in *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, pp. 5137–5142, Oct 2016.

[55] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, "Leveraging smart meter data to recognize home appliances," in *2012 IEEE International Conference on Pervasive Computing and Communications*, pp. 190–197, March 2012.

[56] K. N. Trung, O. Zammit, E. Dekneuvel, B. Nicolle, , and G. Jacquemod, "An innovative non-intrusive load monitoring system for commercial and industrial application," in *The 2012 International Conference on Advanced Technologies for Communications*, pp. 23–27, Oct 2012.

[57] J. Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," in *Proceedings of the 2Nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, BuildSys '15, (New York, NY, USA), pp. 55–64, ACM, 2015.

[58] K. Kumar, R. Sinha, M. G. Chandra, and N. K. Thokala, "Data-driven electrical load disaggregation using graph signal processing," in *2016 IEEE Annual India Conference (INDICON)*, pp. 1–6, Dec 2016.

[59] K. D. Anderson, M. E. Bergs, A. Ocneanu, D. Benitez, and J. M. F. Moura, "Event detection for non intrusive load monitoring," in *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society*, pp. 3312–3317, Oct 2012.

[60] L. Pereira, "Developing and evaluating a probabilistic event detector for non-intrusive load monitoring," in *2017 Sustainable Internet and ICT for Sustainability (SustainIT)*, pp. 1–10, Dec 2017.

[61] L. K. Norford and S. B. Leeb, "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms," *Energy and Buildings*, vol. 24, no. 1, pp. 51 – 64, 1996.

[62] M. Baranski and J. Voss, "Genetic algorithm for pattern detection in nialm systems," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, vol. 4, pp. 3462–3468 vol.4, Oct 2004.

[63] M.-S. Tsai and Y.-H. Lin, "Modern development of an adaptive non-intrusive appliance load monitoring system in electricity energy conservation," *Applied Energy*, vol. 96, pp. 55 – 73, 2012. Smart Grids.

[64] S. B. Leeb, S. R. Shaw, and J. L. Kirtley, "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," *IEEE Transactions on Power Delivery*, vol. 10, pp. 1200–1210, Jul 1995.

[65] D. Luo, L. Norford, S. Shaw, S. Leeb, R. Danks, and G. Wichenko, "Monitoring hvac equipment electrical loads from a centralized location - methods and field test results," in *ASHRAE Transactions*, vol. 108, pp. 841–857, 01 2002.

[66] M. Berges, E. Goldman, H. Matthews, and L. Soibelman, "Learning systems for electric consumption of buildings," in *Proceedings of the 2009 ASCE International Workshop on Computing in Civil Engineering*, vol. 346, 08 2009.

[67] Y. Jin, E. Tebekaemi, M. Berges, and L. Soibelman, "Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4340–4343, May 2011.

[68] K. Nguyen, E. Dekneuvel, B. Nicoll, O. Zammit, C. Van, and G. Jacquemod, "Event detection and disaggregation algorithms for nialm system," in 2^{nd} *International Non-Intrusive LoadMonitoring (NILM) Workshop*, 06 2014.

[69] L. Pereira and N. Nunes, "An experimental comparison of performance metrics for event detection algorithms in nilm," in 4^{th} *International Workshop on Non-Intrusive Load Monitoring (NILM)*, March 2018.

[70] C. Klemenjak and P. Goldsborough, "Non-intrusive load monitoring: A review and outlook," *CoRR*, vol. abs/1610.01191, 2016.

[71] J. O. Smith, *Transient and Steady-State Signals*. Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, 2017.

[72] L. Farinaccio and R. Zmeureanu, "Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses," *Energy and Buildings*, vol. 30, no. 3, pp. 245 – 259, 1999.

[73] M. Marceau and R. Zmeureanu, "Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings," *Energy Conversion and Management*, vol. 41, no. 13, pp. 1389 – 1403, 2000.

[74] J. Z. Kolter, S. Batra, and A. Y. Ng, "Energy disaggregation via discriminative sparse coding," in *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, NIPS'10, (USA), pp. 1153–1161, Curran Associates Inc., 2010.

[75] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proceedings of the 2011 SIAM International Conference on Data Mining*, pp. 747–758, April 2011.

[76] A. Marchiori, D. Hakkarinen, Q. Han, and L. Earle, "Circuit-level load monitoring for household energy management," *IEEE Pervasive Computing*, vol. 10, pp. 40–48, Jan 2011.

[77] M. Zeifman, "Disaggregation of home energy display data using probabilistic approach," *IEEE Transactions on Consumer Electronics*, vol. 58, pp. 23–31, February 2012.

[78] S. Drenker and A. Kader, "Nonintrusive monitoring of electric loads," *IEEE Computer Applications in Power*, vol. 12, pp. 47–51, Oct 1999.

[79] M. B. Figueiredo, A. de Almeida, and B. Ribeiro, *An Experimental Study on Electrical Signature Identification of Non-Intrusive Load Monitoring (NILM) Systems*, pp. 31–40. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[80] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito, "Nonintrusive appliance load monitoring based on integer programming," in *2008 SICE Annual Conference*, pp. 2742–2747, Aug 2008.

[81] T. Kato, H. S. Cho, D. Lee, T. Toyomura, and T. Yamazaki, *Appliance Recognition from Electric Current Signals for Information-Energy Integrated Network in Home Environments*, pp. 150–157. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[82] R. Dong, L. Ratliff, H. Ohlsson, and S. S. Sastry, "A dynamical systems approach to energy disaggregation," in *52nd IEEE Conference on Decision and Control*, pp. 6335–6340, Dec 2013.

[83] W. L. Chan, A. T. P. So, and L. L. Lai, "Harmonics load signature recognition by wavelets transforms," in *DRPT2000. International Conference on Electric Utility Deregulation and Restructuring and Power Technologies. Proceedings (Cat. No.00EX382)*, pp. 666–671, 2000.

[84] H.-H. Chang, "Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses," *Energies*, vol. 5, no. 11, pp. 4569–4589, 2012.

[85] N. C. F. Tse, L. Zhou, and L. L. Lai, "Wavelet-based algorithm for nonstationary power system, waveform analysis," in *2008 International Conference on Wavelet Analysis and Pattern Recognition*, vol. 2, pp. 729–735, Aug 2008.

[86] J. Li, S. West, and G. Platt, "Power decomposition based on svm regression," in *2012 Proceedings of International Conference on Modelling, Identification and Control*, pp. 1195–1199, June 2012.

[87] T. Hassan, F. Javed, and N. Arshad, "An empirical investigation of v-i trajectory based load signatures for non-intrusive load monitoring," in *2014 IEEE PES General Meeting | Conference Exposition*, pp. 1–1, July 2014.

[88] N. Iksan, J. Sembiring, N. Haryanto, and S. H. Supangkat, "Appliances identification method of non-intrusive load monitoring based on load signature of v-i trajectory," in *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*, pp. 1–6, Nov 2015.

[89] S. Gupta, M. S. Reynolds, and S. N. Patel, "Electrisense: Single-point sensing using emi for electrical event detection and classification in the home," in *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, UbiComp '10, (New York, NY, USA), pp. 139–148, ACM, 2010.

[90] S. R. Shaw, S. B. Leeb, L. K. Norford, and R. W. Cox, "Nonintrusive load monitoring and diagnostics in power systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, pp. 1445–1454, July 2008.

[91] S.-J. Huang, C.-T. Hsieh, L.-C. Kuo, C.-W. Lin, C.-W. Chang, and S.-A. Fang, "Classification of home appliance electricity consumption using power signature

and harmonic features," in *2011 IEEE Ninth International Conference on Power Electronics and Drive Systems*, pp. 596–599, Dec 2011.

[92] A. Cole and A. Albicki, "Nonintrusive identification of electrical loads in a three-phase environment based on harmonic content," in *Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference [Cat. No. 00CH37066]*, vol. 1, pp. 24–29 vol.1, 2000.

[93] A. I. Cole and A. Albicki, "Data extraction for effective non-intrusive identification of residential power loads," in *IMTC/98 Conference Proceedings. IEEE Instrumentation and Measurement Technology Conference. Where Instrumentation is Going (Cat. No.98CH36222)*, vol. 2, pp. 812–815 vol.2, May 1998.

[94] Z. Wang and G. Zheng, "Residential appliances identification and monitoring by a nonintrusive method," *IEEE Transactions on Smart Grid*, vol. 3, pp. 80–92, March 2012.

[95] M. Berges, L. Soibelman, and H. S. Matthews, "Leveraging data from environmental sensors to enhance electrical load disaggregation algorithms," in *Proceedings of the 13th International Conference on Computing in Civil and Building Engineering*, (Nottingham, UK), jun 2010.

[96] M. Bergés and A. Rowe, "Appliance classification and energy management using multi-modal sensing," in *Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, BuildSys '11, (New York, NY, USA), pp. 51–52, ACM, 2011.

[97] R. Fisera and K. Macek, "Virtual sub-metering via combined classifiers," in *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, vol. 1, pp. 126–131, Sept 2011.

[98] M. A. Guvensan, Z. C. Taysi, and T. Melodia, "Energy monitoring in residential spaces with audio sensor nodes: Tinyears," *Ad Hoc Networks*, vol. 11, no. 5, pp. 1539 – 1555, 2013.

[99] R. Bonfigli, S. Squartini, M. Fagiani, and F. Piazza, "Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview," in *2015 IEEE 15th*

International Conference on Environment and Electrical Engineering (EEEIC), pp. 1175–1180, June 2015.

[100] M. Baranski and V. J, "Nonintrusive appliance load monitoring based on an optical sensor," in *2003 IEEE Bologna Power Tech Conference Proceedings,*, vol. 4, pp. 8 pp. Vol.4–, June 2003.

[101] D. Egarter, V. P. Bhuvana, and W. Elmenreich, "Appliance state estimation based on particle filtering," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, BuildSys'13, (New York, NY, USA), pp. 24:1–24:2, ACM, 2013.

[102] Y. F. Wong, T. Drummond, and Y. A. Sekercioglu, "Real-time load disaggregation algorithm using particle-based distribution truncation with state occupancy model," *Electronics Letters*, vol. 50, pp. 697–699, April 2014.

[103] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Improving nonintrusive load monitoring efficiency via a hybrid programing method," *IEEE Transactions on Industrial Informatics*, vol. 12, pp. 2148–2157, Dec 2016.

[104] M. Sanquer, F. Chatelain, M. El-Guedri, and N. Martin, "Hierarchical bayesian learning for electrical transient classification," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3442–3446, May 2013.

[105] G.-Y. Lin, S.-C. Lee, J. Y.-J. Hsu, and W.-R. Jih, "Applying power meters for appliance recognition on the electric panel," in *2010 5th IEEE Conference on Industrial Electronics and Applications*, pp. 2254–2259, June 2010.

[106] M. Figueiredo, A. de Almeida, and B. Ribeiro, "Home electrical signal disaggregation for non-intrusive load monitoring (nilm) systems," *Neurocomputing*, vol. 96, pp. 66 – 73, 2012. Adaptive and Natural Computing Algorithms.

[107] M. Azaza and F. Wallin, "Supervised household's loads pattern recognition," in *2016 IEEE Electrical Power and Energy Conference (EPEC)*, pp. 1–5, Oct 2016.

[108] S. Kang and J. W. Yoon, "Classification of home appliance by using probabilistic knn with sensor data," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1–5, Dec 2016.

[109] T. Onoda, H. Murata, G. Ratsch, and K. R. Muller, "Experimental analysis of support vector machines with different kernels based on non-intrusive monitoring data," in *Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on*, vol. 3, pp. 2186–2191, 2002.

[110] Y. H. Lin and M. S. Tsai, "Applications of hierarchical support vector machines for identifying load operation in nonintrusive load monitoring systems," in *2011 9th World Congress on Intelligent Control and Automation*, pp. 688–693, June 2011.

[111] L. Jiang, S. Luo, and J. Li, "An approach of household power appliance monitoring based on machine learning," in *2012 Fifth International Conference on Intelligent Computation Technology and Automation*, pp. 577–580, Jan 2012.

[112] D. Srinivasan, W. S. Ng, and A. C. Liew, "Neural-network-based signature recognition for harmonic source identification," *IEEE Transactions on Power Delivery*, vol. 21, pp. 398–405, Jan 2006.

[113] M. Singh, S. Kumar, S. Semwal, and R. S. Prasad, *Residential Load Signature Analysis for Their Segregation Using Wavelet—SVM*, pp. 863–871. New Delhi: Springer India, 2015.

[114] A. Zoha, A. Gluhak, M. Nati, M. A. Imran, and S. Rajasegarar, "Acoustic and device feature fusion for load recognition," in *2012 6th IEEE International Conference Intelligent Systems*, pp. 386–392, Sept 2012.

[115] Y.-X. Lai, C.-F. Lai, Y.-M. Huang, and H.-C. Chao, "Multi-appliance recognition system with hybrid svm/gmm classifier in ubiquitous smart home," *Information Sciences*, vol. 230, no. Supplement C, pp. 39 – 55, 2013. Mobile and Internet Services in Ubiquitous and Pervasive Computing Environments.

[116] C. Duarte, P. Delmar, K. W. Goossen, K. Barner, and E. Gomez-Luna, "Non-intrusive load monitoring based on switching voltage transients and wavelet transforms," in *2012 Future of Instrumentation International Workshop (FIIW) Proceedings*, pp. 1–4, Oct 2012.

[117] H. Altrabalsi, J. Liao, L. Stankovic, and V. Stankovic, "A low-complexity energy disaggregation method: Performance and robustness," in *2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG)*, pp. 1–8, Dec 2014.

[118] H. Altrabalsi, V. Stankovic, and J. L. andLina Stankovic, "Low-complexity energy disaggregation using appliance load modelling," *AIMS Energy*, vol. 4, p. 1, 2016.

[119] S. Su, Y. Yan, H. Lu, L. Kangping, S. Yujing, W. Fei, L. Liming, and R. Hui, "Non-intrusive load monitoring of air conditioning using low-resolution smart meter data," in *2016 IEEE International Conference on Power System Technology (POWERCON)*, pp. 1–5, Sept 2016.

[120] K. M. Rao, D. Ravichandran, and K. Mahesh, "Non-intrusive load monitoring and analytics for device prediction," in *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2016*, pp. 132–136, March 2016.

[121] L. GAO, B. YIN, and Z.-c. ZHU, "Load identification of non-intrusive load-monitoring system based on time-frequency analysis and pso-svm," *DEStech Transactions on Engineering and Technology Research*, 04 2017.

[122] K. T. Chui, M. D. Lytras, and A. Visvizi, "Energy sustainability in smart cities: Artificial intelligence, smart monitoring, and optimization of energy consumption," *Energies*, vol. 11, p. 2869, Oct 2018.

[123] S. Alshareef and W. G. Morsi, "Application of wavelet-based ensemble tree classifier for non-intrusive load monitoring," in *2015 IEEE Electrical Power and Energy Conference (EPEC)*, pp. 397–401, Oct 2015.

[124] M. Nguyen, S. Alshareef, A. Gilani, and W. G. Morsi, "A novel feature extraction and classification algorithm based on power components using single-point monitoring for nilm," in *2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 37–40, May 2015.

[125] J. M. Gillis, S. M. Alshareef, and W. G. Morsi, "Nonintrusive load monitoring using wavelet design and machine learning," *IEEE Transactions on Smart Grid*, vol. 7, pp. 320–328, Jan 2016.

[126] H. H. Chang, L. S. Lin, N. Chen, and W. J. Lee, "Particle swarm optimization based non-intrusive demand monitoring and load identification in smart meters," in *2012 IEEE Industry Applications Society Annual Meeting*, pp. 1–8, Oct 2012.

[127] H. H. Chang, L. S. Lin, N. Chen, and W. J. Lee, "Particle-swarm-optimization-based nonintrusive demand monitoring and load identification in smart meters," *IEEE Transactions on Industry Applications*, vol. 49, pp. 2229–2236, Sept 2013.

[128] H. H. Chang, K.-L. Lian, Y.-C. Su, and W. J. Lee, "Energy spectrum-based wavelet transform for non-intrusive demand monitoring and load identification," in *2013 IEEE Industry Applications Society Annual Meeting*, pp. 1–9, Oct 2013.

[129] S. Biansoongnern and B. Plangklang, "Nonintrusive load monitoring (nilm) using an artificial neural network in embedded system with low sampling rate," in *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 1–4, June 2016.

[130] R. Bonfigli, A. Felicetti, E. Principi, M. Fagiani, S. Squartini, and F. Piazza, "Denoising autoencoders for non-intrusive load monitoring: Improvements and comparative evaluation," *Energy and Buildings*, vol. 158, pp. 1461 – 1474, 2018.

[131] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, (USA), pp. 807–814, Omnipress, 2010.

[132] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajic, "Ampds: A public dataset for load disaggregation and eco-feedback research," in *2013 IEEE Electrical Power Energy Conference*, pp. 1–6, Aug 2013.

[133] F. C. C. Garcia, C. M. C. Creayla, and E. Q. B. Macabebe, "Development of an intelligent system for smart home energy disaggregation using stacked denoising autoencoders," *Procedia Computer Science*, vol. 105, no. Supplement C, pp. 248 – 255, 2017. 2016 IEEE International Symposium on Robotics and Intelligent Sensors, IRIS 2016, 17-20 December 2016, Tokyo, Japan.

[134] T. T. H. Le, J. Kim, and H. Kim, "Classification performance using gated recurrent unit recurrent neural network on energy disaggregation," in *2016 International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 1, pp. 105–110, July 2016.

[135] A. Zoha, A. Gluhak, M. Nati, and M. A. Imran, "Low-power appliance monitoring using factorial hidden markov models," in *2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 527–532, April 2013.

[136] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent lstm network," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 63–67, Dec 2015.

[137] L. Mauch and B. Yang, "A novel dnn-hmm-based approach for extracting single loads from aggregate power signals," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2384–2388, March 2016.

[138] P. P. M. do Nascimento, *Applications of Deep Learning Techniques on NILM*. Rio de Janeiro, Brasil: Master Thesis. Rio de Janeiro: UFRJ/COPPE, 2016., 4 2016.

[139] W. He and Y. Chai, "An empirical study on energy disaggregation via deep learning," in *2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016)*, Atlantis Press, 2016/11.

[140] J. Kim, T.-T.-H. Le, and H. Kim, "Nonintrusive load monitoring based on advanced deep learning and novel signature," *Computational intelligence and neuroscience*, vol. 2017, pp. 4216281; 4216281–4216281, 2017.

[141] C. Zhang, M. Zhong, Z. Wang, N. H. Goddard, and C. A. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 2604–2611, 2018.

[142] D. Murray, L. Stankovic, V. Stankovic, S. Lulic, and S. Sladojevic, "Transferability of neural networks approaches for low-rate energy disaggregation," in *2019 International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2019, Conference date: 12-05-2019 Through 17-05-2019*, 2 2019.

[143] K. S. Barsim and B. Yang, "On the feasibility of generic deep disaggregation for single-load extraction," *CoRR*, vol. abs/1802.02139, 2018.

[144] V. Stankovic, J. Liao, and L. Stankovic, "A graph-based signal processing approach for low-rate energy disaggregation," in *2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES)*, pp. 81–87, Dec 2014.

[145] M. Aiad and P. H. Lee, "Unsupervised approach for load disaggregation with devices interactions," *Energy and Buildings*, vol. 116, no. Supplement C, pp. 96 – 103, 2016.

[146] G. Elafoudi, L. Stankovic, and V. Stankovic, "Power disaggregation of domestic smart meter readings using dynamic time warping," in *2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pp. 36–39, May 2014.

[147] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Power disaggregation for low-sampling rate data," in 2^{nd} *International Workshop on Non-Intrusive Load Monitoring (NILM)*, June 2014.

[148] T. Zia, D. Bruckner, and A. Zaidi, "A hidden markov model based procedure for identifying household electric loads," in *IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society*, pp. 3218–3223, Nov 2011.

[149] Z. Guo, Z. J. Wang, and A. Kashani, "Home appliance load modeling from aggregated smart meter data," *IEEE Transactions on Power Systems*, vol. 30, pp. 254–262, Jan 2015.

[150] S. Makonin, F. Popowich, I. V. Baji?, B. Gill, and L. Bartram, "Exploiting hmm sparsity to perform online real-time nonintrusive load monitoring," *IEEE Transactions on Smart Grid*, vol. 7, pp. 2575–2585, Nov 2016.

[151] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. Rizzoli, "A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring," *Applied Energy*, vol. 185, no. Part 1, pp. 331 – 344, 2017.

[152] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, "Non-intrusive load monitoring by using active and reactive power in additive factorial hidden markov models," *Applied Energy*, vol. 208, no. Supplement C, pp. 1590 – 1607, 2017.

[153] M. Wytock and J. Zico Kolter, "Contextually Supervised Source Separation with Application to Energy Disaggregation," *ArXiv e-prints*, Dec. 2013.

[154] H. Gonçalves, A. Ocneanu, and M. Bergés, "Unsupervised disaggregation of appliances using aggregated consumption data," in 1^{st} *KDD 2011 Workshop Data Mining Applications in Sustainability*, Aug 2011.

[155] K. He, D. Jakovetic, V. Stankovic, and L. Stankovic, "Post-processing for event-based non-intrusive load monitoring," in *4th International Workshop on Non-Intrusive Load Monitoring ; Conference date: 07-03-2018 Through 08-03-2018*, pp. 1–4, 3 2018.

[156] S. Makonin, *Real-Time Embedded Low-Frequency Load Disaggregation.* PhD thesis, Simon Fraser University, School of Computing Science, August 2014.

[157] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining.* Norwell, MA, USA: Kluwer Academic Publishers, 1998.

[158] D. L. Olson and D. Delen, *Advanced Data Mining Techniques.* Springer Publishing Company, Incorporated, 1st ed., 2008.

[159] A. Faustine, N. H. Mvungi, S. Kaijage, and K. Michael, "A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem," *CoRR*, vol. abs/1703.00785, 2017.

[160] J. Liao, L. Stankovic, and V. Stankovic, "Detecting household activity patterns from smart meter data," in *2014 International Conference on Intelligent Environments*, pp. 71–78, June 2014.

[161] N. Batra, H. Dutta, and A. Singh, "Indic: Improved non-intrusive load monitoring using load division and calibration," in *2013 12th International Conference on Machine Learning and Applications*, vol. 1, pp. 79–84, Dec 2013.

[162] N. Batra, A. Singh, and K. Whitehouse, "If you measure it, can you improve it? exploring the value of energy disaggregation," in *Proceedings of the 2Nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, BuildSys '15, (New York, NY, USA), pp. 191–200, ACM, 2015.

[163] H. Dong, B. Wang, and C.-T. Lu, "Deep sparse coding based recursive disaggregation model for water conservation," in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13, pp. 2804–2810, AAAI Press, 2013.

[164] D. Piga, A. Cominola, M. Giuliani, A. Castelletti, and A. E. Rizzoli, "Sparse optimization for automated energy end use disaggregation," *IEEE Transactions on Control Systems Technology*, vol. 24, pp. 1044–1051, May 2016.

[165] M. Müller, *Information Retrieval for Music and Motion*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.

[166] K. A. Nguyen, H. Zhang, and R. A. Stewart, "Development of an intelligent model to categorise residential water end use events," *Journal of Hydro-environment Research*, vol. 7, no. 3, pp. 182 – 201, 2013.

[167] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, pp. 43–49, Feb 1978.

[168] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[169] K. Nguyen, R. Stewart, and H. Zhang, "An intelligent pattern recognition model to automate the categorisation of residential water end-use events," *Environmental Modelling & Software*, vol. 47, pp. 108 – 127, 2013.

[170] F. Gullo, G. Ponti, A. Tagarelli, S. liritano, M. Ruffolo, and D. Labate, "Low-voltage electricity customer profiling based on load data clustering," in *Proceedings of the 2009 International Database Engineering & Applications Symposium*, IDEAS '09, (New York, NY, USA), pp. 330–333, ACM, 2009.

[171] F. Kupzog, T. Zia, and A. A. Zaidi, "Automatic electric load identification in self-configuring microgrids," in *AFRICON 2009*, pp. 1–5, Sept 2009.

[172] A. A. Zaidi, F. Kupzog, T. Zia, and P. Palensky, "Load recognition for automated demand response in microgrids," in *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, pp. 2442–2447, Nov 2010.

[173] K. Basu, V. Debusschere, A. Douzal-Chouakria, and S. Bacha, "Time series distance-based methods for non-intrusive load monitoring in residential buildings," *Energy and Buildings*, vol. 96, pp. 109 – 117, 2015.

[174] T. Teeraratkul, D. O?Neill, and S. Lall, "Shape-based approach to household electric load curve clustering and prediction," *IEEE Transactions on Smart Grid*, vol. 9, pp. 5196–5206, Sep. 2018.

[175] B. Liu, W. Luan, and Y. Yu, "Dynamic time warping based non-intrusive load transient identification," *Applied Energy*, vol. 195, pp. 634 – 645, 2017.

[176] W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, "An extensible approach for non-intrusive load disaggregation with smart meter data," *IEEE Transactions on Smart Grid*, vol. 9, pp. 3362–3372, July 2018.

[177] L. Stankovic, G. Elafoudi, C. Tachtatzis, I. Andonovic, G. Cassels, and A. Fickling, "Decision support system for proactive maintenance of earthworks assets," in *World Water Congress XV*, May 2015.

[178] G. Elafoudi, V. Stankovic, L. Stankovic, D. Pappusetti, and H. Kalva, "Evaluation of signal processing methods for attention assessment in visual content interaction," in *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops: ICIAP 2015 International Workshops, BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, September 7-8, 2015, Proceedings*, (Cham), pp. 580–588, Springer International Publishing, 2015.

[179] R. Sellers, N. Dixon, T. Dijkstra, D. Gunn, J. Chambers, P. Jackson, and P. Hughes, "Electrical resistivity as a tool to identify areas of progressive failure within uk infrastructure embankments," in *IAEG Congress 2010, Auckland, New Zealand*, EAGE, Sept. 2010.

[180] D. Gunn and R. Ogilvy, "Alert-me : new technologies for embankment warning systems," *Rail Technology Magazine*, 2010.

[181] A. Perrone, V. Lapenna, and S. Piscitelli, "Electrical resistivity tomography technique for landslide investigation: A review," *Earth-Science Reviews*, vol. 135, pp. 65 – 82, 2014.

[182] F. Wenner, "A method for measuring earth resistivity," *Journal of the Washington Academy of Sciences*, vol. 5, no. 16, pp. 561–563, 1915.

[183] IEEE, "IEEE recommended practice for grounding of industrial and commercial power systems - redline," *IEEE Std 142-2007 (Revision of IEEE Std 142-1991) - Redline*, pp. 1–215, Nov 2007.

[184] IEEE, "IEEE guide for measuring earth resistivity, ground impedance, and earth surface potentials of a grounding system - redline," *IEEE Std 81-2012 (Revision of IEEE Std 81-1983) - Redline*, pp. 1–192, Dec 2012.

[185] V. Pangonilo, "Earth rod electrode resistance." https://pangonilo.com/index.php?sdmon=files/Earth_Rod_Resistance.pdf, 2018. Accessed 20 June 2018.

[186] C. N. Nwankwo, D. Ogagarue, and F. Ezeoke, "Investigation of variation in resistivity with depth and soil dielectric constant in parts of rivers state, southern nigeria," *British Journal of Applied Science & Technology*, vol. 3, no. 3, pp. 452–461, 2013.

[187] J.-Y. Wang, "Pupil dilation and eye tracking," in *A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide*, pp. 185–204, 01 2011.

[188] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *International Journal of Human-Computer Studies*, vol. 59, no. 1, pp. 185 – 198, 2003.

[189] C. Privitera, L. W. Renninger, T. Carney, S. Klein, and M. Aguilar, "The pupil dilation response to visual detection," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 6806, 03 2008.

[190] F. T. Oliveira, A. Aula, and D. M. Russell, "Discriminating the relevance of web search results with measures of pupil size," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, (New York, NY, USA), pp. 2209–2212, ACM, 2009.

[191] C. M. Privitera, L. W. Renninger, T. Carney, S. R. Klein, and M. Aguilar, "Pupil dilation during visual target detection.," *Journal of vision*, vol. 10 10, p. 3, 2010.

[192] J. Beatty, "Task-evoked pupillary responses, processing load, and the structure of processing resources," in *Psychological Bulletin*, vol. 91, pp. 276–292, 03 1982.

[193] G. Hossain and M. Yeasin, "Understanding effects of cognitive load from pupillary responses using hilbert analytic phase," in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 381–386, June 2014.

[194] J. Klingner, R. Kumar, and P. Hanrahan, "Measuring the task-evoked pupillary response with a remote eye tracker," in *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, ETRA '08, (New York, NY, USA), pp. 69–72, ACM, 2008.

[195] S. P. Marshall, "Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity."

[196] S. P Marshall, "Identifying cognitive state from eye metrics," in *Aviation, space, and environmental medicine*, vol. 78, pp. B165–75, 06 2007.

[197] Tobii Technology AB, *User's Manual Tobii X2-60 Eye Tracker*. Tobii Technology AB, Karlsrovägen 2D, Box 743 S-182 17, Danderyd, Stockholm, Sweden, 1.0.3 ed., June 2014.

[198] A. Voßkühler, V. Nordmeier, L. Kuchinke, and A. M. Jacobs, "Ogama (open gaze and mouse analyzer): Open-source software designed to analyze eye and mouse movements in slideshow study designs," *Behavior Research Methods*, vol. 40, pp. 1150–1162, Nov 2008.

[199] B. Wilson, "House." https://www.flickr.com/photos/billy_wilson/4073334407/in/photostream/, 27 Jun 2009. Creativity Commons Attribution-NonCommercial 2.0 Generic (CC BY-NC 2.0), https://creativecommons.org/licenses/by-nc/2.0/, Accessed 6 June 2018.

[200] K. Gaensler, "The Parton." https://www.flickr.com/photos/gaensler/7921310066/, 24 May 2012. Creativity Commons Attribution-NonCommercial-ShareAlike 2.0 Generic (CC BY-NC-SA 2.0), https://creativecommons.org/licenses/by-nc-sa/2.0/, Accessed 6 June 2018.

[201] L. Haines, "Bartestree Cottage." https://www.flickr.com/photos/leshaines123/8657745074/, 5 Feb 2009. Creativity Commons Attribution 2.0 Generic (CC BY 2.0), https://creativecommons.org/licenses/by/2.0/, Accessed 6 June 2018.

[202] Marketing Deluxe, "Small luxury hotels of the world grossarler hof groarltal salzburger land sterreich doppelzimmer standard fichte." https://www.flickr.com/photos/marketing-deluxe/8093497858/, 25 Sept. 2012. Creativity Commons Attribution-NonCommercial-ShareAlike 2.0 Generic (CC BY-NC-SA 2.0), https://creativecommons.org/licenses/by-nc-sa/2.0/, Accessed 6 June 2018.

[203] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, pp. 70–73, Jun 1967.

[204] F. J. HARRIS, "Chapter 3 - multirate fir filters for interpolating and desampling," in *Handbook of Digital Signal Processing* (D. F. Elliott, ed.), pp. 173 – 287, San Diego: Academic Press, 1987.

[205] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, pp. 613–627, May 1995.

[206] Computer Business Review, "What is consumer internet of things ?." https://www.cbronline.com/what-is/what-is-consumer-internet-of-things-4926794/, 2019. Accessed 8 March 2019.

[207] X. Zhu, S. K. Mukhopadhyay, and H. Kurata, "A review of rfid technology and its managerial applications in different industries," *Journal of Engineering and Technology Management*, vol. 29, no. 1, pp. 152 – 167, 2012. Creating competitive edge in operations and service management through technology and innovation.

[208] L. Ruiz-Garcia and L. Lunadei, "The role of rfid in agriculture: Applications, limitations and challenges," *Computers and Electronics in Agriculture*, vol. 79, no. 1, pp. 42 – 50, 2011.

[209] S. Zhang and H. Zhang, "A review of wireless sensor networks and its applications," in *2012 IEEE International Conference on Automation and Logistics*, pp. 386–389, Aug 2012.

[210] J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Terzis, and M. Welsh, "Wireless sensor networks for healthcare," *Proceedings of the IEEE*, vol. 98, pp. 1947–1960, Nov 2010.

[211] H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2688 – 2710, 2010.

[212] D. Basu, G. Moretti, G. S. Gupta, and S. Marsland, "Wireless sensor network based smart home: Sensor selection, deployment and monitoring," in *2013 IEEE Sensors Applications Symposium Proceedings*, pp. 49–54, Feb 2013.

[213] M. E. Brak and M. Essaaidi, "Wireless sensor network in smart grid technology: Challenges and opportunities," in *2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, pp. 578–583, March 2012.

[214] Y. Zhang, X. Li, S. Zhang, and Y. Zhen, "Wireless sensor network in smart grid: Applications and issue," in *2012 World Congress on Information and Communication Technologies*, pp. 1204–1208, Oct 2012.

[215] C. G. Keller, T. Dang, H. Fritz, A. Joos, C. Rabe, and D. M. Gavrila, "Active pedestrian safety by automatic braking and evasive steering," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 1292–1304, Dec 2011.

[216] Waymo, "Our journey." https://waymo.com/journey/, 2019. Accessed 8 March 2019.

[217] Waymo, "Waymo safety report: On the road to fully self-driving." https://storage.googleapis.com/sdc-prod/v1/safety-report/waymo-safety-report-2017-10.pdf, 2017. Accessed 8 March 2019.

[218] Uber Advanced Technologies Group, "A principled approach to safety." https://uber.app.box.com/v/UberATGSafetyReport, 2018. Accessed 8 March 2019.

[219] Tesla, "Autopilot." https://www.tesla.com/en_GB/autopilot?redirect=no, 2018. Accessed 8 March 2019.

[220] Cruise Automation, "Mission." https://getcruise.com/mission, 2018. Accessed 8 March 2019.

[221] General Motors, "Self-driving safety report." https://www.gm.com/content/dam/company/docs/us/en/gmcom/gmsafetyreport.pdf, 2018. Accessed 8 March 2019.

[222] Navya, "Navya safety report: The autonom era." https://navya.tech/wp-content/uploads/2019/01/NAVYA-Safety-Report-01.09.2019-1.pdf, 2018. Accessed 8 March 2019.

[223] M. V. Rajasekhar and A. K. Jaswal, "Autonomous vehicles: The future of automobiles," in *2015 IEEE International Transportation Electrification Conference (ITEC)*, pp. 1–6, Aug 2015.

[224] D. Bajpayee and J. Mathur, "A comparative study about autonomous vehicle," in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1–6, March 2015.

[225] Apur: Atelier Parisien D'Urbanism , "Impacts and potential benefits of autonomous vehicles: from an international context to grand paris." https://www.apur.org/sites/default/files/documents/publication/etudes/impacts_potential_benefits_autonomous_vehicles.pdf, Oct. 2018. Accessed 8 March 2019.

[226] D. Yadron and D. Tynan, "Tesla driver dies in first fatal crash while using autopilot mode." https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk, July 2016. Accessed 8 March 2019.

[227] S. Levin and J. C. Wong, "Self-driving uber kills arizona woman in first fatal crash involving pedestrian." https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe, Mar. 2018. Accessed 8 March 2019.

[228] National Highway Traffic Safety Administration , "Preparing for the future of transportation: Automated vehicle 3.0." https://www.transportation.gov/sites/dot.gov/files/docs/policy-initiatives/automated-vehicles/320711/preparing-future-transportation-automated-vehicle-30.pdf, Oct. 2018. Accessed 8 March 2019.

[229] Law Commission and Scottish Law Commission, "Automated vehicles: A joint preliminary consultation paper." https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2018/11/6.5066_LC_AV-Consultation-Paper-5-November_061118_WEB-1.pdf, Nov 2018. Accessed 8 March 2019.

[230] Mobivia, "The connected car: Towards an eu regulation ensuring effective and fair access to in-vehicle data." http://ec.europa.eu/competition/information/digitisation_2018/contributions/mobivia.pdf, Sep. 2018. Accessed 8 March 2019.

[231] Y. Zhang, B. Chen, and X. Lu, "Intelligent monitoring system on refrigerator trucks based on the internet of things," in *Wireless Communications and Applications* (P. Sénac, M. Ott, and A. Seneviratne, eds.), (Berlin, Heidelberg), pp. 201–206, Springer Berlin Heidelberg, 2012.

[232] Y. Álvarez López, J. Franssen, G. Álvarez Narciandi, J. Pagnozzi, I. González-Pinto Arrillaga, and F. Las-Heras Andrés, "RFID technology for management and tracking: e-health applications," *Sensors (Basel, Switzerland)*, vol. 18, p. 2663, Aug. 2018.

[233] Z. Pang, Q. Chen, J. Tian, L. Zheng, and E. Dubrova, "Ecosystem analysis in the design of open platform-based in-home healthcare terminals towards the internet-of-things," in *2013 15th International Conference on Advanced Communications Technology (ICACT)*, pp. 529–534, Jan 2013.

[234] D. Niyato, E. Hossain, and S. Camorlinga, "Remote patient monitoring service using heterogeneous wireless access networks: architecture and optimization," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 412–423, May 2009.

[235] R. R. Harmon, E. G. Castro-Leon, and S. Bhide, "Smart cities and the internet of things," in *2015 Portland International Conference on Management of Engineering and Technology (PICMET)*, pp. 485–494, Aug 2015.

[236] K. K. Patel and S. M. Patel, "Internet of things-iot: Definition, characteristics, architecture, enabling technologies, application & future challenges," *International Journal of Engineering Science and Computing (IJESC*, vol. 6, pp. 6122–6131, May 2016.

[237] I. Ud Din, M. Guizani, S. Hassan, B. Kim, M. Khurram Khan, M. Atiquzzaman, and S. H. Ahmed, "The internet of things: A review of enabled technologies and future challenges," *IEEE Access*, vol. 7, pp. 7606–7640, 2019.

[238] M. A. Mahmud, A. Abdelgawad, K. Yelamarthi, and Y. A. Ismail, "Signal processing techniques for iot-based structural health monitoring," in *2017 29th International Conference on Microelectronics (ICM)*, pp. 1–5, Dec 2017.

[239] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiqa, and I. Yaqoob, "Big iot data analytics: Architecture, opportunities, and open research challenges," *IEEE Access*, vol. 5, pp. 5247–5261, 2017.

[240] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, pp. 22–32, Feb 2014.

[241] H. N. Saha, S. Auddy, A. Chatterjee, S. Pal, S. Sarkar, R. Singh, A. K. Singh, P. Sharan, S. Banerjee, R. Sarkar, and A. Maity, "Iot solutions for smart cities," in *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, pp. 74–80, Aug 2017.

[242] F. Li, W. Qiao, H. Sun, H. Wan, J. Wang, Y. Xia, Z. Xu, and P. Zhang, "Smart transmission grid: Vision and framework," *IEEE Transactions on Smart Grid*, vol. 1, pp. 168–177, Sep. 2010.

[243] Y. Saleem, N. Crespi, M. H. Rehmani, and R. Copeland, "Internet of things-aided smart grid: Technologies, architectures, applications, prototypes, and future research directions," *CoRR*, vol. abs/1704.08977, 2017.

[244] Q. Ou, Y. Zhen, X. Li, Y. Zhang, and L. Zeng, "Application of internet of things in smart grid power transmission," in *2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing*, pp. 96–100, June 2012.

[245] X. Chen, L. Sun, H. Zhu, Y. Zhen, and H. Chen, "Application of internet of things in power-line monitoring," in *2012 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, pp. 423–426, Oct 2012.

[246] K. Sohraby, D. Minoli, B. Occhiogrosso, and W. Wang, "A review of wireless and satellite-based m2m/iot services in support of smart grids," *Mobile Networks and Applications*, vol. 23, pp. 881–895, Aug 2018.

[247] D. C. Mocanu, E. Mocanu, P. H. Nguyen, M. Gibescu, and A. Liotta, "Big iot data mining for real-time energy disaggregation in buildings," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 003765–003769, Oct 2016.

[248] M. Nardello, M. Rossi, and D. Brunelli, "A low-cost smart sensor for non intrusive load monitoring applications," in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, pp. 1362–1368, June 2017.

[249] M. A. Mengistu, A. A. Girmay, C. Camarda, A. Acquaviva, and E. Patti, "A cloud-based on-line disaggregation algorithm for home appliance loads," *IEEE Transactions on Smart Grid*, pp. 1–1, 2018.

[250] K. Sornalatha and V. R. Kavitha, "Iot based smart museum using bluetooth low energy," in *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, pp. 520–523, Feb 2017.

[251] R. Goswami, "How iot will redefine the theme park experience." https://bit.ly/2TNzpdo, Feb. 2017. Accessed 8 March 2019.

[252] B. Marr, "Disney uses big data, iot and machine learning to boost customer experience." https://bit.ly/2O1BMEj, Aug. 2017. Accessed 8 March 2019.

[253] K. Smith, S. V. Ravulapati, Y. Daga, and D. S. Naga, "The internet of things: the personal and social domain," *European Journal of Management Studies*, vol. 20, no. 1, pp. 43–50, 2015.

[254] Tile, "How it works." https://www.thetileapp.com/en-gb/how-it-works. Accessed 8 March 2019.

[255] Apple, "Find my iphone." https://support.apple.com/explore/find-my-iphone-ipad-mac-watch. Accessed 8 March 2019.

[256] Apple, "Find, lock, or erase a lost android device." https://support.google.com/accounts/answer/6160491?hl=en. Accessed 8 March 2019.

[257] Z. Ren, X. Liu, R. Ye, and T. Zhang, "Security and privacy on internet of things," in *2017 7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 140–144, July 2017.

[258] T. Xu, J. B. Wendt, and M. Potkonjak, "Security of iot systems: Design challenges and opportunities," in *2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 417–423, Nov 2014.

[259] D. Boneh and R. J. Lipton, "A revocable backup system," in *Proceedings of the 6th Conference on USENIX Security Symposium, Focusing on Applications of Cryptography - Volume 6*, SSYM'96, (Berkeley, CA, USA), pp. 9–9, USENIX Association, 1996.

[260] Z. N. J. Peterson, R. Burns, J. Herring, A. Stubblefield, and A. Rubin, "Secure deletion for a versioning file system," in *Proceedings of the USENIX Conference on File And Storage Technologies (FAST)*, pp. 143–154, December 2005.

[261] S. Bin, L. Yuan, and W. Xiaoyi, "Research on data mining models for the internet of things," in *2010 International Conference on Image Analysis and Signal Processing*, pp. 127–132, April 2010.

[262] A. Singh and S. Sharma, "Analysis on data mining models for internet of things," in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 94–100, Feb 2017.

[263] A. Savaliya, A. Bhatia, and J. Bhatia, "Application of data mining techniques in iot: A short review," *International Journal of Scientific Research in Science, Engineering and Technology(IJSRSET)*, vol. 4, pp. 218–228, Jan.-Feb. 2018.

[264] P. Wlodarczak, M. Ally, and J. Soar, "Data mining in iot: Data analysis for a new paradigm on the internet," in *Proceedings of the International Conference on Web Intelligence*, WI '17, (New York, NY, USA), pp. 1100–1103, ACM, 2017.

[265] C. Tsai, C. Lai, M. Chiang, and L. T. Yang, "Data mining for internet of things: A survey," *IEEE Communications Surveys Tutorials*, vol. 16, pp. 77–97, Jan. 2014.

[266] A. Shobanadevi and G. Maragatham, "Data mining techniques for iot and big data ? a survey," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 607–610, Dec 2017.

[267] O. Matei, C. Anton, S. Scholze, and C. Cenedese, "Multi-layered data mining architecture in the context of internet of things," in *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*, pp. 1193–1198, July 2017.

[268] P. Gupta and R. Gupta, "Data mining framework for iot applications," *International Journal of Computer Applications*, vol. 174, pp. 4–7, Sep 2017.

[269] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth, "Machine learning for internet of things data analysis: a survey," *Digital Communications and Networks*, vol. 4, no. 3, pp. 161 – 175, 2018.

[270] P. Ray, "A survey on internet of things architectures," *Journal of King Saud University - Computer and Information Sciences*, vol. 30, no. 3, pp. 291 – 319, 2018.

[271] P. Brezany, I. Janciak, and A. Min Tjoa, "Gridminer: a fundamental infrastructure for building intelligent grid systems," in *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, pp. 150–156, Sep. 2005.

[272] A. Congiusta, D. Talia, and P. Trunfio, "Distributed data mining services leveraging wsrf," *Future Gener. Comput. Syst.*, vol. 23, pp. 34–41, Jan. 2007.

[273] V. Stankovski, M. Swain, V. Kravtsov, T. Niessen, D. Wegener, M. Rhm, J. Trnkoczy, M. May, J. Franke, A. Schuster, and W. Dubitzky, "Digging deep into the data mine with dataminin ggrid," *IEEE Internet Computing*, vol. 12, pp. 69–76, Nov 2008.

[274] F. Chen, P. Deng, J. Wan, D. Zhang, A. V. Vasilakos, and X. Rong, "Data mining for the internet of things: Literature review and challenges," *Int. J. Distrib. Sen. Netw.*, vol. 2015, pp. 12:12–12:12, Jan. 2015.

[275] S. Guha, R. Rastogi, and K. Shim, "Cure: An efficient clustering algorithm for large databases," in *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, SIGMOD '98, (New York, NY, USA), pp. 73–84, ACM, 1998.

[276] S. Guha, R. Rastogi, and K. Shim, "Cure: An efficient clustering algorithm for large databases," *Inf. Syst.*, vol. 26, pp. 35–58, Mar. 2001.

[277] M. W. Berry and M. Browne, *Understanding Search Engines: Mathematical Modeling and Text Retrieval (Software, Environments, Tools), Second Edition.* Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2005.

[278] H. Bangui, M. Ge, and B. Buhnova, "Exploring big data clustering algorithms for internet of things applications," in *Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security, IoTBDS 2018, Funchal, Madeira, Portugal, March 19-21, 2018.*, pp. 269–276, 2018.

[279] D. Pandove and S. Goel, "A comprehensive study on clustering approaches for big data mining," in *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, pp. 1333–1338, Feb 2015.

[280] H. Andrews and C. Patterson, "Singular value decompositions and digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, pp. 26–53, February 1976.

[281] S. Tanwar, T. Ramani, and S. Tyagi, "Dimensionality reduction using pca and svd in big data: A comparative case study," in *Future Internet Technologies and Trends* (Z. Patel and S. Gupta, eds.), (Cham), pp. 116–125, Springer International Publishing, 2018.

[282] D. García, I. Díaz, D. Pérez, A. A. Cuadrado, M. Domínguez, and A. Morán, "Interactive visualization for nilm in large buildings using non-negative matrix factorization," *Energy and Buildings*, vol. 176, pp. 95 – 108, 2018.

[283] K. J. Cios, W. Pedrycz, R. W. Swiniarski, and L. A. Kurgan, *Data Mining: A Knowledge Discovery Approach*. Berlin, Heidelberg: Springer-Verlag, 2007.

[284] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*. IEEE, 2003.

[285] Z. Zheng, H. Chen, and X. Luo, "A supervised event-based non-intrusive load monitoring for non-linear appliances," *Sustainability*, vol. 10, p. 1001, Mar 2018.

[286] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[287] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 3rd ed., 2011.

[288] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, Mar 1986.

[289] J. R. QUINLAN, ed., *C4.5: Programs for Machine Learning*. San Francisco (CA): Morgan Kaufmann, 1993.

[290] G. V. Kass, "An exploratory technique for investigating large quantities of categorical data," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 29, no. 2, pp. 119–127, 1980.

[291] M. van Diepen and P. H. Franses, "Evaluating chi-squared automatic interaction detection," *Information Systems*, vol. 31, no. 8, pp. 814 – 831, 2006.

[292] J. Mei, D. He, R. G. Harley, and T. G. Habetler, "Random forest based adaptive non-intrusive load identification," in *2014 International Joint Conference on Neural Networks (IJCNN)*, pp. 1978–1983, July 2014.

[293] Z. Yan and C. Xu, "Combining knn algorithm and other classifiers," in *9th IEEE International Conference on Cognitive Informatics (ICCI'10)*, pp. 800–805, July 2010.

Appendix A

Appliance Types

According to Hart in [24] and Baranski and Voss in [100], consumer appliances can be classified in four classes depending on their operational states. These four classes are as follows:

- **On/Off appliances**

  This class includes most of the common household appliances, such as a toaster, kettle or a light bulb. These appliances usually can be switched on using simple two-point control. During their ON state, they produce a specific amount of power, which can be detected in the mains with the increased consumed power, but the duration of the ON state can vary depending on user's specifications. As an example, a toaster's ON state can be affected by the toasting settings and function selected (e.g. defrosting), while a kettle's ON state by the volume of the water that requires boiling.

- **Finite State Machines (FSM)**

  These are multi-state appliances with a finite number of operating states and according to Hart in [24] can be represented using *Finite State Machines (FSM)* model, thus their name. The switching pattern of these appliances is repeatable, and each state has a specific power consumption. This class includes appliances such as washing machine or tumble dryer. For example, a washing machine's operation has three states/cycles, washing, rinsing and spinning, which are repeating every time a household user operates the appliance (daily or weekly). The duration of each cycle and the total duration of washing machine's operation can vary depending on the programme and temperature selected.

- **Continuously Variable Consumer Devices**

  These are appliances with variable power consumption and no periodic pattern of changing states or power. For this reason, their detection and therefore disaggregation using a NILM algorithm is almost impossible. Examples of devices in this class include dimmable lights and power tools.

- **Permanent Consumer Devices**

  According to Baranski and Voss in [100], these are devices that remain on for 24 hours a day, 7 days a week, and they have approximately constant active and reactive power consumption. Examples of devices in this class include hard-wired smoke alarms and gas alarms.

Appendix B

Supplementary Results for Chapter 5

This appendix provides the results presented in Chapter 5 with benchmarks the state-of-the-art NILM methods used in Chapter 4.

B.1 Disaggregation Results for House 1 from REDD Dataset

Table B.1 shows the classification accuracy obtained using F_M for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REDD House 1, as presented in [5]. Table B.2 shows the estimation accuracy per appliance using Acc_i for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REDD House 1, as presented in [5].

Table B.1: Classification Accuracy using F_M of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REDD House 1, as presented in [5].

	DTW			*UGSP*[4]			*SGSP*[33]			*DT*[32]			
Apps	(1)	(2)	(3)	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.32	–	0.32	0.12	0.10	0.52	0.10	0.16	**0.63**	0.05	0.05	0.57	0.14
F	**0.80**	**0.80**	**0.80**	0.52	0.51	0.63	0.08	0.13	0.58	0.42	0.43	0.63	0.51
KO	**0.86**	0.83	**0.86**	0.10	0.17	0.47	0.04	0.14	0.55	0.02	0.03	0.47	0.68
MW	**0.78**	**0.78**	**0.78**	0.39	0.21	0.40	0.32	0.27	0.57	0.21	0.24	0.43	0.19
WD	**0.79**	**0.79**	**0.79**	0.44	0.49	0.61	0.53	0.52	0.71	0.46	0.42	0.65	0

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, WD=Washer Dryer. (1) represents the original DTW-based method in Chapter 4, (2) the DTW+kM, and (3) the kDTW, that uses DTW as a final refinement step. **Bold** values represent the method/methods with the highest obtained classification accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Table B.2: Estimation Accuracy using Acc_i of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REDD House 1, as presented in [5].

	DTW			*UGSP*[4]			*SGSP*[33]			*DT*[32]			
Apps	(1)	(2)	(3)	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.64	–	0.64	0.23	0.21	0.66	0.16	0.24	**0.72**	0.55	0.51	0.58	0.21
F	0.80	0.80	0.80	0.68	0.65	0.91	0.48	0.51	**0.93**	0.47	0.47	0.88	0.59
KO	0.74	0.74	0.74	0.16	0.25	0.83	0.09	0.11	0.84	0.24	0.27	0.80	**0.92**
MW	0.63	0.63	0.63	0.40	0.26	0.66	0.37	0.46	**0.69**	0.58	0.61	0.62	0.40
WD	0.52	0.52	0.52	0.58	0.61	0.95	0.64	0.61	**0.96**	0.76	0.73	0.89	−1.99

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, WD=Washer Dryer. (1) represents the original DTW-based method in Chapter 4, (2) the DTW+kM, and (3) the kDTW, that uses DTW as a final refinement step. **Bold** values represent the method/methods with the highest obtained estimation accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Figure B.1 shows the comparative results, as found in Table B.1, in terms of classification accuracy, using F_M, between the proposed methods of this chapter, the DTW-based method from Chapter 4 and the UGSP [4], SGSP [33], DT [32], with and without BR [58] and P [5], and FHMM [34], as presented in [5].

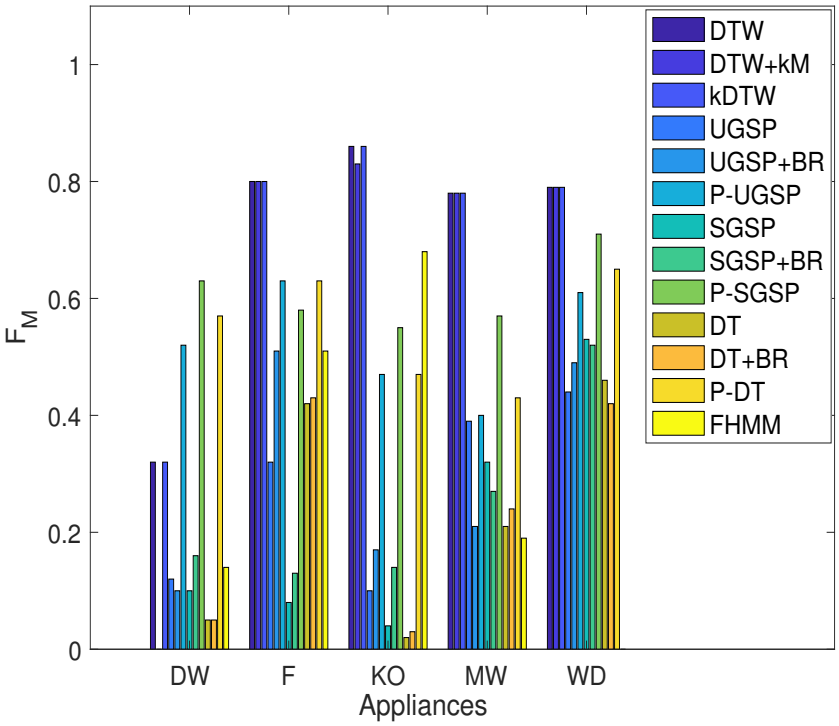


Figure B.1: Performance evaluation using F_M for the proposed DTW+kM and kDTW (Table B.1) with benchmarks the DTW-based method proposed in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REDD House 1, as presented in [5].

Figure B.2 shows the comparative results, as found in Table B.2, in terms of estimation accuracy using Acc_i, between the proposed methods of this chapter, the DTW-based method from Chapter 4 and the UGSP [4], SGSP [33], DT [32], with and without BR [58] and P [5], and FHMM [34], as presented in [5].

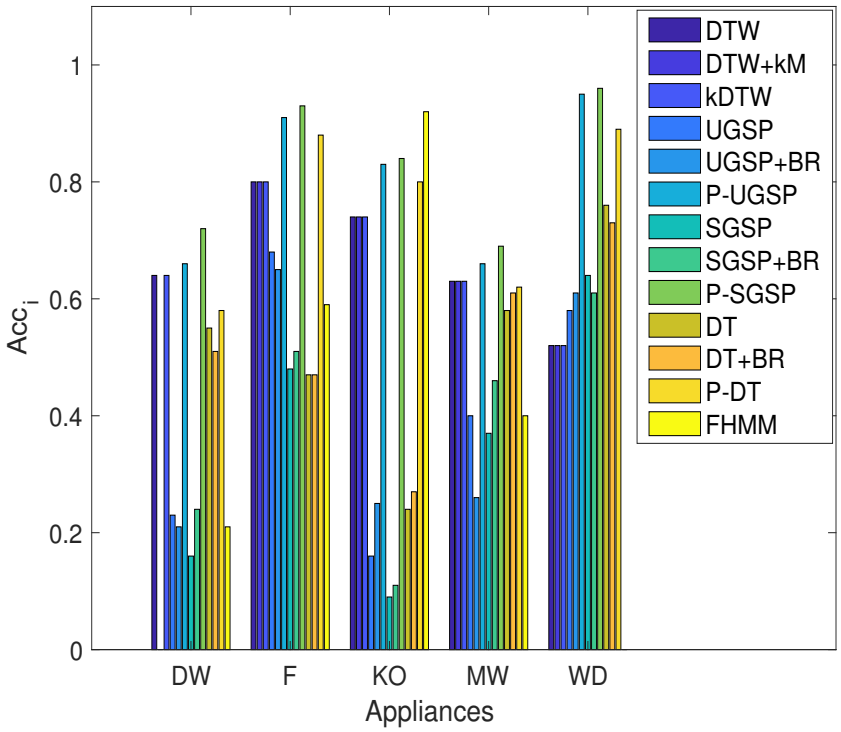


Figure B.2: Performance evaluation Acc_i for the proposed DTW+kM and kDTW (Table B.2) with benchmarks the DTW-based method proposed in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REDD House 1, as presented in [5].

B.2 Disaggregation Results for House 2 from REDD Dataset

Table B.3 shows the classification accuracy obtained using F_M for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for house 2 from REDD [1] , as presented in [5]. Table B.4 shows the estimation accuracy per appliance using Acc_i for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for house 2 from REDD dataset [1], as presented in [5].

Table B.3: Classification Accuracy using F_M of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REDD House 2, as presented in [5].

	DTW			*UGSP*[4]			*SGSP*[33]			*DT*[32]			
Apps	(1)	(2)	(3)	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.36	0.20	0.36	0	0.21	**0.82**	0.12	0.21	0.80	0.08	0.16	0.77	0
F	**0.82**	0.75	**0.82**	0.12	0.11	0.24	0.17	0.16	0.31	0.30	0.28	0.29	**0.82**
KO1	0.84	0.75	0.84	0.07	0.13	0.82	0.01	0.09	**0.90**	0.18	0.23	0.85	0.52
KO2	0.79	**0.82**	**0.82**	0.33	0.37	0.60	0.45	0.51	0.53	0.24	0.33	0.43	0.38
MW	**0.90**	0.87	**0.90**	0.31	0.34	0.69	0	0.01	0.83	0	0.05	0.78	0.24
S	0.53	0.52	**0.57**	0	0	0.26	0	0	0.23	0.01	0.01	0.33	0.01

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, S=Stove. (1) represents the original DTW-based method in Chapter 4, (2) the DTW+kM, and (3) the kDTW, that uses DTW as a final refinement step. **Bold** values represent the method/methods with the highest obtained classification accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Table B.4: Estimation Accuracy using Acc_i of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REDD House 2, as presented in [5].

	DTW			*UGSP*[4]			*SGSP*[33]			*DT*[32]			
Apps	(1)	(2)	(3)	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.61	0.56	0.61	0.50	0.52	**0.78**	0.32	0.33	0.77	0.15	0.13	0.64	−2.13
F	0.69	0.69	0.69	0.04	0.01	0.55	0.07	0.07	0.64	0.09	0.07	**0.72**	0.69
KO1	0.41	0.41	0.41	0.17	0.24	**0.89**	0.09	0.15	0.86	0.32	0.34	0.86	−0.4
KO2	**0.81**	**0.81**	**0.81**	0.58	0.06	0.72	0.55	0.49	0.73	0.59	0.54	0.66	0.12
MW	0.64	0.64	0.64	0.54	0.56	0.82	0.4	0.21	**0.85**	0.65	0.63	0.76	0.11
S	**0.75**	0.73	**0.75**	0.02	0	0.45	0.02	0.04	0.44	0.24	0.19	0.43	−4.49

Note: Apps=Appliances, DW=Dishwasher, F=Fridge, KO=Kitchen Outlet, MW=Microwave, S=Stove. (1) represents the original DTW-based method in Chapter 4, (2) the DTW+kM, and (3) the kDTW, that uses DTW as a final refinement step. **Bold** values represent the method/methods with the highest obtained estimation accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Figure B.3 shows the comparative results, as found in Table B.3, in terms of classification accuracy, using F_M, between the proposed methods of this chapter, the DTW-based method from Chapter 4 and the UGSP [4], SGSP [33], DT [32], with and without BR [58] and P [5], and FHMM [34], as presented in [5].

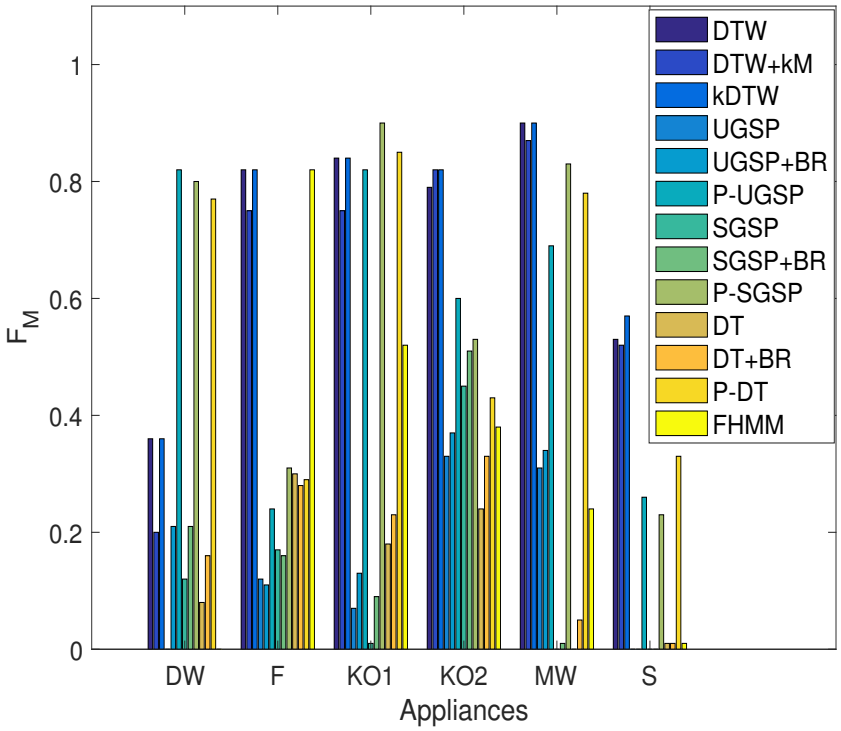


Figure B.3: Performance evaluation using F_M for the proposed DTW+kM and kDTW (Table B.3) with benchmarks the DTW-based method proposed in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REDD House 2, as presented in [5].

Figure B.4 shows the comparative results, as found in Table B.4, in terms of estimation accuracy using Acc_i, between the proposed methods of this chapter, the DTW-based method from Chapter 4 and the UGSP [4], SGSP [33], DT [32], with and without BR [58] and P [5], and FHMM [34], as presented in [5].

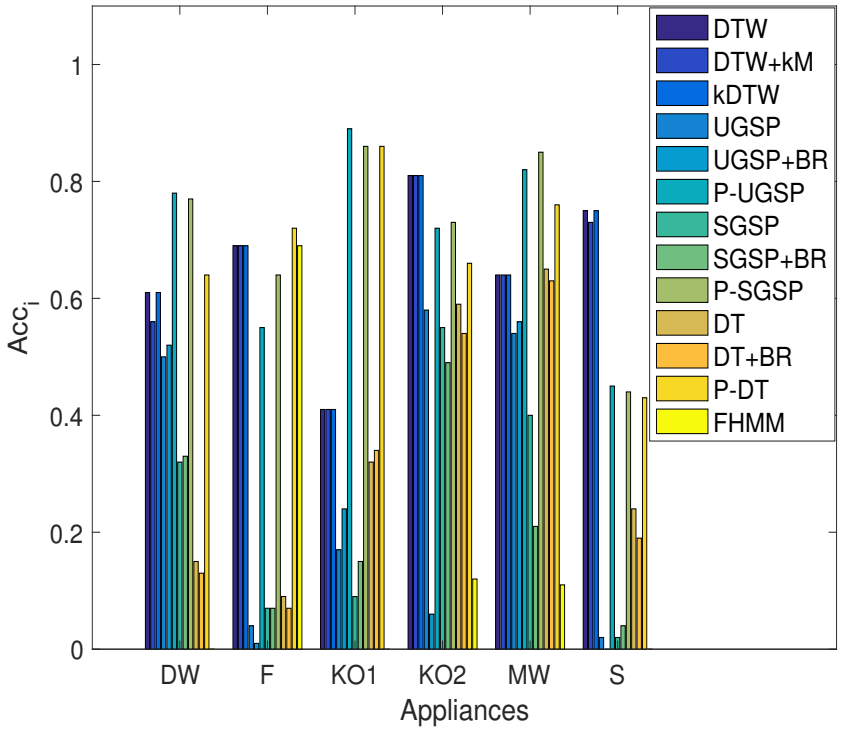


Figure B.4: Performance evaluation using Acc_i for the proposed DTW+kM and kDTW (Table B.4) with benchmarks the DTW-based method proposed in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REDD House 2, as presented in [5].

B.3 Disaggregation Results for House 6 from REDD Dataset

Table B.5 shows the classification accuracy obtained using F_M for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, GSP [144], GSP+FS [33],GSP+SA [33], SGSP [33], UGSP [4], DT [32], HMM [35], for REDD House 6, as presented in [4, 33].

Table B.5: Classification Accuracy using F_M for the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, GSP [144], GSP+FS [33],GSP+SA [33], SGSP [33], UGSP [4], DT [32], HMM [35], for REDD House 6, as presented in [4, 33].

	DTW			*GSP*						
Apps	*DTW*	*DTW+kM*	*kDTW*	*GSP*	*GSP+FS*	*GSP+SA*	*SGSP*	*UGSP*	*DT*	*HMM*
AC	**0.93**	0.67	**0.93**	0.49	0.74	0.73	0.89	0.88	0.89	0.12
B	0.67	0.33	**0.75**	–	–	–	–	–	–	–
F	0.79	0.70	0.8	0.54	0.77	0.77	0.77	0.82	**0.99**	0.88
KO1	**1**	–	**1**	0.92	0.77	0.77	0.91	0.87	0	0
S	0.91	0.73	**1**	**1**	0.75	0.75	0.92	0.54	0.67	0

Note: Apps=Appliances, AC= Air Conditioning, B=Bathroom GFI, F=Fridge, KO=Kitchen Outlet, S=Stove. **Bold** values represent the method/methods with the highest obtained classification accuracy per appliance. The KO1, as found in REDD dataset [1], represents the microwave appliance presented in [4, 33].

Figure B.5 shows the comparative results, as found in Table B.5, in terms of classification accuracy, using F_M, between the proposed DTW+kM and kDTW methods, the DTW-based method from Chapter 4 and GSP [144], GSP+FS [33],GSP+SA [33], SGSP [33], UGSP [4], DT [32], HMM [35], for REDD House 6, as presented in [4, 33].

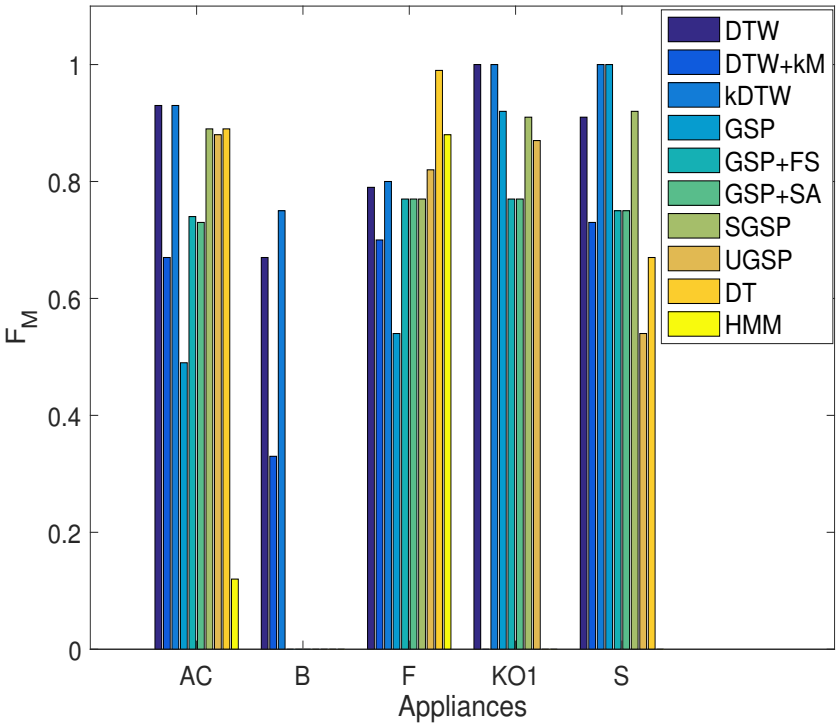


Figure B.5: Performance evaluation using F_M for the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, GSP [144], GSP+FS [33],GSP+SA [33], SGSP [33], UGSP [4], DT [32], HMM [35], as presented in [4, 33], for the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method presented in Chapter 4, for REDD House 6.

B.4 Disaggregation Results for House 2 from REFIT Dataset

Table B.6 shows the classification accuracy obtained using F_M for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for house 2 from REFIT dataset [2, 3], as presented in [5]. Table B.7 shows the estimation accuracy per appliance using Acc_i for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for house 2 from REFIT dataset [2, 3], as presented in [5].

Table B.6: Classification Accuracy using F_M of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REFIT House 2, as presented in [5].

	DTW			*UGSP*[4]			*SGSP*[33]			*DT*[32]			
Apps	(1)	(2)	(3)	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.27	–	0.27	0.32	0.41	**0.79**	0.54	0.52	0.73	0.64	0.66	0.73	0.23
FFZ	**0.61**	0.27	0.41	0.47	0.52	0.42	0.32	0.33	0.59	0.30	0.54	0.33	0.50
K	**0.96**	0.84	**0.96**	0.66	0.68	0.88	0.77	0.67	0.90	0.45	0.42	0.76	0.06
MW	**0.85**	0.71	**0.85**	0.42	0.48	0.73	0.35	0.44	0.84	0.32	0.31	0.80	0.09
T	**0.83**	0.62	**0.83**	0.40	0.25	0.54	0.52	0.53	0.64	0.6	0.54	0.55	–
WM	**0.69**	0.66	**0.69**	0.24	0.21	0.23	0.32	0.29	0.43	0.37	0.32	0.48	0.09

Note: Apps=Appliances, DW=Dishwasher, FFZ=Fridge-Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine. (1) represents the original DTW-based method in Chapter 4, (2) the DTW+kM, and (3) the kDTW, that uses DTW as a final refinement step. **Bold** values represent the method/methods with the highest obtained classification accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Table B.7: Estimation Accuracy using Acc_i of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REFIT House 2, as presented in [5].

	DTW			*UGSP*[4]			*SGSP*[33]			*DT*[32]			
Apps	(1)	(2)	(3)	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.34	–	0.34	0.33	0.40	0.42	0.40	0.43	**0.67**	0.28	0.31	0.61	0.30
FFZ	0.69	0.63	0.63	0.42	0.34	0.77	0.58	0.56	**0.80**	0.37	0.32	0.73	0.24
K	**0.91**	0.88	**0.91**	0.49	0.68	0.83	0.51	0.51	0.85	0.41	0.39	0.76	−0.34
MW	**0.79**	0.76	**0.79**	0.51	0.48	0.64	0.55	0.53	0.65	0.33	0.42	0.64	−3.17
T	**0.85**	0.80	**0.85**	0.37	0.48	0.62	0.31	0.36	0.66	0.26	0.22	0.58	–
WM	**0.73**	**0.73**	**0.73**	0.46	−1.59	0.43	0.47	0.19	0.48	0.44	0.17	0.35	−1.84

Note: Apps=Appliances, DW=Dishwasher, FFZ=Fridge-Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine. (1) represents the original DTW-based method in Chapter 4, (2) the DTW+kM, and (3) the kDTW, that uses DTW as a final refinement step. **Bold** values represent the method/methods with the highest obtained estimation accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Figure B.6 shows the comparative results, as found in Table B.6, in terms of classification accuracy, using F_M, between the proposed methods of this chapter, the DTW-based method from Chapter 4 and the UGSP [4], SGSP [33], DT [32], with and without BR [58] and P [5], and FHMM [34], as presented in [5].

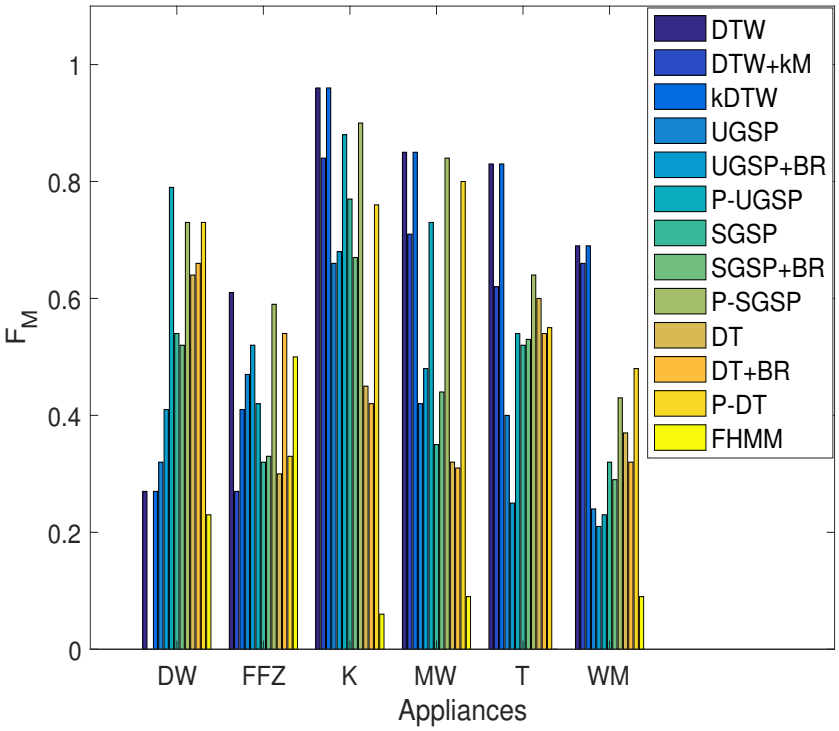


Figure B.6: Performance evaluation using F_M for the proposed DTW+kM and kDTW (Table B.6) with benchmarks the DTW-based method proposed in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REFIT House 2, as presented in [5].

Figure B.7 shows the comparative results, as found in Table B.7, in terms of estimation accuracy using Acc_i, between the proposed methods of this chapter, the DTW-based method from Chapter 4 and the UGSP [4], SGSP [33], DT [32], with and without BR [58] and P [5], and FHMM [34], as presented in [5].

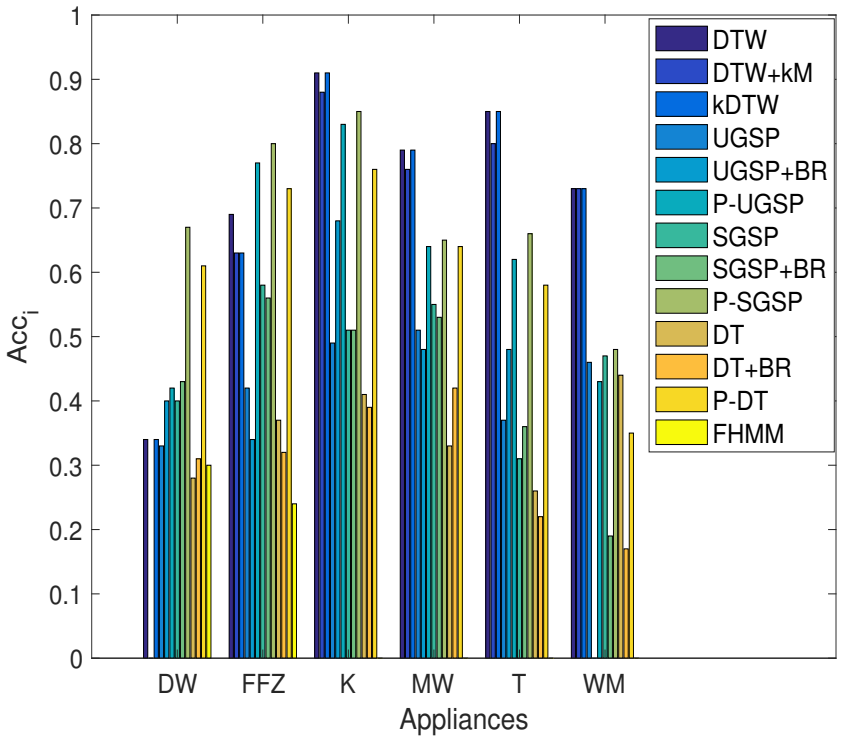


Figure B.7: Performance evaluation using Acc_i for the proposed DTW+kM and kDTW (Table B.7) with benchmarks the DTW-based method proposed in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REFIT House 2, as presented in [5].

Table B.8 shows the normalised total power consumption estimation error (TER) for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for house 2 from REFIT dataset [2, 3], as presented in [5].

Table B.8: Normalised total power consumption estimation error (TER) for the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REFIT House 2, as presented in [5].

	DTW			*UGSP*[4]			*SGSP*[33]			*DT*[32]			
Apps	(1)	(2)	(3)	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
DW	0.43	–	0.43	0.73	**0.08**	0.66	0.47	0.13	0.33	0.64	0.35	0.49	0.75
FFZ	0.70	0.73	0.73	**0.10**	0.53	0.31	0.19	0.42	0.27	0.47	0.67	0.31	0.38
K	**0.019**	0.031	**0.019**	0.34	0.06	0.05	0.27	0.11	0.04	0.31	0.25	0.15	0.41
MW	**0.09**	**0.09**	**0.09**	0.43	0.17	0.34	0.52	0.16	0.28	0.66	0.59	0.63	13.63
T	**0.21**	0.33	**0.21**	0.83	0.81	0.82	0.9	0.88	0.73	0.78	0.71	0.42	7.53
WM	0.42	0.42	0.42	0.17	0.40	0.09	0.32	0.52	**0.10**	0.28	0.33	0.22	3

Note: Apps=Appliances, DW=Dishwasher, FFZ=Fridge-Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine. (1) represents the original DTW-based method in Chapter 4, (2) the DTW+kM, and (3) the kDTW, that uses DTW as a final refinement step.**Bold** values represent the method/methods with the lowest TER error per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

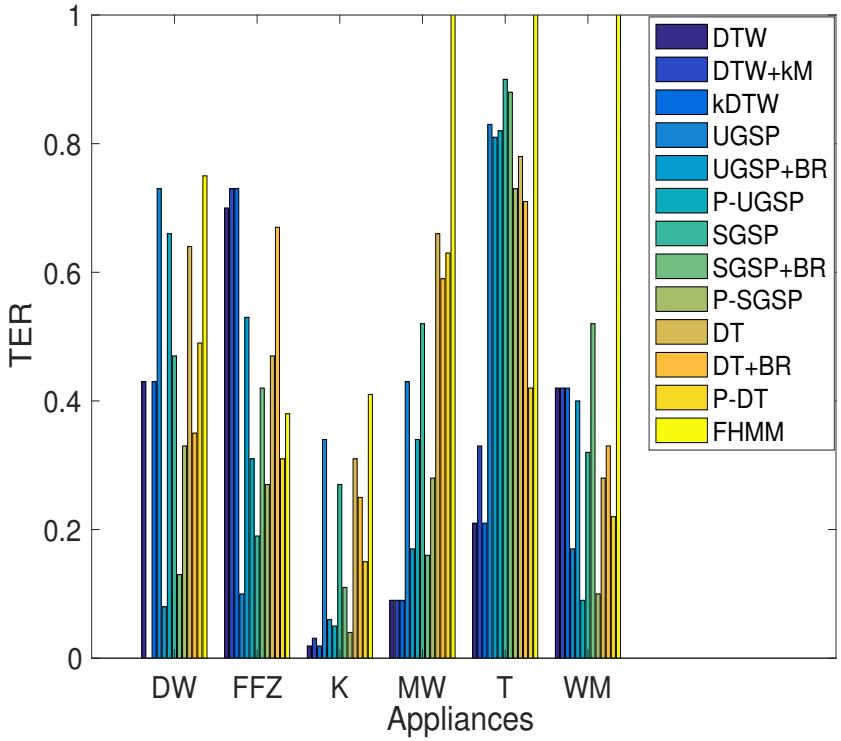


Figure B.8: Performance evaluation using TER for the proposed DTW-based NILM method, based on Table B.8, for the proposed DTW+kM and kDTW, with benchmarks the DTW-based method proposed in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REFIT House 2, as presented in [5].

Figure B.8 provides a graphical representation of the normalised total power consumption estimation error (TER) for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for house 2 from REFIT dataset [2, 3], as presented in [5].

B.5 Disaggregation Results for House 17 from REFIT Dataset

Table B.9 shows the classification accuracy obtained using F_M for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for house 17 from REFIT dataset [2, 3], as presented in [5]. Table B.10 shows the estimation accuracy per appliance using Acc_i for the proposed DTW+kM and kDTW methods, proposed in Chapter 5, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for for house 17 from REFIT dataset [2, 3], as presented in [5].

Table B.9: Classification Accuracy using F_M of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REFIT House 17, as presented in [5].

	DTW			*UGSP*[4]			*SGSP*[33]			*DT*[32]			
Apps	(1)	(2)	(3)	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
FZ	0.52	0.48	0.48	0.39	0.40	0.60	0.47	0.42	0.78	0.56	0.53	**0.80**	0.74
K	0.82	0.82	0.82	0.79	0.61	0.84	0.81	0.77	**0.96**	0.62	0.63	0.95	0.37
MW	0.68	0.68	0.68	0.15	0.24	0.55	0.28	0.31	0.77	0.11	0.13	**0.79**	–
T	**0.88**	**0.88**	**0.88**	0.23	0.07	0.81	0.35	0.36	0.76	0.39	0.33	0.46	–

Note: Apps=Appliances, FZ=Freezer, K=Kettle, MW=Microwave, T=Toaster, WM=Washing Machine. (1) represents the original DTW-based method in Chapter 4, (2) the DTW+kM, and (3) the kDTW, that uses DTW as a final refinement step. **Bold** values represent the method/methods with the highest obtained classification accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Table B.10: Estimation Accuracy using Acc_i of the proposed DTW+kM and kDTW methods, with benchmarks the proposed DTW-based method in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34], for REFIT House 17, as presented in [5].

	DTW			*UGSP*[4]			*SGSP*[33]			*DT*[32]			
Apps	(1)	(2)	(3)	*UGSP*	*UGSP+BR*	*P-UGSP*	*SGSP*	*SGSP+BR*	*P-SGSP*	*DT*	*DT+BR*	*P-DT*	*FHMM*[34]
FZ	0.65	0.64	0.64	0.04	0.43	0.64	0.21	0.26	0.63	0.26	0.21	0.68	**0.71**
K	**0.90**	**0.90**	**0.90**	0.12	0.66	0.79	0.32	0.31	0.80	0.43	0.43	0.77	0.2
MW	**0.68**	**0.68**	**0.68**	0.47	0.23	0.55	0.47	0.54	0.55	0.59	0.57	0.61	–
T	**0.88**	**0.88**	**0.88**	0.37	0.23	0.80	0.42	0.39	0.81	0.54	0.50	0.71	–

Note: Apps=Appliances, FZ=Freezer, K=Kettle, MW=Microwave, T=Toaster. (1) represents the original DTW-based method in Chapter 4, (2) the DTW+kM, and (3) the kDTW, that uses DTW as a final refinement step. **Bold** values represent the method/methods with the highest obtained estimation accuracy per appliance. BR is the base load removal method, proposed in [58] and P is the GSP based pre-processing and refinement method proposed in [5].

Figure B.9 shows the comparative results, as found in Table B.9, in terms of classification accuracy, using F_M, between the proposed methods of this chapter, the DTW-based method from Chapter 4 and the UGSP [4], SGSP [33], DT [32], with and without BR [58] and P [5], and FHMM [34], as presented in [5].

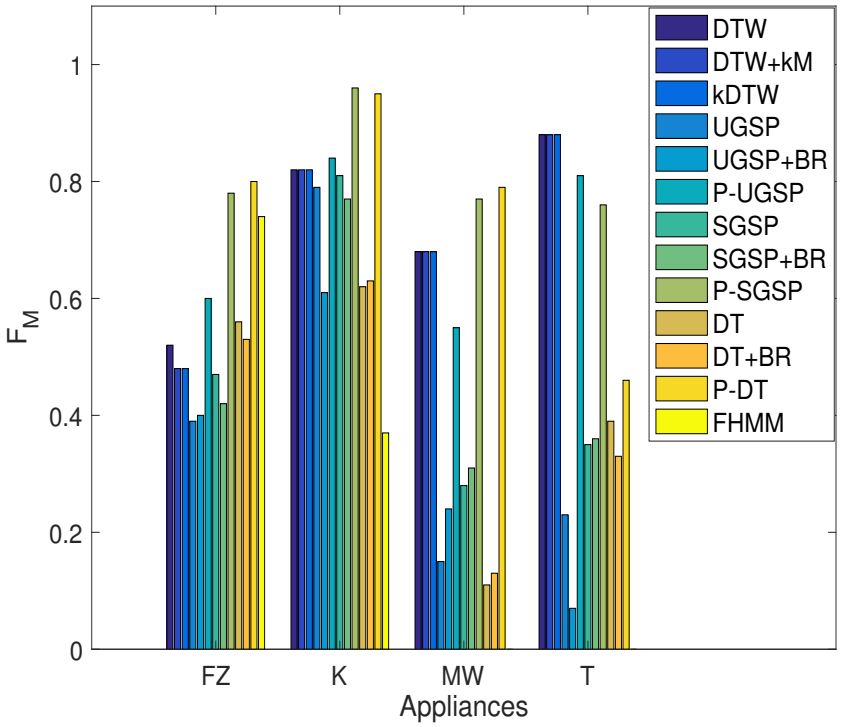


Figure B.9: Performance evaluation using F_M for the proposed DTW+kM and kDTW (Table B.9) with benchmarks the DTW-based method proposed in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REFIT House 17, as presented in [5].

Figure **??** shows the comparative results, as found in Table B.7, in terms of estimation accuracy using Acc_i, between the proposed methods of this chapter, the DTW-based method from Chapter 4 and the UGSP [4], SGSP [33], DT [32], with and without BR [58] and P [5], and FHMM [34], as presented in [5].

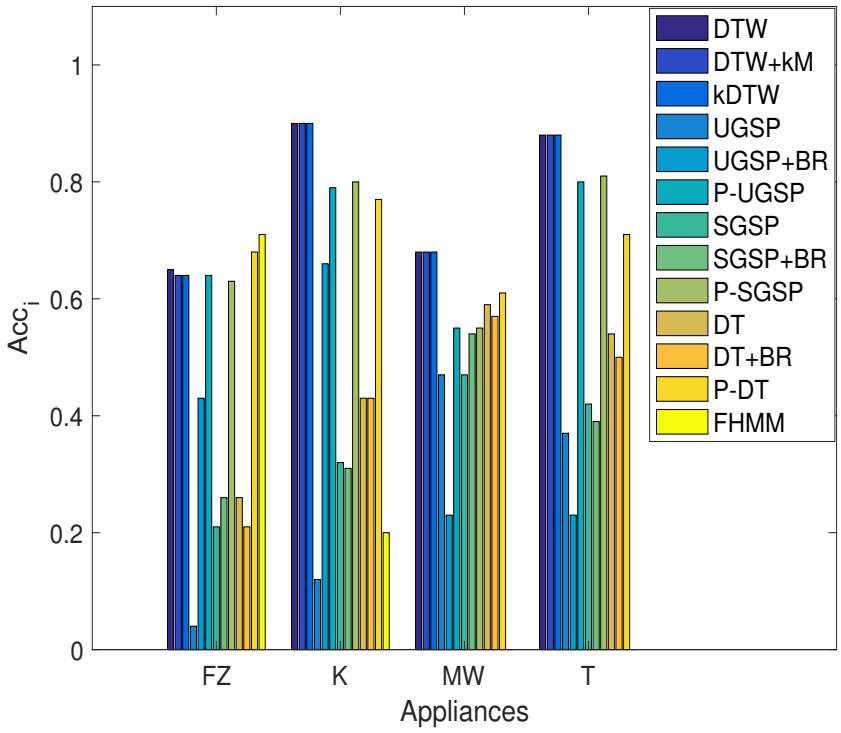


Figure B.10: Performance evaluation using Acc_i for the proposed DTW+kM and kDTW (Table B.10) with benchmarks the DTW-based method proposed in Chapter 4, UGSP [4], SGSP [33], DT [32], BR [58], P [5] and FHMM [34] for REFIT House 17, as presented in [5].

Appendix C

Supplementary IoT Background

C.1 Introduction

The advancement of various technologies, such as sensors, software, data storage and data analysis, have led researchers to conceive the idea of *Internet of Things* (IoT), which could potentially create a link between all technologies and fields. A further explanation of the concept behind IoT, its structure, applications and challenges can be found in Section C.2.

C.2 Internet of Things

Internet of Things (IoT) represents one of the most trending concepts in current technology and research community. IoT is a global "cloud-like" network, that could interconnect various industries through the utilisation of smart sensors, *Radio Frequency Identification* (RFID), actuators, and other enabling technologies that are currently used in each industry[7–10]. A graphical example of IoT can be found in Figure C.1.

The applications of IoT are numerous and according to Atzori *et al.* [7], they can be classified in four domains: (a) transportation and logistics, (b) healthcare, (c) smart environment and (d) personal and social domain. Although these applications will be highly beneficial to all parties involved (namely individuals, industries, etc), many challenges arise in terms of: (a) how the data will be stored (b) how the data will be processed and (c) how secure this network will be in order to maintain privacy [7–10].

Figure C.1: The Internet of Things [206].

C.2.1 IoT Enabling Technologies

The technologies that can enable the actualisation of IoT can be summarised in the following, as seen in [7–10]: (a) *Radio Frequency Indentification* (RFID), (b) *Wireless Sensor Networks* (WSN), (c) *Middleware*, (d) *Data Storage and analytics* and (e) *Data Visualisation*.

C.2.1.1 Radio Frequency Identification

Radio Frequency Identification (RFID) is the technology that allows the identification and tracking of objects using tags with embedded microchips. The RFID process requires a sensor that uses radio waves in order to "read" the tags [207].

RFID technology has been applied successfully in various sectors, such as logistics and supply chain management, pharmaceutical industry, agriculture, retail, healthcare etc. RFID is used in retail industry for inventory and replenishment purposes, in addition to loss prevention purposes with RFID enabled security tags. In agriculture, RFID can be used in order to identify and track livestock in farms and increase efficiency and production of *precision agriculture* [208]. Another interesting application is the implementation of RFID technology in biometric passports, which allows the access of ID, biometric and travelling data, without physically accessing the passport. Libraries are using RFID systems in order to have an accurate management system, that allows

book tracking, and therefore reduce theft and losses [207].

Essentially, all the above examples represent some of the objects that can be monitored through the incorporation of RFID technology in the IoT.

C.2.1.2 Wireless Sensor Networks

Wireless sensor networks (WSN) consist of a variety of sensor nodes distributed in a wide area that form a wireless network with the purpose of monitoring different environmental and application-specific parameters [7–10, 209]. According to Zhang *et al.* [209], WSNs applications can be separated into monitoring and tracking related applications, depending on whether they are used for monitoring various features, or tracking objects, animals, etc, through cooperation with RFID systems.

WSNs can be used for monitoring different assets, such as earthworks, bridges, embankments, by monitoring various environmental parameters and essentially provide predictive and early notification methods through data mining for avoiding future failures due to floods, seismic activities and other natural extreme phenomena. Environmental monitoring can also lead to agricultural advancements, such as increased and better production and assist in the field of precision agriculture. Furthermore, by monitoring soil, water and atmospheric parameters, climate changes and pollution can be studied, and therefore useful information can be inferred [209].

Moreover, in healthcare, WSNs can assist in more efficient and real-time monitoring of patients and therefore better medical services [209]. At home monitoring through a sensor and actuator network implemented in various areas of the residence can be useful, especially for elderly patients and patients with serious heart or breathing conditions, or mobility issues. Through this network, readings from medical devices available in the patient's residence could be obtained and through motion sensors a daily activity could be recorded. This data could provide notification of unusual events, such as unusual measurements obtained from the medical devices and extended lack of motion presence, which could mean, for example, a serious fall/injury [210, 211].

Similarly, WSNs can be incorporated in the concept of smart homes and Smart Grid. For the case of smart homes, the application is very similar with the at home patient monitoring, as similar features can be monitored, namely motion, temperature, etc [212] using various sensors and actuators. Another application is in the monitoring of

renewable energy installations, such as wind and solar farms, monitoring of power quality, transmission line and cabling, residential energy management, building automation, and many more [213, 214].

C.2.1.3 Middleware

Middleware is a software layer between different software applications, that allows the integration of different technologies and development of services. For the case of IoT and with the variety of objects/devices connected, middleware can assist in the smooth application-specific development [7, 10].

According to Atzori *et al.* [7] and Lee *et al.* [10], the middleware architecture for the IoT can be described as a *Service Oriented Architecture* (SOA) approach. SOA is a flexible architecture that allows services to be provided through a network. Its flexibility is mainly based on the easy and fast adaptation when changes occur, software and hardware reuse, as each service, both new and old, do not require a specific technology [7]. These advantages make SOA an optimal solution for IoT, where heterogenous objects/devices, applications and services are required to exist in the same network.

C.2.1.4 Data Storage

Data storage is a very important part of the IoT, as an enormous amount of diverse data will be produced. Therefore data storage requires massive capacity and fast processing speed for real-time data analysis and notifications [8, 10].

C.2.1.5 Data Analytics

Data mining and machine learning algorithms are required for a further understanding of the collected data and the deduction of useful information [8, 10]. More recently, the idea of Cloud based data storage and data analytics seemed a more appropriate solution especially for the case of real-time data processing and analysis [8, 10].

C.2.1.6 Data Visualisation

Data Visualisation is an essential part of IoT applications, at it allows the representation of data in a more interactive way, that could be accessible to end-users and third parties, such as utilities and policy makers [8, 10].

C.2.2 IoT Applications

As IoT has a plethora of applications, therefore it is important to present an overview of these applications, in order to enable the understanding of the importance of IoT implementation. The applications will be presented under the main categories proposed in [7], as already discussed in Section C.2.

C.2.2.1 Transportation and Logistics Domain

The incorporation of smart systems in both vehicles and transport infrastructure allows the implementation of the IoT concept in both transportation and logistics industries and can be highly beneficial for our society. These applications can be categorised in those related to transportation and logistics and supply chain management, as found in [7].

- *Transportation*

  WSNs for monitoring transport infrastructure, namely railways, roads and canals, could provide live information about traffic congestion, accidents and even infrastructure failures, that can affect the smooth operation of the transport network [7–9]. Navigation systems could be automatically updated using both network and vehicle tracking information and thus provide the most convenient routes to network users, which can positively affect both duration, safety and cost of transportation [7, 8]. Apart from the economical and social impact, this application could have environmental benefits, as it could reduce both emissions and traffic noise by minimising traffic congestion [8].

  Vehicle tracking could also assist in the allocation of the most suitable parking space and the mobile billing for any services related with use of the transport network (parking and toll fees) [7, 9]. Furthermore, the use of smart cameras

in vehicles, which is already used for parking assistance, could be able to detect vehicles or other "*motion-based objects*" and recognise pedestrians, which would reduce collisions and fatalities [9, 215]. Keller *et al.* have proposed in [215] a prototype system for object/human detection and "*automatic braking and evasive steering*" when the object/human is in proximity.

Autonomous vehicles (AV) or *self-driving vehicles* represent a further expansion of the above applications, as they incorporate a variety of sensors and advanced control systems, in order to assist (as in [9, 215]) or even replace the driver in the fully automated AVs, with companies such as Waymo (Google), Uber, Tesla, Cruise Automation and Navya already leading the AV field [216–222]. AVs could potentially reduce traffic congestion and traffic accidents, as according to researches both are often related to human error, especially for the case of accidents, more than 90% occur due to the driver's error [223–225]. Despite the potential of AVs, there have been a number of mainly minor incidents/accidents involving material damage and minor injuries, and two fatal incidents [226, 227], thus new legislations are required to ensure the safety and smooth operation of AVs in coordination with the existing non-automated vehicles and current transport infrastructure, privacy and cybersecurity [225, 228–230].

- *Logistics and Supply Chain Management*

  Similarly, IoT could be applied in the field of logistics and supply chain management, which is closely related to transportation. The increased use of RFID tags and *Near Field Communications* (NFC) allows real-time monitoring of all individual parts of the supply chain, such as raw materials, production, packaging and inventory, transportation, storage, distribution, sales and returns of products [7, 9].

  Product and transportation information could be obtained in real-time, which will allow all involved parties (manufacturing and transportation companies) to react faster to changes related to the products or the transportation process [7]. In terms of transportation, the dynamic updates of transport network and any traffic problems would allow companies to optimise both transportation of goods and delivery schedule, by adapting to any changes fed by the transport network

[7–9]. Furthermore, *Enterprise Resource Planning* (ERP) systems will be able to provide real-time information to companies and retail staff, which will lead to better customer service and product replenishment [7].

Another important application of IoT is in the field of *Food Supply Chain* (FSC), which involves production, distribution and consumption of food products, especially those that can degrade easily, when the correct conservation standards are not met, sensors can be used for monitoring parameters, such as temperature and humidity, to ensure the quality of the products throughout the FSC [7, 9, 10, 231]. Furthermore, *precision agriculture* has gained a lot of interest in the research community, with various smart agriculture systems being proposed in recent works, such as [13–15], for monitoring various environmental parameters and smart solutions for performing agricultural tasks, such as weeding, spraying, moisture sensing.

C.2.2.2 Healthcare Domain

Healthcare domain represents one of the most challenging and *"sensitive"* domains, as any malfunctions and delays in the healthcare system could affect both patients and healthcare personnel. According to [7–10], the implementation and usage of IoT in healthcare could provide an improvement in the services and can lead to a more personalised healthcare system for patients.

Atzori *et al.* [7] have categorised the IoT applications in healthcare into the following four categories: tracking of objects and people, identification and authentication of people, data collection and sensing.

- *Tracking*: Tracking can provide real-time information about the location of both people and objects, in order to improve the operation of healthcare facilities and assist with access identification to designated and restricted areas [7]. Inventory and equipment tracking can provide information about maintenance of equipment, availability and monitoring of equipment and medical products [7, 232].

- *Identification and authentication*: According to Atzori *et al.* [7], patient identification can minimise the mistakes regarding wrong drugs/doses/operations etc, and can provide an up-to-date electronic medical record and furthermore can

prevent mother-infant mismatching incidents. For the case of healthcare assets, such as instruments, equipments, materials, and medical products, identification and authentication can be used for security purposes, in order to avoid thefts or losses of the assets [7].

- *Data collection*: Data can be obtained automatically, through the integration of RFID and WSN technologies with existing health information systems and clinical application technologies, in order to obtain a more efficient healthcare system, and could be potentially applied for disease spread modelling and containment [7–9].

- *Sensing*: The application of sensors in healthcare, as discussed in [7], can facilitate in a more accurate diagnosis and real-time monitoring of patient's health indicators, such as heart rate, temperature, and could be applied even for *In-Home Healthcare* (IHH) services [233, 234].

  Another interesting sensor approach has been proposed in [17], where the authors have developed an *"intelligent patient bed"* for disabled patients. A contactless eye tracker has been implemented for the communication between the patient and the smart bed. The proposed system senses the gaze position and through a GUI system, the patient's intention is recognised and the specific action/service is performed automatically. For example, the system allows a call to the nurse and even controls the position of the bed.

In general the incorporation of various sensors in healthcare for both hospitals and in-home care, can allow a more automated, intelligent, personalised and efficient healthcare system, which can be fulfilled with the interconnection of those smart systems through IoT.

C.2.2.3 Smart Environment Domain

Smart environment domain is one of the most extended application domains in IoT, and can be categorised in *smart cities*, *smart industry*, *smart grid* and *smart leisure environment*.

- *Smart Cities*

  Smart cities represent the cities of the future, where *information and communications technology* (ICT) will be used in order to interconnect urban infrastructures and services and therefore improve the quality of life in an urban environment [235]. The IoT framework and the related technologies are essential for the development of smart cities, as they will achieve the interconnection of all devices and services of the city in "a cloud-like" network, making accessible services and data to all parties involved [235].

  The implementation of sensors and WSNs in houses and buildings could provide automated adjustment of the room temperature depending on the weather conditions and predefined user preferences [7, 8]. Remote control of home or office appliances, such as air conditioners, fridge, and washing machine could increase energy saving and reduce accidents caused by appliances that have been left switched on due to user's error [7, 8, 236]. According to Patel *et al.* [236], smart fridges could allow an inventory check, that would notify users about products that are expired or close to expire and provide a to buy list depending on user's specifications. Smartphone applications could be used to remotely adjust oven temperature and monitor washing machine cycles [236]. Security systems, such as cameras, alarms and motion sensors integrated in doors and windows, could provide real-time notifications to both owners and police/security services, in order to prevent intrusions [7, 236].

  The interconnection of different surveillance systems in the context of smart cities, such as CCTVs and traffic cameras, would increase public safety and even assist in the creation of a unified smart emergency system in case of an accident, natural disaster, crime or even disease outbreak [235–237]. Although the advantages of such a system are important, many concerns could arise regarding the protection of people's privacy and their sensitive information. These IoT challenges will be further discussed in Section C.2.3.

  Monitoring of the structural health of various infrastructures, such as bridges, buildings, embankments, is possible with the use of sensors and WSNs in IoT and can improve security and safety and predict catastrophic failures. For example,

in [238], a system was proposed for monitoring the pressure of a building using piezoelectric transducers (PZT) and predict catastrophic damage. The authors have used Butterworth for cleaning the noise of the PZT signals, *Cross Correlation* for detecting damage and a mathematical model for identifying the location and size of the damage, which were all validated successfully in a lab environment [238].

Another IoT application in a smart city will be the implementation of smart traffic management systems, that will provide real-time information to users allowing their secure and efficient transportation around the city, by incorporating all the smart attributes of IoT based transportation, as discussed in Section C.2.2.1 [237, 239, 240]. Smart traffic management system will receive traffic information from moving vehicles through GPS systems and smartphone monitoring applications, and will be able to use inductive loop detectors, audio detectors and video detectors [237, 240]. Inductive loop detectors are detecting moving vehicles, when they are passing through the loop's magnetic area and audio detectors can estimate traffic and traffic congestion by analysing audio signals produced by the vehicles in the area of interest [237, 240]. Similarly, video detectors are using the feed of cameras attached to various spots on the motorways/roads [237]. Traffic lights will be able to communicate with sensors that can detect vehicles, cyclists and pedestrians and therefore automatically altering the cycle of the traffic lights by assessing the speed of approaching vehicles, and generally the traffic condition, in order to minimise traffic congestion and accidents [239].

Water management is an essential part of a smart city infrastructure, and it requires the efficient monitoring of the water supply network and the quality of drinking water, which can be achieved through IoT technologies [8]. Sensors can be deployed in all parts of the water network and can monitor water levels, water infrastructures and sewage system, providing real-time information about consumption, quality, flow and pressure and reporting the condition of the water infrastructures and identify future failures [8, 235]. Wastage of water will be reduced, as with the sensors' assistance, pipeline leakages would be identified and fixed, customers would be notified for excessive use, which could be the result of

a forgotten tab, and water theft would be prevented [241].

Another important everyday problem in our cities is waste managment, where in many cases the waste and recycling bins are overflowing due to the current waste managements systems, that normally allow collection either at specific times or specific days per area. Introduction of intelligent waste and recycling containers could provide real-time information about the level of waste and through this information the waste collection services could optimise their everyday collection routes, prioritising on the "almost full" containers and "ignoring" the containers that are not in full capacity [236, 240]. This smart waste management system could reduce the cost of waste collection and improve the quality of recycling, as consumers tend to dispose recycling packages in normal waste bins, when recycling bins are full [240].

- *Smart Industry*

  Deployment of RFID tags and WSNs in an industrial environment could increase the automation in production, and therefore its efficiency [7]. RFID readers could identify the production stage, and, through WSNs, automated machines could receive notifications to collect the parts for the next production stage [7]. Wireless networks incorporated into the machines could monitor their workload, and could generate emergency shutdown, if this exceeds the predefined thresholds. ERP systems will receive real-time updates, and will have access to information about the production process [7].

  Furthermore, these sensors, by monitoring the equipment's condition, could potentially provide predictions of possible malfunctions and maintenance notifications, before any critical failure occurs [236]. Environmental parameters monitoring, such as temperature, toxic and oxygen levels, could ensure the safety inside the industrial plants and minimise occupational accidents, especially for industries that manufacture or use hazardous materials [236].

- *Smart Grid*

  Smart Grid is the concept of an intelligent electricity network, that will be able to integrate both existing and future transmission infrastructures and incorporate

multiple renewable resources [242]. According to Li *et al.* [242], smart grid will use a digital platform, that will allow monitoring of several elements of the energy network, such as power plants, transmission lines, transmission towers, distribution centres and consumers' premises and therefore could be easily integrated in the IoT framework [243]. Intelligent monitoring devices provide a *self-healing* attribute to smart grid, as they allow detection and recovery from faults, and pre-failure maintenance through early notifications [8, 11, 236, 242–246].

Additionally, smart meters allow a real-time bi-directional communication between customers and power distributors, providing important information about consumption, pricing, maintenance works scheduled in the customers' area, etc [8, 11, 236, 242–246]. Power generation is more efficient, as the network uses distributed energy resources, that can be allocated accordingly depending on the energy demand at any given time, through real-time updates from the customers' premises [8, 11, 236, 242–246]. In case of a local power generation failure, smart grid can provide alternative power supply through the distributed system.

Furthermore, the NILM concept represents one of the most interesting aspects of smart grid deployment, as it utilises data acquired from *Advanced Metering Infrastructure* (AMI) meters and smart sensors, in order to obtain useful information about the energy consumption of appliances for residential and office buildings. As already discussed in Section 1.3, NILM methods aim to disaggregate power consumption, identify and predict appliance use and even build an activity profile for households. The majority of proposed NILM methods, as it will be discussed in more details in Chapter 2, are focused on offline disaggregation, but with the advancement in IoT technologies and the processing capabilities of both smart meters and database servers, a real-time disaggregation either performed at the smart meters level or at the database/cloud level could present a better fit for an IoT based smart grid framework, as found in [129, 247–249].

- *Smart Leisure Environment*

  Leisure activities are an important part of individual lives and therefore an improvement of people's experiences through these activities is deemed as necessary. IoT technology could change the way people experience simple activities, such as

visiting a museum, a theme park or exercising in a gym.

According to Atzori *et al.* [7], a smart museum could potentially allow the local adjustment of the environmental conditions, in order to represent different historical periods, such Egyptian period, Ancient Greek or even ice age, and thus provide a unique experience for all visitors. Another interesting approach has been published in [250], where the authors have proposed a wearable device, that will be able to track the visitors throughout the museum using *Bluetooth Low Energy* transmitters, that are placed in different locations of the museum. Using image processing, they were able to identify the visitors' position and any observed exhibit, which allows their cloud processing centre to provide automatically relevant information and provide a personalised experience depending on the visitor's age [250].

Similarly, the incorporation of IoT technology into theme parks has an increased interest in recent years, as it could potentially provide a more personalised and enjoyable experience for visitors, through wearable devices or smart phone applications [251]. Disney, for example, has launched *MagicBand*, which is a wristband that uses RFID and long-range radio technology, which has the functionality of hotel keys, credit cards, tickets, etc [10, 252]. Through sensors that are implemented throughout the amusement park, MagicBand transmits real-time data about the location of the visitors, the chosen activities etc., which, for example, can allow the company to allocate more staff if overcrowded. According to Marr [252], Disney has applied for a patent regarding a system, that could recognise visitors through their shoe soles, using various sensors and cameras, which could assist further to a more personalised experience.

Furthermore, in a gym environment, personalised exercise routines could be uploaded to the smart training machines, which could identify each individual user through an RFID tag incorporated either in a wearable or a smart gym card [7]. The machines would be able to monitor the user's health indicators throughout the training session, which would allow the trainers to identify overtraining, or even cases where the user does not seem to put substantial effort through the session [7]. Through the historical monitoring of these data, this system could

create a personalised health profile and therefore allow the identification of any irregularities, that could potentially indicate a health problem.

C.2.2.4 Personal and Social Domain

The applications of IoT in the personal and social domain, according to Atzori *et al.* [7], allow the user to connect with other users through social networks and can be categorised into: *social networking*, *historical queries*, *losses* and *thefts*.

- *Social Networking*

  IoT in social networking could allow an automated update of the users' social networking applications and website through their social activities. The users could be able to predefine who will have access to their updates, and which events and activities will be automatically uploaded [7, 253]. This concept already exists in many social networking interfaces, but requires the users to update manually their events and activities. Through the advancement in smart devices, sensors and software applications, the user's devices could automate this procedure, without any involvement from the user's side.

  Smith *et al.* [253] have discussed the possibility of the incorporation of wearables, that can update social network status and interconnect with other users, which with the evolution of smart watches, such Apple Watch, could be accomplished in the near future. Currently some smart watches allow the use of applications such as Facebook Messenger and LinkedIn Pulse, but without providing the full capabilities of these networking applications.

- *Historical Queries*

  Queries of historical data of objects and events could allow users to study their past activities and behaviour [7, 253]. This could be beneficial, as users could identify how they use their time, and therefore change behaviours that are deemed unnecessary and reallocate more efficiently their everyday tasks, especially work related ones. This type of idea is already available through companies, such as Google, where search data can be used for personalised recommendations and advertisements, but the individual users do not have access to their data [7, 253].

A GUI interface could allow the users to access historical data through graphical representation, which is generally a more comprehensive format of reviewing data, especially for everyday users [7, 253].

- *Losses*

  This application would allow the user to locate devices and objects that are equipped with a form of RFID or sensor technology, such as locating a lost smartphone, laptop and even wallets and car keys [7, 253]. The user might have the option to define conditions about the location of their objects, and receive updates accordingly [7].

  Since [7], a variety of applications and devices offer the ability of tracking and locating lost items, such as Tile, Find My iPhone (Apple) and Find My Device (Google) [253–256]. Tile offers a variety of trackers that can be attached to everyday objects, such as keys, phones, laptops, wallets and luggages, and can locate the position of the object. They use Bluetooth technology, which has a limited range from 30m to 90m, depending on the tracker's model, thus they do not provide an overall tracking service that a GPS based-tracking system can provide [254].

  On the other hand, both Find My iPhone and Find My Device applications, provide the capability of locating the registered devices, such as smart phones, smart watches and tablets, as long as the devices are switched on and are connected to mobile data or Wi-Fi [255, 256]. Additionally, the user has the option of locking and even erasing any of the registered devices, in case of loss or theft [255, 256]. Apple provides the Find My Mac service which can locate, lock and erase any registered Mac computer, that connects to the internet through Wi-Fi [255].

- *Thefts*

  As already discussed in the above section, both for the case of loss and theft of an object, a tracking feature is required to monitor the object's location. This application can be enhanced to provide real-time notifications to the registered owner or security guards, when the object is removed without authorisation from a restricted area, that can be predefined by the user [7, 253].

C.2.3 IoT Challenges

Although the applications of IoT and its advantages are numerous, as already presented in Section C.2.2, there are some challenges that have arisen from both the architecture of IoT and the outcome data. These challenges can be summarised in the following categories: (a) *Data Management and storage*, (b) *data mining*, (c) *security* and (d) *privacy*, which will be further discussed in the following sections.

C.2.3.1 Data Management and Storage

As already mentioned through out this chapter, IoT, through the interconnected sensors and devices, will create a tremendous amount of data, that require enormous storage capacity. The question that arises is whether the current data storage systems and centres can facilitate this data collection process. According to Lee *et al.* [10], the cost of an IoT data storage system will result to limited company investments, thus data collection process should be prioritised depending on the needs that emerge. Additionally, the authors claim that more distributed data centres will be able to process data with higher efficiency and less processing time.

C.2.3.2 Data mining

Collecting and storing the IoT data might be treated as one of the most challenging aspects of IoT, but unfortunately the data themselves do not provide any useful information, especially for individuals without the relevant technical and scientific expertise. Inferring this useful information is a task that requires data mining techniques. Identifying events through event detection and signal processing methods, together with pattern matching/recognition and machine learning techniques, would be able to resolve this. It is essential though to find algorithmic solutions that fit each unique application and provide increased performance and low complexity implementation [8, 10].

C.2.3.3 Security

IoT, as already defined, is a network that interconnects various devices and sensors, and similarly with every other network, it is prone to security threats, even hackers' and cyber criminals' attacks. The IoT devices are distributed in various locations and due to their number makes it impossible to physically monitor and ensure the safety of each of them at all times, thus they can be possibly "attacked" and destroyed [7]. Furthermore, the IoT network is mainly a wireless network, and this makes it prone to eavesdropping and other types of attacks, such as *Denial of service* (DoS), *Distributed Denial of Service* (DDos), and malicious code injection [10, 257].

The majority of the devices do not have the capability for the implementation of complex data encryption schemes, which would fortify them in the case of a malicious attack. This make them vulnerable to attacks, and therefore endanger the IoT network as a whole [7, 10, 257].

According to Atzori *et al.* [7] and Xu *et al.* [258], another important security aspect in IoT is *authentication* and *data integrity*. *Authentication* requires the exchange of predefined messages between IoT nodes, which needs infrastructures and servers that would facilitate that. An example of devices, that cannot support this authentication process, are the passive RFID tags, which can only exchange a limited number of messages with the relevant servers or nodes [7, 258].

Another important question in terms of IoT security is how *data integrity* can be ensured, which essentially means, how the involved technology-oriented parties can keep the data safe from modifications from adversaries, and how they can assure the system will be able to detect such attacks [7]. Sensors in IoT, and especially RFID systems when connected to the internet, can present an "easy target" for data modification adversaries, as they are normally "unattended" [7].

C.2.3.4 Privacy

In Section C.2.2, a variety of applications have been presented, that aim to monitor personal information and activities, in addition to sensitive corporate and government-oriented data. Due to the nature of these data, many concerns have arisen regarding the maintenance of *privacy* for both individuals and companies. Individuals will not have a total control of what is monitored in IoT, even when they do not use actively any IoT service [7, 10]. For this purpose, it is important to introduce legislation, that will allow people to have knowledge and control of what can be monitored, who is performing the monitoring process and when and who can have access to this information. The collected data should only be collected from authorised parties, and should not be stored indefinitely in the data centres [7].

On the other hand, sensor networks, that monitor buildings, assets and transportation, collect information about every object/individual entering the network coverage area. This is a case where individuals do not have control over the personal data that can be possibly collected. Similarly with what have been discussed above, policy makers need to ensure that the data are "anonymised" in order to protect people's privacy, and can only be used through official legal processes, when required [7].

Another interesting way for ensuring privacy and reducing the storage cost is the process of *data revocation* or else known *data forgetting* [7, 8, 258], were data are periodically deleted, as seen in [259, 260], where by discarding encryption keys that are not required anymore, the data encrypted using this key are essentially "forgotten" and could potentially provide a secure overwrite. This method can be used in cases where and encryption key is created per file, which no data are shared between files [260].

After reviewing all aspects of IoT in terms of architecture, applications and challenges, it is important to present an overview of data processing, data analysis, signal processing and machine learning techniques, that could be used for processing IoT information.

C.3 IoT Data Analytics

IoT, as already discussed in Section C.2, allows the incorporation of smart sensors implemented for different applications, and produces a large amount of diverse data. Thus, it is essential to investigate ways of processing these data through data mining techniques and propose models suitable for IoT [237, 261–270].

C.3.1 Data Mining Models in IoT

According to [261–263], data mining models for IoT can be categorised in: *multi-layer model*, *distributed model*, *grid-based model* and *Big Data model*.

C.3.1.1 Multi-layer Model

The *Multi-layer data mining model* for IoT, according to [261–263], consists of four layers: *data collection layer*, *data management layer*, *event processing layer* and *data mining service layer*. *Data collection layer* is responsible for collecting the smart data obtained by the different IoT-enabled devices, such as RFID, sensors, etc. Each type of data, due to their diversity, require different data collection approach. Throughout this layer, various issues need to be addressed, such as energy conservation, data filtering, missing and repeated readings and communication [261–263].

The following layer is the *data management layer*, where the collected data are stored and managed in distributed or centralised databases and data warehouses. The data are saved there, after identification, data abstraction and compression and can be used and further analysed in the next layer called *event processing layer* [261–263]. This layer is responsible for analysing events that occur in the different data obtained through IoT devices. According to [261–263], after event filtering, the data can be organised, and analysed according to the occurring events.

The final layer of this model is the *data mining service layer*, that utilises the information extracted through the data management and event processing layers [261–263]. Various data mining techniques, such as classification, clustering and patterns mining, are used in order to obtain useful information that can be used in different IoT applications, such as supply chain management, inventory and optimisation [261].

C.3.1.2 Distributed Model

IoT-enabled devices produce a massive amount of data, that are distributed around the IoT network, and they are both time and position sensitive, thus the adoption of a centralised data mining model would be very challenging [261–263]. Those massive datasets are stored in various distributed databases and require real-time pre-processing, which could be highly expensive for the case of the centralised architecture, due to the hardware requirements [261–263]. Furthermore, the concentration of the data in central nodes could create concerns about data security and privacy, fault-tolerance, business competition and about the correct enforcement of laws and regulations regarding data and privacy [261–263]. The resources of the distributed nodes are limited, thus it is important to allocate these resources more efficiently, which does not happen when used only for energy-costly transmissions of data to the central nodes [261–263]. Estimates of parameters would be sufficient for the central nodes, therefore data pre-processing and data mining could be performed in the distributed nodes, which would then forward only the necessary information [261–263].

As discussed in [261–263], the *distributed data mining model* will consist of a global control node, that will identify the suitable data mining algorithms and datasets that require mining. The distributed nodes will receive these "*guidelines*" and will collect the relevant data and perform data pre-processing using filtering, data abstraction and data compression. The processed data are saved in local databases, and will be further analysed using event filtering, event detection and data mining, creating local models using the "*guidelines*" defined by the global node [261–263]. Bi-directional communication will be established between the local nodes, which will allow data and knowledge exchange. The global node will be able to control the processes in the local nodes and acquire useful information through the local model, when required [261–263]. The distributed model requires less storage capacity, due to the distributed data storage, and less computational power, as the data pre-processing and mining are performed locally in a smaller scale compared to the case of the centralised model [261–263].

C.3.1.3 Grid-based Model

Grid computing utilises distributed resources, such as computing resources, data resources and various devices, in order to perform heterogeneous, large-scale and high performance tasks [261, 262]. The IoT framework consists of various distributed IoT-enabled devices, that could be used as resources similarly to the grid computing concept and therefore allow data mining process in a grid-based model [261, 262].

Brezany *et al.* [271] proposed a grid-based infrastructure called *GridMiner*, that allows online analytical processing and data mining using distributed resources. Another interesting approach was proposed in [272], where the authors designed and implemented a *Web Services Resource Framework* (WSRF)-based grid, which treats the various grid applications, such as data mining, as web services.

Furthermore, Stankovski *et al.* [273], developed the *DataMiningGrid*, which can incorporate various programs and applications in a single but flexible and scalable framework. It has a *service-oriented architecture* (SOA) and can support both *Open Grid Service Architecture* (OGSA) and WRSF. Based on this model, Bin *et al.* [261] proposed a DataMiningGrid-based data mining model suitable for IoT, that can incorporate the various hardware resources of IoT, such as RFID tags and readers, WSNs, etc., and can support a plethora of IoT software applications, such as event processing algorithms, and data mining applications.

C.3.1.4 Big Data Model

Big Data represent the data that are either too large or complex to be analysed using traditional data processing methods, which is the case for IoT data. In recent years, many companies, such as IBM, Microsoft, Tera data and Amazon, have designed and implemented big data systems, and have released their products in order to enhance their services [266]. *MapReduce*, *Apache Hadoop* and *Apache Stark* are some of the main methods that are currently used for big data analytics [266].

MapReduce can process large-scale data in the form of 'n' number of clusters, and uses two kernel elements -namely mappers and reducers in the programming model [266]. Map function creates a set of key/value pairs and reduce function merges all the intermediate values of the key. Every node of both map and reduce is independent

from the other parallel nodes, that use different data and key [266]. Similarly, *Apache Hadoop* is also designed to process large-scale data, by combining a *Hadoop Distributed File System* (HDFS) and the *MapReduce* programming framework [266]. HDFS file system splits the large data in a cluster node by creating small blocks, that are then distributed to the other nodes. *MapReduce* is then used for the smaller blocks, which reduces the complexity of data processing for big data [266].

Furthermore, *Apache Spark* is an extension of the *Apache Hadoop* and allows the use of general purpose programming language for processing big data on cluster nodes [266]. Various machine learning algorithms and interactive data analysis tools can be used through this framework. *Apache Spark* has introduced the use of *resilient distributed datasets* (RDD), parallel operation for data, shared variables and *spark context* [266].

Driven by the concept of big data in the IoT and in order to ensure the security of the framework, Singh *et al.* [262] proposed a 5-layer *big data mining model*. The first layer represents the devices present in IoT, such as sensors, RFID, and cameras, that create and collect data in real-time. The following layer integrates the raw data, in the form of structured, semistructured and unstructured data. This is followed by the data gathering layer, where real-time data are parsed, analysed and merged. Data processing is performed using various big data open source solutions, such as Hadoop and HDFS. The service layer provides as services various data mining methods, such as classification, clustering and association analysis. All layers have an enhanced security, in order to ensure privacy and minimise illegal data access [262].

C.3.2 Data Collection in IoT

Data collection is one of the most important aspects in IoT, as all the smart IoT-enabled devices will produce a plethora of data. According to Wlodarczak *et al.* [264], data collection in IoT is performed at the device level, as the majority of the devices will have the capability of collecting their corresponding data and transmitting them for storage to distributed databases and data warehouses. In order to reduce the amount of data transmission and optimise the use of the available resources, techniques, such as data aggregation, can be implemented for the devices that produce a large amount of data [261, 262, 264].

C.3.3 Data Pre-processing in IoT

The obtained IoT data, similarly with all types of data, may exhibit inaccuracies and repetitions, that may be caused by a failure during the data collection process or even by malicious attacks, as already discussed in Section C.2.3.3. The pre-processing steps for IoT would include removal of noisy, inconsistent and incomplete data and filtering of duplicate data [262, 268]. Furthermore, data reduction techniques are required in order to achieve more efficient event detection and data mining processes, as handling massive data is extremely complex, as seen in Section C.3.1. Data reduction can be performed by using techniques to remove redundant data or even data aggregation, when required [261, 262, 265, 268].

C.3.4 Event Detection and Filtering in IoT

Event detection and filtering are essential for processing IoT data, as they can extract useful information from them. Event detection allows the categorisation of activities/events occurring in the data, such as the operation of a smart home appliance and the change of the RFID status of a material/product. These events can be further filtered using event filtering, in order to obtain the valuable events, that can be used as features in the data mining process [261, 262]. A more detailed explanation of event detection for the case of NILM applications can be found in Chapter 2.

C.3.5 Data Mining Algorithms for IoT applications

Data Mining is the process, that allows the extraction of useful information from raw datasets using intelligent methods/algorithms to identify patterns and trends [237]. Data mining process has to be customised, in order to accommodate the variability of IoT data, which can be achieved by techniques ranging from traditional signal processing methods to machine learning. Although the range of available data mining techniques is quite vast, this section is focused on introducing *clustering*, *classification* and *association rules*, which are the most common methods used for data mining and could be suitable for the case of IoT, as seen in [237, 239, 265, 266, 268, 269, 274].

C.3.5.1 Clustering

Clustering separates data into different groups and does not require prior knowledge of the data characteristics, thus it can be categorised as *unsupervised learning* method [239, 268, 274]. Members of the same *cluster* exhibit similar patterns, whereas their patterns are different in the same sense when compared to the rest of the clusters [239, 268, 274].

According to Chen *et al.* [274], the most common categories in clustering are the *hierarchical clustering* and the *partitioning clustering*. *Hierarchical clustering* combines objects into clusters, which are further grouped into larger clusters, in order to create a *hierarchy tree* [268, 274]. This clustering method can be performed either using an *agglomerative* (bottom-up) or *divisive* (top-bottom) approach. *Agglomerative clustering* starts with one object clusters and after each iteration it combines clusters with similar patterns into larger clusters, until all objects are members of one of the larger clusters [274]. Variations of hierarchical agglomerative clustering have been used for NILM applications in both [154] and [88]. *Divisive clustering*, on the other hand, starts with one cluster containing all data, and continues by splitting this cluster into smaller clusters, depending on the similarity of patterns [274].

Clustering Using Representatives (CURE) [275, 276] and *Singular Value Decomposition* (SVD) [277] are some of the most commonly used hierarchical clustering algorithms [274]. According to Bangui *et al.* [278], CURE could be a suitable clustering method for IoT, as it can be efficiently used for clustering large data, as seen in [275, 276, 279]. SVD is generally used in signal processing, image processing [280] and big data analytics [281], thus it could be applied for clustering IoT data [274]. More recently, Garcia *et al.* [282] have applied SVD-based initialisation in order to perform NILM using *Non-negative matrix factorisation* (NMF).

Partitioning clustering algorithms, as defined in [274],“ *discover clusters either by relocating points between subsets or by identifying areas heavily populated with data*”. One of the most commonly used partitioning clustering method is *k-means*, which can be used for big data analytics [237, 239, 266] and have been applied in various NILM methods [87, 117, 154]. K-means clustering aims to group data points into clusters, normally by using the distance between data points, in order to identify similarity in

the clusters [269].

Another interesting clustering approach is the *Density Based Spatial Clustering of Applications with Noise* (DBSCAN), where clusters are areas with high object density, and the areas with low density (small number of objects) are treated as outliers [269, 283, 284]. In the context of IoT, DBSCAN has been used in [285] as a part of a NILM application. More specifically, Zheng *et al.* [285] have performed event detection using a windowed DBSCAN-based clustering method, where both active power and fundamental current RMS curve were used in the format of a predefined "*window*".

C.3.5.2 Classification

Classification is using different classifier models in order to predict the future patterns by categorising them in predefined classes [268, 274]. The identification of the predefined classes, normally occurs through training or a priori knowledge, and due to this attribute, classification is generally categorised as a *supervised learning technique* [268, 274]. Some of the most common classification methods include *decision tree*, *k-Nearest Neighbour*, *Support Vector Machine* and *Bayesian Networks*.

- *Decision Tree*

  A *decision tree* (DT) is a tree-like structure, that consists of nodes and branches connecting the nodes and uses a top-down approach, where the top node, also known as *root*, contains the training data/attributes that are going to be used for classification [265, 274, 283, 284, 286, 287]. The final nodes in a DT architecture are called *leaves* and they represent the different classes created using the already defined attributes [283, 284, 287]. The remaining nodes throughout the tree are known as *decision nodes* and are used for testing the different attributes [265, 274, 283, 284, 287]. Each *branch* represents the different outcomes possible from each node and every path in the DT represents different *classification rules*.

  Iterative Dichotomiser 3 (ID3) is a simple DT algorithm originally developed by Quinlan in [288], that allows classification using different attributes. This algorithm can only check one attribute at a time, which can result in large and incomprehensible trees [274, 283, 284]. Quinlan proposed C4.5 in [289], which is an improved version of ID3 and allows the use of both numerical and categorical

attributes, handling of missing/incomplete values and can support *pruning* [274, 283, 284]. *Pruning* is the process of discarding subtrees, that do not assist in the classification process, which can reduce complexity and increase the classification accuracy by reducing the predicted error rate [283, 284]. Both ID3 and C4.5 are selecting the attributes that increase the *information gain* [283, 284].

Chi-squared automatic interaction detector (CHAID) is a DT method that splits the data in exclusive and exhaustive subsets depending on the dependent variable, using a number of independent variables, known as *predictors* [274, 290, 291]. CHAID uses the predictors in order to identify the best way the categories can be merged using *Bonferroni testing* [274, 290, 291]. Another interesting DT algorithm is the *Classification and Regression Trees* (CART), which can build both classification and regression trees, depending on the type of the attributes, categorical or continuous [274]. CART uses the *Gini diversity index* as a spitting criterion, instead of information gain, as it can be computed much faster [284]. *Gini index* is calculated for all possible features, and the feature with the minimum value is the point where the DT splits [284]. It can support both multi-class problems and *pruning* by using *twoing splitting criterion* and a method called *minimal cost complexity pruning* respectively [284]. Furthermore, CART can identify and remove outliers and noisy data and handle missing and incomplete data [284].

Variations of the above DT algorithms can be used for classification purposes in IoT applications, such as NILM. Berges *et al.* [66] have used a DT classifier for pattern matching using both transient and steady state features, whereas the authors in [32] have proposed a low complexity DT-based NILM method with high disaggregation accuracy. A combined DT induction and AdaBoost algorithm scheme was implemented in [123] and a similar DT method, using greedy Hunt's algorithm was proposed in [124]. Mei *et al.* [292] have used a *Random forest*-based NILM technique for load identification. *Random forests* combine the idea of bagging (*Bootstrap Aggregating*) and CART algorithm, and allow the creation of random DT in every iteration of the bagging algorithm [287]. More details regarding the NILM methods using DTs, can be found in Section 2.6.1.5.

- *k-Nearest Neighbour*

  The *k-Nearest Neighbour* (k-NN) algorithm is used for both classification and regression purposes, by searching the k-nearest points of the observed data. The metrics normally used in order to identify the closest neighbours is the *euclidean distance* for continuous data and *Hamming distance* for discrete data [274, 283, 286]. There is a plethora of variations of the traditional k-NN algorithm, but as the purpose of this thesis is not to investigate those improved k-NN methods, the reader can refer to [274] for more details. Furthermore, k-NN can be combined with different machine learning and classification techniques, such as C4.5, Bayes classifiers, *support vector machine*, fuzzy logic, *artificial neural networks* (ANN) and *Hidden Markov-Models* (HMM) [47, 107, 293].

  K-NN-based algorithms can be applied for classification using big data, as seen in [237, 239, 269], and therefore could be a potential solution for classifying IoT data. Moreover, many NILM methods are using k-NN either as the only classifier or as a part of combined scheme for load classification. 1-NN-based NILM methods have been proposed in [27, 63, 66, 89, 114], 5-NN in [79, 106], a 10-NN in [114] and a k-NN-based approach in [98]. Kolter *et al.* [47] have implemented a combined scheme using both k-NN and HMM and Azaza *et al.* [107] have used a combination of *dynamic fuzzy c-means clustering* (dFCM) and k-NN. Furthermore, a *Probabilistic K-Nearest Neighbour* (PKNN) method was proposed in [108], which identifies the nearest neighbour through its probability. A more in depth review of the above NILM methods can be found in Section 2.6.1.3.

- *Support Vector Machine*

  Support Vector Machine (SVM) is a supervised learning algorithm, that uses labeled training data for learning purposes and can be applied for both classification and regression [274, 284, 286]. SVMs use decision planes for classification, where the decision boundaries define the different classes of data samples [284].

  Similarly with k-NN, SVMs can be applied successfully in the field of big data analytics and IoT applications, such as NILM [237, 239, 269]. A plethora of SVM NILM methods have been implemented using different kernels such as

linear, *radial-basis functions* (RBF), polynomial and Gaussian for classification and comparison purposes [25, 79, 81, 87, 98, 106, 108, 109, 112, 114, 116, 119, 120]. A *Hierarchical SVM* (HSVM) was proposed in [110] and a *Multiple class*-SVM (M-SVM) using Gaussian kernel function in [111]. Lai *et al.* [115] have proposed a hybrid *Support Vector Machine/Gaussian Mixture Model* (SVM/GMM) and Altrabalsi *et al.* [117, 118] have used a hybrid NILM method based on a combination of SVM and k-means. A further discussion of the different SVM NILM methods can be found in 2.6.1.4.

- *Bayesian Networks and Classifiers*

  Bayesian networks are directed acyclic networks, where each node represents a random variable with probabilistic distribution, that is defined through *Bayes Theorem* [274, 284, 286]. The edges of the network show the conditional dependencies, whereas non connected nodes are conditionally independent [274, 284, 286]. The *Bayes classifiers* can be applied for classification purposes in order to accommodate complex data, such as big data and consequently data derived from various IoT applications [237, 239, 269, 274].

  More specifically, in the NILM research, which will be discussed in 2.6.1.2, *Bayes classifiers*, such as *Naive Bayes classifiers* have been used for load identification and classification performance comparison, such as in [76, 88, 113, 120]. Berges *et al.* [66] have implemented a *Gaussian Naive Bayesian* (GNB) classification approach, whereas the authors in [104] have used a *hierarchical Bayesian* approach. Another interesting Bayesian-based method is the method proposed in [105], which incorporates *Dynamic Bayesian Networks* (DBN) and a Bayes filter with GMM for NILM purposes.

C.3.5.3 Association Rules

Association rule mining is an unsupervised data mining technique and focuses on identifying *associations*, such as relationships and dependencies, in large data [237, 265, 266, 274, 283, 284]. It is commonly used in *market basket analysis* and *transaction data analysis*, as seen in [274, 283, 284]. *Apriori algorithm* is using a bottom-up approach in order to identify frequent items in a dataset. Apriori algorithm assumes

that all subsets of frequent itemsets are frequent and for every infrequent itemset, its supersets are also infrequent [283]. *Support* and *confidence* are employed as measures, in order to select appropriate association rules [283, 284]. Improvements of the apriori algorithm include *hashing*, *transaction removal*, *data set partitioning*, *sampling* and *mining frequent itemsets without generation of candidate itemsets*, which can be found in [283].

If the items of the dataset follow a hierarchical topology, the association rules will have stronger support on the top and weaker closer to the bottom of the hierarchy [284]. *Multilevel association rules* (MAR) are suitable for these cases, as they follow a top-bottom approach, and can be separated into the methods using *uniform support* and those using *reduced support* [284]. *Uniform support* methods use the same minimum support threshold for all levels of the hierarchy, whereas *reduced support* methods use level-specific minimum support thresholds [284]. Normally, lower levels will have smaller thresholds [284]. *Multidimensional and quantitative association rules* can also be employed for cases of multidimensional datasets and non-Boolean data [284].

Frequent pattern growth (FP-growth) algorithm is another common algorithm for performing association rules data mining, which compresses the sets of frequent items into a tree structure, known as *frequent pattern tree* (FP-tree) [283, 284]. It is a more scalable, efficient and faster solution for mining frequent items than apriori algorithm, as it searches the compressed FP-tree for identifying the frequent items, instead the whole dataset [283]. Furthermore, FP-growth has the ability to remove infrequent items, which is not the case of the original apriori algorithm [283].

Both apriori and FP-growth based data mining methods have been already implemented for mining big data, such as *FiDoop* using *MapReduce*, *FIM* on *Hadoop* and *R-Apriori* [266]. Similarly with the market-basket and transaction applications, association rules mining could be used for IoT applications, such as identifying customer behaviour using RFID data and providing personalised services [265].

C.4 Summary

The focus of this chapter was to provide a brief overview of the IoT concept, the technologies that enable its implementation, the applications and challenges that emerge, and the data mining and data processing techniques, that can be applied for extracting meaningful information from the IoT obtained data.

IoT interconnects various fields and technologies in the form of a distributed network. RFID technology is one of the IoT enabling technologies, as it is widely used for identification and tracking of objects and even people in a plethora of industries, such as transportation, logistics, healthcare and agriculture, to name a few. Similarly, WSNs and their distributed nature are suitable for IoT, as they allow the continuous monitoring of various parameters depending on the application and do not require manual intervention for acquiring the measured data. Furthermore, a middleware based on a SOA approach would be able to support the diversity of IoT objects, devices, applications and services. Distributed and cloud-based data storage systems, appropriate data mining methods and data visualisation are an integral part of the IoT, due to the quantity and variability of the data.

IoT applications can generally be categorised in: (a) transportation and logistics, (b) healthcare, (c) smart environment and (d) personal and social domain, as defined in [7]. Examples of applications in all different domains have been discussed throughout this chapter, and allow the reader to understand both the usefulness of an IoT implementation and the challenges that arise with that. IoT challenges include data management and storage, data mining, security and privacy. Due to the nature of the IoT data, sensitive and important information will be exchanged throughout the network, from both private and commercial users, thus it is important to fortify the security of the IoT from various malicious attacks and users' privacy.